UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Structural studies of the SARS-CoV-2 M protein

Permalink

https://escholarship.org/uc/item/1gg8g6t9

Author

Dolan, Kimberly

Publication Date

2024

Peer reviewed|Thesis/dissertation

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

Structural studies of the SARS-CoV-2 M protein

By

Kimberly Dolan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stephen Brohawn, Chair
Professor James Cate
Professor Eva Nogales
Professor Robert Saxton

Summer 2024

Structural studies of the SARS-CoV-2 M protein

©2024

By
Kimberly Dolan

Abstract

Structural studies of the SARS-CoV-2 M protein

By

Kimberly Dolan

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Stephen Brohawn, Chair

The coronavirus genome encodes four structural proteins that are incorporated into virions, spike (S), envelope (E), nucleocapsid (N), and membrane (M). M is the most abundant structural protein and is known to drive viral particle assembly and budding through direct protein-protein interactions. Throughout the COVID-19 pandemic caused by the virus SARS-CoV-2, M has displayed high sequence conservation, making it a compelling target for vaccines and therapeutics. Earlier studies of viral particles from related betacoronaviruses had suggested that M may be able to adopt two distinct conformations – a compact and elongated form – yet prior to 2019, the structure and molecular basis for its functions in infectious particle formation were unknown. The work presented in this dissertation focuses on the structural characterization of the M protein from SARS-CoV-2 to gain insight into how M functions in the viral life cycle and subsequently leverage our findings to develop new avenues to combat disease.

First, we develop routines to express and purify full-length M protein and determine its structure in nanodiscs to describe its architecture in a lipid environment using cryo-electron microscopy (cryo-EM). We show that M is a 50 kDa homodimer that is structurally homologous to an accessory SARS-CoV-2 protein, ORF3a, suggesting that the two share a common ancestor. Subtle structural differences between the two proteins lend insight into how they have evolved to fill drastically different roles in the coronavirus life cycle. One striking feature of M is the overwhelmingly electropositive cytosolic surface that is likely important for protein-protein interactions in viral assembly, as well allowing for close contact with the dense core of viral RNA. We performed molecular dynamics simulations and showed that this conformation is stable in a simple lipid bilayer. Together, these results produced the first atomic model of a coronavirus M protein and provide insight into roles for M in viral assembly and structure.

The next chapter of this thesis describes the structural basis and biophysical characterization of a neutralizing antibody that targets the N-terminus of the SARS-CoV-

2 M protein. We used biolayer-interferometry (BLI) to show that the SARS37 neutralizing antibody specifically recognizes the alternate long conformation of M, rather than the short form we had previously characterized in nanodiscs. We used cryo-EM to determine the structure of the SARS37 Fab in complex with long form M and show that the epitope involves residues from the N-termini of both subunits, as well as the outer surface of the transmembrane domain through the insertion of a long CDR3 loop into the lipid bilayer. In agreement with our BLI data, we found that the complex formed between the SARS37 Fab and M in the short conformation was structurally heterogeneous and unstable. Importantly, we find that the residues of M that the Fab recognizes are highly conserved, remaining unchanged as the virus has evolved. Given the strong and stable preference for the long conformation, we infer that this form of M is present in circulating viral particles presented to the immune system and can be targeted to prevent further infection.

The last chapter of this dissertation investigates the role specific M-lipid interactions may play in the process of viral assembly. We show M directly binds a Golgi resident anionic lipid, ceramide-1-phosphate (C1P) and investigate the effect that C1P binding has on M conformational dynamics. We find an overall stabilizing effect of C1P binding to M in the short conformation and describe a specific C1P binding site that involves M residues spanning the transmembrane, hinge and C-terminal domains. Comparing our C1P bound structure with the previously determined apo short conformation M structures, we observe a previously uncharacterized potassium ion binding site and a series of side chain rotamer changes distributed across the C-terminal domain and dimer interface. We design point mutants at these sites and use AlphaFold to show that the predicted models exist in a continuum between the established short and long conformations. The residues involved are highly conserved amongst sarbecoviruses, suggesting a conserved potential principle underlying M protein conformational dynamics.

Table of Contents

Abstract **1**
Table of Contents **i**
List of Figures **iii**
Dedication **v**
Acknowledgements **vi**

Chapter One: Introduction **1**
1.1 The Coronavirus family and a brief history of pandemics 1
1.2 The Coronavirus M protein: roles in viral assembly, morphogenesis and budding 2

Chapter Two: Structure of the SARS-CoV-2 M protein in lipid nanodiscs **5**
2.1 Abstract 5
2.2 Introduction 5
2.3 Results 6
2.4 Discussion 12
2.5 Materials and Methods 14
2.5.1 Cloning and protein expression 14
2.5.2 Protein purification 15
2.5.3 Nanodiscs formation 15
2.5.4 Cryo-EM sample preparation and data collection 16
2.5.5 Cryo-EM data processing 16
2.5.6 Modeling, refinement, and analysis 17
2.5.7 Fluorescence size exclusion chromatography (FSEC) 17
2.5.8 Molecular dynamics 17
2.6 Data and reagent availability 18
2.7 Acknowledgements 18
2.8 Contributions 19
2.9 Supplemental tables and figures 19

Chapter Three: Structural insight into a neutralizing antibody that targets the SARS-CoV-2 M protein **25**
3.1 Introduction 25
3.2 Results 27
3.3 Discussion 30
3.4 Materials and Methods 31
3.4.1 Cloning, expression and purification of SARS-CoV-2 M protein 31
3.4.2 Cloning, expression and purification of conformationally selective Fabs 32
3.4.3 M protein – Fab complex purication 33
3.4.4 Cryo-EM sample preparation and data collection 33
3.4.5 Cryo-EM data processing 33
3.4.6 Modeling, refinement and analysis 34

3.4.7 Biolayer interferometry 34
3.5 Acknowledgements 35
3.6 Contributions 35
3.7 Supplemental tables and figures 35

Chapter Four: Direct SARS-CoV-2 M-lipid interaction controls M conformational dynamics and virus assembly 39
4.1 Introduction 39
4.2 Results and Discussion 40
4.3 Conclusion 47
4.4 Materials and Methods 48
4.4.1Cloning, expression, and purification of SARS-CoV-2 M protein and Fabs 48
4.4.2 M protein-Fab complex formation and purification 48
4.4.3 Cryo-EM sample preparation and data collection 48
4.4.4 Cryo-EM data processing 49
4.4.5 Modeling, refinement, and analysis 49
4.5 Supplemental tables and figures 50

Chapter Five: Final conclusions and outlook 56

References 59

List of Figures and Tables

Figure 1.1 Summary of key M protein roles in coronavirus particle assembly. 3
Figure 2.1 Structure of SARS-CoV-2 membrane (M) in lipid nanodiscs. 7
Figure 2.2 SARS-CoV-2 membrane (M) and ORF3a proteins are structurally homologous. 8
Figure 2.3 An enclosed polar pocket between cytosolic domains in membrane (M). 9
Figure 2.4 An electropositive cytosolic surface in membrane (M). 10
Figure 2.5 Molecular dynamics simulations of membrane (M). 12
Supplemental Table 2.1 Cryo-EM data collection, refinement, and validation statistics. 19
Supplemental Figure 2.1 Purification and reconstitution of membrane (M). 20
Supplemental Figure 2.2 Cryo-EM processing and validation. 21
Supplemental Figure 2.3 Comparison of experimentally determined and predicted membrane (M) protein structures. 22
Supplemental Figure 2.4 Sequence alignment of membrane (M) proteins across Coronaviridae. 23
Supplemental Figure 2.5 Analysis of pocket size and lipid interactions from molecular dynamics simulation. 24
Figure 3.1 The M protein N-terminal peptide does not co-crystalize with the SARS37 Fab. 26
Figure 3.2 SARS37 antibody binds SARS-CoV-2 M with higher affinity to the long conformation. 27
Figure 3.3 Structure of the SARS37:M:IcFab complex. 28
Figure 3.4 The SARS37 binding interface is highly conserved. 29
Supplemental Table 3.1 Cryo-EM data collection, refinement and validation statistics. 35
Supplemental Figure 3.1 Cryo-EM data, processing pipeline, and validation For SARS37:M:IcFab complex in LMNG/CHS. 37
Supplemental Figure 3.2 Cryo-EM map for SARS37:M:scFab complex in LMNG/CHS. 38
Figure 4.1 M short conformation is stabilized by C1P binding. 41
Figure 4.2 Differences between our M:scFab + C1P model to 7VGS are distributed throughout the transmembrane and C-terminal domains. 43
Figure 4.3 Conformational change between subunits is directed by a trio of residues at the CTD dimer interface. 45
Figure 4.4 AlphaFold predicted models of CTD mutants show a distribution of conformations between long and short. 46
Supplemental Table 4.1 Cryo-EM data collection, refinement and validation statistics. 50
Supplemental Figure 4.1 The SARS-CoV-2 M protein binds ceramide-1-phosphate (C1P). 51

Supplemental Figure 4.2 Cryo-EM data, processing pipeline, and validation for M:scFab complex in LMNG/CHS/C1P. 52
Supplemental Figure 4.3 Cryo-EM data, processing pipeline, and validation for M:lcFab complex in LMNG/CHS/C1P. 53
Supplemental Figure 4.4 Comparison of atomic displacement parameters (B-factors) for +/- C1P short conformation models. 54
Supplemental Figure 4.5 AlphaFold predicted models of wild-type M dimers plus ions. 54
Supplemental Figure 4.6 Overlay of short and long form M monomers. 55
Supplemental Figure 4.7 Differences in electrostatic surface potential between conformations. 55

Dedication

“You stand at the tip of the high board. High above the sea…” - Samuel Beckett

To Ryan, Cathy, Dad, and Regina – for your endless love and support.

Acknowledgements

To my advisor, Steve Brohawn, thank you for everything. I started my PhD in August of 2018, and it would be an understatement to say that these intervening years have been tumultuous – at home and the world over. So much of your experience in graduate school is dictated by the lab you choose to join, and I have been reminded at every turn that I made the best possible decision in joining the Brohawn lab. It has been a wonderful place to do science and that starts at the top with Steve, who holds us all to high standards and has the patience to allow us to meet them at our own pace. Thank you as well to Brohawn lab members past and present – especially Christopher Hoel, David Kern, Baobin Li, Robert Rietmeijer, Kyle Tucker, and Trevor Docter – for these years of advice, support, commiseration, and laughter. It's genuinely bittersweet to be leaving so soon and it has been the privilege of a lifetime to get to do this work that I love with all of you.

I want to thank the many mentors, staff, and organizers who made my participation in the McNair and BUILD EXITO scholar programs possible during my time at Portland State University, especially Dr. Steve Reichow and Dr. Marilyn Rampersad Mackiewicz. Those programs gave me so much that I can hardly put it into words, but without the least exaggeration I can state that I would not be here at Berkeley without them, and I am forever grateful. You opened my eyes to new ambitions and gave me the confidence to trust that I was worthy of pursuing them. Dr. Mackiewicz is a tireless advocate for others while also knowing her own value and refusing to settle in pursuit of her goals; she is truly an inspiration and model for how to be in this world. My time in the Reichow lab was instrumental in learning how to be a biochemist, and I couldn't really appreciate how well it prepared me to succeed in graduate school until I came to Berkeley. Half of my accomplishments to date are a direct result of Steve's advocating on my behalf without me ever needing to ask him for it, and I can't possibly thank him enough.

Lastly, I would like to thank all the people in the greater Berkeley Biophysics and structural biology communities who make this place what it is. The energizing, creative, and collaborative spirit at Berkeley is everything I hoped for when I first arrived, and I feel so lucky to have gotten to be a part of such a vibrant and inspiring scientific community. A big part of that is the many opportunities to share our science outside of our labs through supergroups, retreats, and social events, none of which would be possible without the efforts of so many dedicated student and faculty organizers. Finally, although I think we do a pretty good job of thanking Kate Chase whenever the opportunity arises, it is still never enough. So, thanks again to Kate for her competence and caring.

Chapter One
Introduction

1.1 The Coronavirus family and a brief history of pandemics

Coronaviridae is a large family of related lipid enveloped, positive-sense RNA viruses. The first identified member of the family was reported in a 1931 paper chronicling an outbreak of an acute and highly deadly respiratory disease infecting newly born domesticated chicks in the central and western united states[1,2]. Since then, the family has expanded to include hundreds of identified members that can infect a wide variety of mammals and birds, and cause diseases of varying severity[3,4]. As the *Coronaviridae* family has grown, it has been divided into the subfamilies *Letovirinae* and *Orthocoronovirinae*[5], the latter of which are commonly referred to as coronaviruses (CoVs). Coronavirus infections often include upper respiratory symptoms such as those reported in 1931, but can vary widely and involve multiple organs, such as encephalitis, hepatitis, enteritis, or severe diarrhea and dehydration[6,7].

The coronaviruses subfamily has been further divided into four genera: alpha-, beta-, delta-, and gammacoronavirus[8]. The delta and gamma genera are primarily found circulating in avian and porcine gene pools and have only very recently been reported to be able to infect humans[9,10]. Even before its potential for zoonosis was characterized, porcine deltacoronavirus posed a particular concern in factory farming due to severe economic losses and food chain disruption stemming from its high mortality rate in piglets[11]. Alpha- and betacoronaviruses descend from bat gene pools and are known to infect a much wider array of mammals[8]. The first description of a coronavirus infection in humans was published in 1965[12]. Since then, seven coronaviruses have been found to be transmissible between humans. Alphacoronaviruses NL63 and 229E, and betacoronaviruses HKU1 and OC43 are endemic in humans and are among the viruses collectively associated with the common cold[13]. In recent decades, three highly-pathogenic betacoronaviruses have led to epidemics: severe acute respiratory syndrome coronavirus (SARS-CoV-1)[14], Middle East respiratory syndrome coronavirus (MERS-CoV)[15], and finally, SARS-CoV-2.

SARS-CoV-1, which first emerged in Guangdong, China in 2002[16], was responsible for the 2002-2004 SARS outbreak. The illness is characterized by initial symptoms of headache and body aches, followed by the onset of respiratory symptoms. While the disease is relatively mild in young children, mortality rates in patients over 60 years old of more than 40% have been recorded[17]. SARS is most transmissible when patients are symptomatic and thus the outbreak was largely contained through public health measures such as testing, isolating sick patients, and restricting travel to and from affected areas[18]. Ultimately, more than 8,000 infections and at least 700 SARS-related deaths have been confirmed worldwide[19].

The first recorded case of Middle East respiratory syndrome (MERS) caused by MERS-CoV was in June 2012 in Jeddah, Saudi Arabia[15]. MERS-CoV causes an acute and highly lethal pneumonia as well as renal failure in many cases[20]. Case fatality rates for MERS are estimated at about 35% among laboratory confirmed cases[21]. Since 2012 there have been over 2600 laboratory confirmed infections and 943 associated deaths[22]. Compared to SARS, the person-to-person transmission for MERS is relatively low, however MERS continuously circulates amongst dromedary camels in the Middle East and frequent spillover events persist largely on the Arabian peninsula[23].

Most recently, coronavirus disease 2019 (COVID-19) caused by the virus SARS-CoV-2, emerged in Wuhan, China in late 2019 and had spread worldwide by early 2020. Since then, at least 775 million cases and 7 million deaths have been confirmed, with true counts estimated to be higher[24]. SARS-CoV-2 is closely related to SARS-CoV-1, both belonging to the sarbecovirus subfamily of betacoronaviruses[25]. The symptoms of COVID-19 are typically fatigue, fever, and cough, but are highly variable and can include heart problems, dizziness, pink eye, the loss of taste and smell, or no symptoms at all[26]. In elderly and immunocompromised patients, COVID-19 can cause severe and deadly pneumonia, hypoxia, and multiorgan failure. Understanding the long-term consequences of COVID-19 infection is still in progress, but evidence shows that a significant portion of patients have persistent symptoms leading to a syndrome known as long covid[27]. The COVID-19 pandemic caused massive social and economic disruption worldwide, led to the largest global recession since the Great Depression, and caused significant food and supply shortages[28,29]. The WHO has recently declared an end to the emergency phase of the pandemic, however COVID-19 continues to circulate. It is more difficult to delineate the end of the pandemic itself, and despite extensive immunity from prior infections and vaccines, and particularly with the continuous emergence of immune-evasive variants, SARS-CoV-2 continues to present a significant disruptive force on normal health care and education operations around the globe.

1.2 The coronavirus M protein: roles in viral assembly, morphogenesis and budding

Coronaviruses have a large ~30 Kb genome that encodes 4 structural proteins – spike (S), nucleocapsid (N), envelope (E), and membrane (M) – as well as 16 non-structural proteins (nsp1-16)[30]. Infectious coronavirus particles, or virions, are often described as spherical but show a degree of pleomorphism in their shape and diameter which has been measured ranging from 70 to 120 nM[31,32]. Assembled, mature coronavirus virions contain a dense core of N protein sheltered viral RNA surrounded by a lipid envelope that is packed with structural membrane proteins S, M, and E[31]. Spike protein ectodomains project from the virion surface, giving the viral particles their eponymous crown.

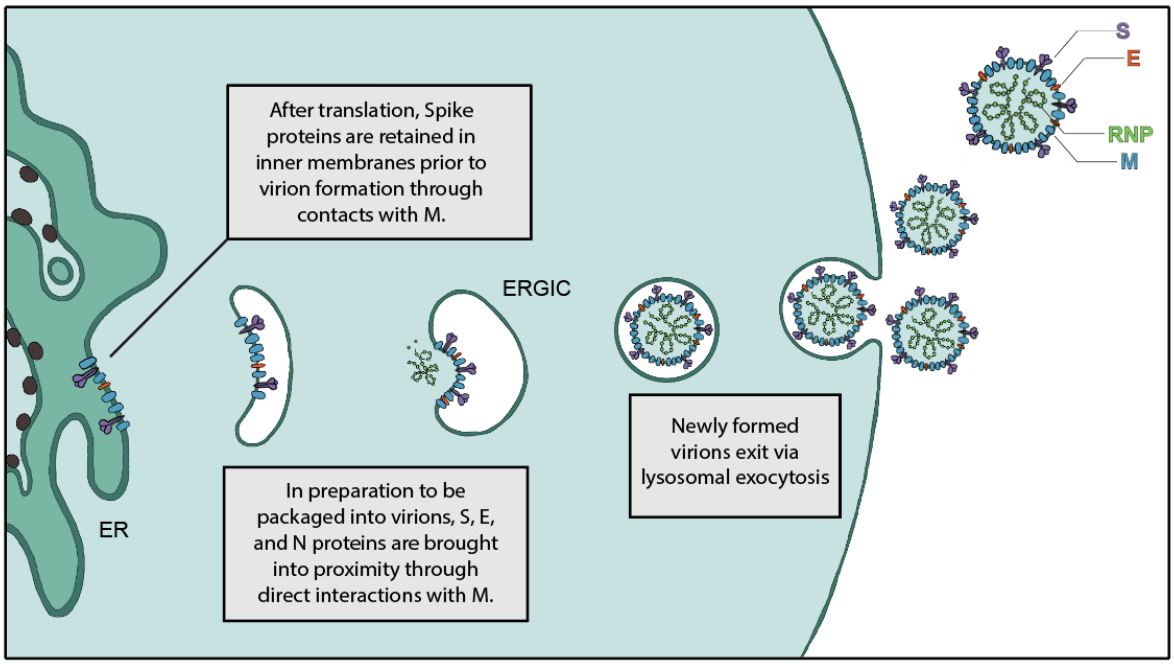


Figure 1.1 Summary of key M protein roles in coronavirus particle assembly.

The coronavirus membrane (M) protein, previously referred to as the E1 or matrix protein, is the most abundant structural membrane protein and performs essential roles in viral assembly, morphogenesis, and egress[33]. While M is not the only membrane protein encoded in the coronavirus genome, I will refer to it as "the" membrane protein throughout this dissertation in keeping with the current naming convention. In infected cells, M directs the other structural proteins, S and E, and ribonucleoprotein complexes (RNP) to the budding site of the ERGIC membrane through direct protein-protein interactions[34–36]. This role is particularly important in the case of Spike proteins, which show a propensity for trafficking away from the budding site to the plasma membrane, leading to the formation of syncytia[37–39]. M proteins interact with the cytoplasmic tail of spike to retain it in inner membranes to be packaged into virions, and co-expression of M along with S results in a significant reduction of syncytia formation[38].

In addition to ensuring that all structural proteins are properly incorporated into newly forming virions, M also dictates virion morphology through protein-protein interactions and conformational changes. Electron microscopy studies on SARS-CoV, murine hepatitis virus (MHV) and feline coronavirus (FCoV) found regions of virions with clear evidence of specific supramolecular organization despite their overall pleomorphism[31,40]. The order appeared to be driven by S:M, M:M, and M:N or M:RNP interactions, and gave rise to a lattice-like structure with an approximate 1:8:4 S_{trimer}:M_{dimer}:N ratio at the envelope interface (Fig 1.1). In virus-like particles (VLPs) prepared by co-expressing S, M, E, and N proteins, these lattices were not seen to be

evenly distributed over the entire particle surface. The particles included small M protein-free areas which could represent budding scars. Low-resolution 2-dimensional class averages revealed that the M protein appeared to be able to adopt two distinct conformations, one compact form that does not appear to contact either N or other M proteins (M_{short}), and the elongated form seen in the lattice-like regions (M_{long}). Viral spike proteins, however, can be seen in regions containing either M conformer, but not in areas without M protein. Regions of the VLPs populated by M_{short} proteins also showed a higher degree of membrane curvature whereas M_{long} appears in areas with a higher degree of rigidity, order, and density of spike proteins[40]. Owing to the relative resolution limits inherent in single particle electron microscopy data collection and processing from 15 years ago compared to today[41], the molecular details of all the interactions described here were unable to be determined at the time.

Finally, M is essential for viral particle budding. While expression of SARS-CoV-2 M alone is unable to generate VLPs, co-expression with N or S proteins are the minimal component required to produce VLPs in multiple expression cell types[43]. In SARS-CoV, VLP formation requires co-expression of either M and N or M and E[44], MERS VLPs can be produced with M, S, and E[45], and for MHV, M and E are required but VLP production is enhanced by the addition of N[46]. The molecular determinants of viral egress are not well understood and how exactly the M protein functions to promote this critical step in the virus life cycle remains an interesting and open question. However, a recent preprint characterizing a small molecule that binds to the transmembrane domain of M_{short}, shows that M can effectively be targeted to inhibit viral egress[42]. A cryo-EM structure of this drug bound to M suggests that drug binding would block the conformational switch from M_{short} to M_{long}, indicating that the ability to undergo this change is an essential step in the virion formation process.

The M protein is indispensable for viral replication, and yet at the outset of the COVID-19 pandemic we had very little insight its structure and function. In this thesis, I describe the structure of the SARS-CoV-2 M protein in lipid nanodiscs and use molecular dynamics simulations to provide critical insight into M function and dynamics (Chapter 2). In Chapter 3 I structurally characterize a neutralizing antibody that targets the M protein and show that it specifically recognizes the M_{long} conformation, demonstrating that this conformation is presented to the immune system during infection and can be targeted to treat disease. Finally, in Chapter 4 I investigate the specific interaction between M and a resident Golgi anionic lipid ceramide-1-phosphate (C1P). I show that C1P binds to a site on M_{short} that bridges the transmembrane, hinge, and C-terminal domains. C1P binding appears to compact and stabilize M_{short} and my structure reveals a novel K+ binding site. I show that the C1P binding site is not present in the M_{long} conformation, suggesting that C1P may be important for regulating the timing of conformational interconversion. Together, this work enhances our understanding of the structure and conformational dynamics of this critical protein.

Chapter Two

Structure of the SARS-CoV-2 M protein in lipid nanodiscs

This chapter is adapted, with permission, from work previously published: Kimberly A Dolan, Mandira Dutta, David M Kern, Abhay Kotecha, Gregory A Voth, Stephen G Brohawn (2022) Structure of SARS-CoV-2 M protein in lipid nanodiscs eLife doi: 10.7554/eLife.81702

2.1 Abstract

SARS-CoV-2 encodes four structural proteins incorporated into virions, spike (S), envelope (E), nucleocapsid (N), and membrane (M). M plays an essential role in viral assembly by organizing other structural proteins through physical interactions and directing them to sites of viral budding. As the most abundant protein in the viral envelope and a target of patient antibodies, M is a compelling target for vaccines and therapeutics. Still, the structure of M and molecular basis for its role in virion formation are unknown. Here, we present the cryo-EM structure of SARS-CoV-2 M in lipid nanodiscs to 3.5 Å resolution. M forms a 50 kDa homodimer that is structurally related to the SARS-CoV-2 ORF3a viroporin, suggesting a shared ancestral origin. Structural comparisons reveal how intersubunit gaps create a small, enclosed pocket in M and large open cavity in ORF3a, consistent with a structural role and ion channel activity, respectively. M displays a strikingly electropositive cytosolic surface that may be important for interactions with N, S, and viral RNA. Molecular dynamics simulations show a high degree of structural rigidity in a simple lipid bilayer and support a role for M homodimers in scaffolding viral assembly. Together, these results provide insight into roles for M in coronavirus assembly and structure.

2.2 Introduction

Coronaviruses encode four structural proteins that are incorporated into mature enveloped virions: the transmembrane spike (S), membrane (M), and envelope (E) proteins and the soluble nucleocapsid (N) protein[30]. S proteins protrude from the virion, creating the eponymous corona in electron micrographs, and mediate fusion of viral and host cell membranes. E proteins form cationic viroporins that promote viral assembly and modulate the host immune response. N is an RNA-binding protein that packages the viral RNA genome. M organizes the assembly and structure of new virions and is essential for virus formation[47–50]. M is the most abundant membrane protein in the viral envelope and anti-M antibodies are found in plasma of patients infected with SARS-CoV-2 and other coronaviruses[51–56]. Based on its functional importance and immunogenicity, M has been proposed as a target for coronavirus vaccines or therapeutics.

In infected cells, M mediates virus assembly and budding by interacting with all other structural proteins and directing their localization to the ER-Golgi intermediate

compartment (ERGIC)[30,34,36,37,44,57–60]. M is proposed to interact with E through its transmembrane region and S and N through a cytosolic C-terminal region[34,36,37,60]. ER export and Golgi localization sequences in M determine its subcellular localization and M, in turn, modulates localization and posttranslational processing of S to promote virion assembly[37,61]. Across a wide range of coronaviruses (including SARS-CoV-2, SARS-CoV-1, MERS, mouse hepatitis virus (MHV), infectious bronchitis virus, and transmittable gastroenteritis virus), M is required for minimal virus-like particle (VLP) formation in transfected cells[43,44,62–64]. M is insufficient for VLP formation alone, however, and co-required components vary in different systems. SARS-CoV-2 VLP formation requires M co-expression with S or N[43,64].

M has further been implicated in modulating host antiviral innate immunity. M inhibits the innate immune response by interfering with MAVS-mediated signaling and interferon production[65,66]. In mouse models of infection, M expression results in lung epithelial cell apoptosis in vitro and in vivo and may contribute to lung injury and pulmonary edema found in severe disease[66].

Despite its essential role in viral assembly and implication in pathogenesis, the molecular determinants of M function remain largely unknown. MHV M was proposed to adopt long and compact structures that differentially facilitate membrane bending and recruitment of other structural proteins based on low resolution tomographic analysis[40]. Intriguingly, a structural and evolutionary relationship between SARS-CoV-2 M and the accessory viroporin ORF3a was reported[67,68] based on predicted homology to our experimental ORF3a structures[69]. The manner in which distinct functional roles for M and ORF3a can be achieved in the context of a shared architecture remains to be determined. Here, we report the cryo-EM structure of SARS-CoV-2 M in lipid nanodiscs and perform molecular dynamics simulations to provide insight into M structure, function, and dynamics.

2.3 Results

We determined the structure of SARS-CoV-2 M in lipid nanodiscs (Fig. 2.1). Full-length M was expressed in *Spodoptera frugiperda* (Sf9) cells with a cleavable C-terminal GFP tag. Gel filtration chromatography of protein extracted in DDM/CHS detergent shows M runs predominantly as a single species consistent with a 50 kDa homodimer. We do not observe evidence of specific higher order oligomerization at low concentrations by fluorescence size exclusion chromatography or at higher concentrations in large scale purifications (Supplemental Fig. 2.1). SARS-CoV-2 ORF3a, in contrast, assembles into stable homodimers and homotetramers under similar conditions[69].

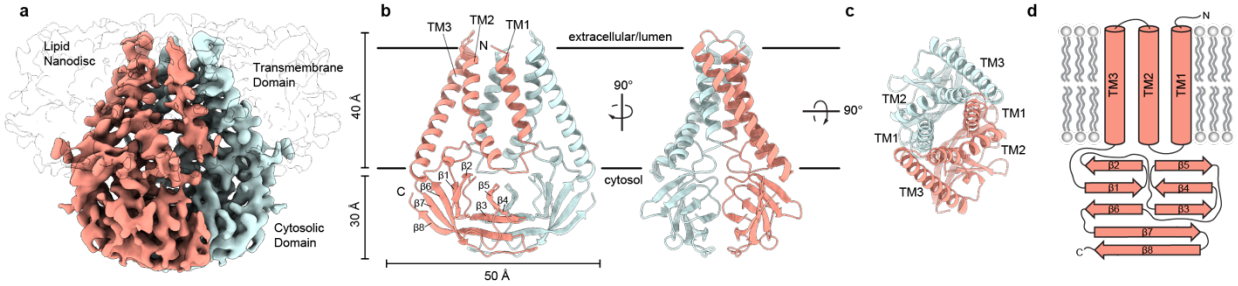


Figure 2.1 Structure of SARS-CoV-2 membrane (M) in lipid nanodiscs. (a) A 3.5 Å resolution cryo-EM map of SARS-CoV-2 M in MSP1E3D1 nanodiscs viewed from the membrane. One subunit is colored pink, and the second subunit is colored blue. Density corresponding to the lipid nanodisc is shown transparent. **(b,c)** Model of M viewed **(b)** from the membrane in two rotations and **(c)** from the extracellular or lumenal side. **(d)** Cartoon schematic of an M monomer with secondary structure elements indicated.

We reconstituted homodimeric SARS-CoV-2 M in nanodiscs made from the scaffold protein MSP1E3D1 and lipids (DOPE:POPC:POPS in a 2:1:1 ratio) and determined its structure by cryo-EM (Fig. 2.1, Supplement Fig. 2.2, Table 2.1). The majority of M (189 of 222 amino acids per subunit) was de novo modeled in the cryo-EM map (Fig. 2.1, Supplemental Fig. 2.2). The N-terminus (amino acids 1-16) and C-terminus (amino acids 205-222) are not resolved in the map and were not modeled. Loops connecting transmembrane helices (amino acids 36-42 and 71-78) are the least well resolved regions of the structure. The relatively weak density is consistent with a lack of stabilizing interactions between these and other M regions and likely indicates they adopt a range of conformations among particles used to generate the final map.

M is ~70 Å tall when viewed from the membrane with a ~40 Å transmembrane spanning region and ~30 Å cytosolic domain (CD) extending into the intracellular solution (Fig. 1). Each subunit contains an extracellular or lumenal N-terminus, three transmembrane helices (amino acids 17-36, 43-71, and 79-105) connected by short linkers, and a beta-strand rich C-terminal cytosolic domain. We note that AlphaFold and RoseTTAFold predicted M structures diverge substantially from the experimental structure (Supplemental Fig. 2.3)[70,71]. In the predicted structures, TM1s are swapped between subunits in addition to differences in the relative positions of transmembrane and cytosolic domains.

Viewed from above, TM1-3 from each subunit is positioned along a flattened ellipse with a long major (~50 Å) and short minor (~16 Å) axis. Within each subunit, TM2-TM3 are closely juxtaposed and tightly packed while TM1-TM2 are more distant and loosely connected. The two subunits assemble along their long axis, with TM1 from one protomer forming extensive interactions with the TM2-TM3 unit of the second protomer. As they project towards the cytoplasm, the three transmembrane helices twist counterclockwise and splay outwards in the inner leaflet, creating an expanded ellipse with ~55 Å and ~40 Å axes at the intracellular leaflet.

The transmembrane region is connected to the cytosolic domain through a tight turn-helix-turn segment comprised of residues 106-116. Within the cytosolic domain, 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, β6, the N-terminal half of β7, and the C-terminal half of β8. The inner sheet is formed by strands β3, β4, β5, the C-terminal half of β7, and the N-terminal half of β8. The inner sheets from each protomer interact through a large (~690 Å2 buried surface area per chain) and complementary interface with residues L138, V139, V143, L145, F193, A195 contributing to a hydrophobic core surrounded by additional polar interactions.

Using Dali[72] to compare the M structure to all experimentally determined protein structures returns SARS-CoV-2 ORF3a[69] as the only structural homolog with a shared fold. Superposition of the two viral proteins reveals a similar fold topology and homodimeric assembly with an overall RMSD of 4.4 Å. Isolated transmembrane and cytosolic domains from individual protomers are better superimposed (RMSD = 2.7 Å and 2.5 Å, respectively) (Fig. 2.2). Substantial differences in M and ORF3a structure are observed in three regions: TM2-TM3, the transmembrane-cytosolic domain junction, and the cytoplasmic domain interface. TM1s of M and ORF3a are well superimposed, but TM2-TM3 of M are splayed further out into the membrane and are less twisted about the two-fold symmetry axis to create a flatter and tighter interaction surface. The angle between TM3 and the cytosolic domain is ~25° more acute in M. The cytosolic domains of M are rotated ~15° away from the symmetry axis relative to ORF3a, shifting the cytosolic domain interface between subunits further from the membrane.

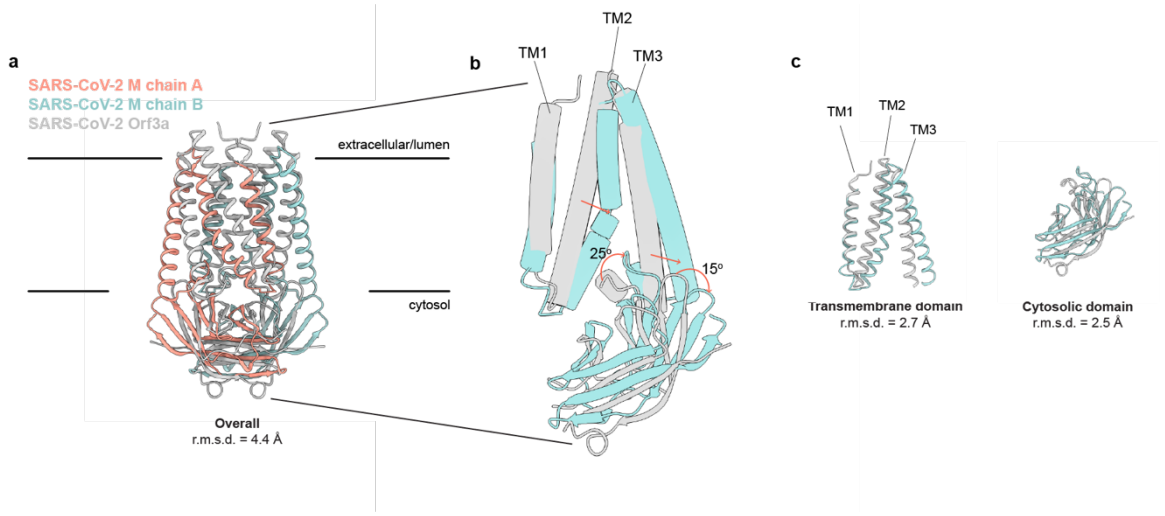


Figure 2.2 SARS-CoV-2 membrane (M) and ORF3a proteins are structurally homologous. (a) Overlay of M and ORF3a structures. M is colored with one subunit pink and the second subunit blue and ORF3a is white. **(b)** Overlay of a single subunit indicating major conformational rearrangements. **(c)** Overlay of isolated transmembrane and cytosolic domains from each protein.

What are the consequences of the structural rearrangement in M relative to ORF3a? Association of subunits in the M homodimer creates a polar and, presumably, water-filled pocket, reminiscent of the polar cavity created between subunits in ORF3a. However, the included volumes are different in several key respects (Fig. 2.3). First, the M pocket is ~⅓ smaller with an enclosed volume of ~840 Å3 compared to ~1300 Å3 in ORF3a. Second, the M pocket is completely sealed by protein to the surrounding membrane and cytoplasm; no openings large enough for water passage are observed connecting the pocket and protein exterior. In contrast, ORF3a displays three pairs of tunnels connecting its internal cavity to the membrane and cytoplasm (two are displayed in Fig. 2.3). Third, the position of the M pocket and ORF3a cavity are different. In M, the gap between subunits is confined to the region between cytosolic domains because transmembrane helices from opposing subunits form tight interactions across the entire lipid bilayer. In ORF3a, the gap extends from the region between cytosolic domains to approximately halfway across the membrane because transmembrane helices are less tightly associated across the membrane inner leaflet.

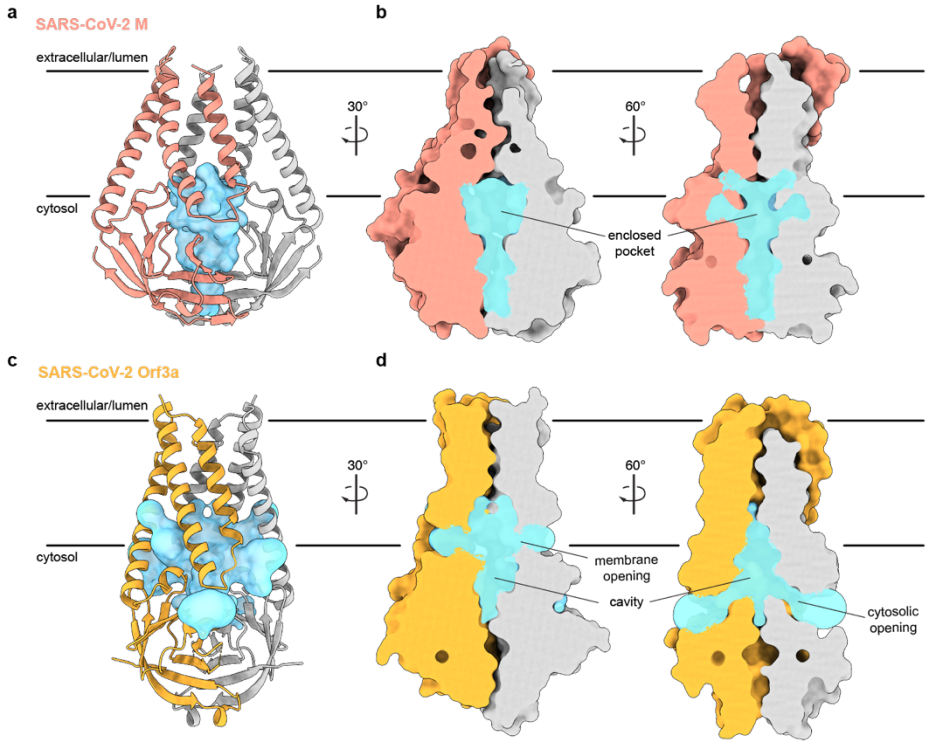


Figure 2.3 An enclosed polar pocket between cytosolic domains in membrane (M). (a) M shown as a cartoon and **(b)** surface with enclosed pocket volume calculated with CASTp36 shown as a blue surface. The enclosed pocket in M is formed between cytosolic domains and is sealed to the surrounding solution by protein. **(c,d)** Same as **(a,b)**, but for SARS-CoV-2 ORF3a. The cavity in ORF3a begins closer to the lipid bilayer, extends approximately halfway across the membrane, and is open to surrounding solution and lipids through multiple openings[69].

Another major difference in M and ORf3a structures is shown in Figure 2.4. The cytosolic domain of M is strikingly electropositive across nearly the entire exposed surface. Electropositive character is contributed by seventeen basic amino acids in three surface patches. The first covers the wide face of the cytosolic domains and consists of eight residues (R44, H125, R131, R146, H148, H155, R158, and R198). The second covers the narrow face of M and consists of four residues (R101, R105, R107, and R174). The third covers the underside of M and consists of five residues (K162, K166, K180, R186, and R200). ORF3a, in contrast, presents mixed electrostatic character with electropositive patches closer to the membrane and electronegative patches towards the cytoplasm. Such uniform electropositivity across the M cytosolic surface could facilitate the close juxtaposition of M present at high concentration in viral envelope with the negatively charged viral RNA genome.

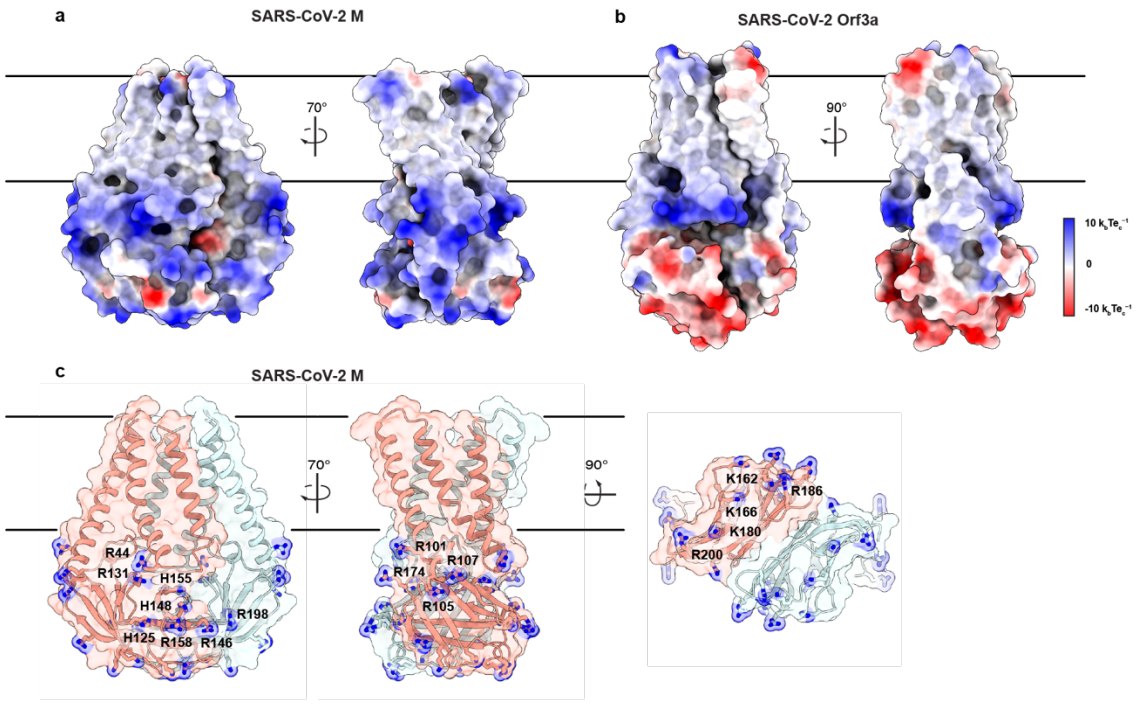


Figure 2.4 An electropositive cytosolic surface in membrane (M). (a,b) Views of the wide and narrow faces of **(a)** M and **(b)** ORF3a colored according to electrostatic surface potential from red (electronegative, –10 kbTec-1) to blue (electropositive, +10 kbTec-1). **(c)** Views of three electropositive surface patches on M cytosolic domains with basic residues labeled and shown as sticks with blue nitrogen atoms.

The large, complementary, and hydrophobic interface between transmembrane and cytosolic regions of subunits in the M structure suggests a structurally rigid core. However, a previous tomographic study of MHV suggesting that M adopts distinct long and compact structures[40], M's structural homology to the viroporin ORF3a[67,69], and the dissociation of cytosolic regions shown in predicted SARS-CoV-2 M structures[70] suggest

the possibility that M can undergo large-scale structural rearrangements. We next performed molecular dynamics (MD) simulations to explore potential conformational dynamics of M.

We equilibrated M in a lipid environment and ran an all-atom MD simulation for 2 μs. Major conformational changes related to protein function typically take place on a timescale ranging from a microsecond to seconds[73–75]. Prior work has reported microsecond simulations can capture large conformational changes of membrane proteins in a lipid bilayer[76–79]. We therefore expected to capture a subset of possible structural changes in our simulation that correspond to relatively frequently accessed states. Overall, we did not observe substantial conformational rearrangements in M during the simulation (Fig. 2.5A,B). Superposition of the experimental and final M structure following the simulation shows minor deviations through most of the protein (overall RMSD of 2.5 Å) (Fig. 2.5A). The largest difference is a shift in TM1 up towards the extracellular/lumenal side by approximately half a helical turn, enabled by rearrangement of the TM1-TM2 linker (Fig. 2.5A). This relatively subtle movement is consistent with weaker density for the TM1-TM2 linker in the cryo-EM map and fewer packing interactions for TM1 than TM2 or TM3. Per residue deviations ranged from ~ 1-4 Å and, aside from the movement of TM1, were similar between subunits and largest in the TM2-TM3 linker, transmembrane to cytosolic region connection, and loops connecting strands in the cytosolic domain. Minimal structural deviation was observed during the simulation within or between subunits as judged by the number of close Cα contacts, the angle between transmembrane and cytosolic regions, the distance between transmembrane regions, or the distance between cytosolic domains (Fig. 5D-H). Consistent with limited movement of the transmembrane region and a lack of evidence for lipid binding in the cryo-EM structure, no obvious enrichment of specific lipids around M was identified following simulation (Supplemental Fig. 2.5). Finally, the internal M pocket remained similar in size and sealed from the surrounding solution throughout the simulation (Fig. 2.5I,J). Taken together, these data suggest that M adopts a stable structure with minimal dynamic rearrangements on the microsecond time scale, at physiological temperature, and in the absence of additional binding proteins. These results are consistent with the absence of other highly populated conformations in our cryo-EM data, though we cannot exclude the possibility that M undergoes rare large-scale conformational changes under these conditions not captured in the simulation.

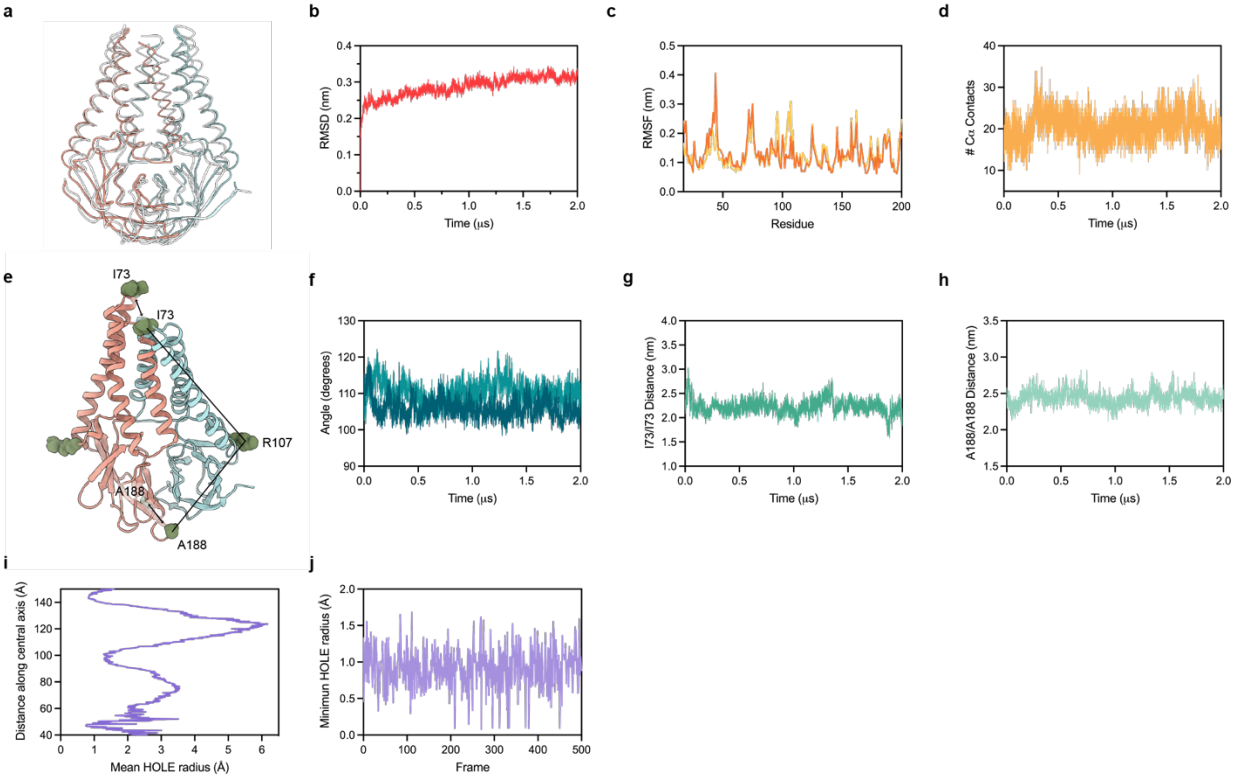


Figure 2.5 Molecular dynamics simulation of membrane (M). (a) Overlay of M cryo-EM structure (colored in pink and blue) and final structure (in white) following 2.0 μs all-atom molecular dynamics simulation. **(b)** Overall RMSD between simulated and initial structure during simulation. **(c)** Root mean square fluctuation of protein residues in the simulation. Orange and yellow colors correspond to individual M protein chains. **(d)** Number of C-alpha contacts between two monomers. **(e)** Structural representation of distances and angles used for calculations in **(f–h)**. **(f)** A188-R107-I73 angle plot for each monomer. One monomer has slightly higher values than the other. **(g)** Center of mass distance between I73 residues at the top of the TM2-TM3 linker. **(h)** Center of mass distance between A188 residues at the base of the cytosolic domains. **(i)** Mean radius of the enclosed pocket in M over the simulation trajectory versus distance along the symmetry axis. At its widest positions, the pocket is wide enough to accommodate two water molecules. **(j)** Minimum hole radius versus the frame number in the simulation. The lack of substantial changes in radius indicates a stable pocket size and shape that does not open to solution during the simulation.

2.4 Discussion

The structure of the SARS-CoV-2 M protein that we have obtained by cryo-EM reveals a homodimeric fold that is structurally homologous to the nonselective Ca2+ permeable cation channel of SARS-CoV-2, ORF3a. As with 3a, each subunit of M contains three transmembrane helices and a C-terminal beta sandwich domain. However, the structure differs from ORF3a in several key ways that provides insight into how these structurally similar proteins can fill drastically different apparent roles in the coronavirus life cycle.

When viewed from the plane of the membrane, M is considerably wider and flatter than ORF3a, due to differences in transmembrane helix packing and a rotation about the central axis of the cytosolic domain. Among the consequences of this flattening out of

M are distinct differences in the dimer interface across the membrane, where M shows a tighter dimer interface closer to the membrane outer leaflet as well as a gap between cytosolic domain subunits that forms an enclosed pocket lined by polar residues. In ORF3a, transmembrane regions are less closely opposed and a gap between subunits extends from halfway across the membrane to halfway down the cytosolic domains. The result is a larger cavity that is open to the membrane and cytoplasm. Mutations in the ORF3a cavity alter ion channel activity, consistent with the cavity forming part of the conduction path. Tight subunit association may therefore be important for the structural role of M, while loose subunit association that creates a large open cavity may be essential for the viroporin activity of ORF3a.

In further contrast to ORF3a, which was seen to form stable tetramers through electrostatic interactions between neighboring dimers, we see no evidence that M forms higher order oligomers under similar experimental conditions. Surface characteristics of the M dimer lend credence to the possibility that M exists solely as a dimer in the membrane–one striking feature of the M C-terminal beta sandwich domain is the presence of three sizable patches of positive charge that dominate its solvent exposed surface. Molecular dynamics simulations of M show that the dimer is stable and does not readily adopt alternate conformations at physiological temperature over the 1.6 μs trajectory. These data suggest M could play a structural role in mediating morphological changes in host cell membranes through interactions with other SARS-CoV-2 structural proteins and perhaps negatively charged lipid headgroups or viral RNA.

A complementary study published during review of this work includes two structures of SARS-CoV-2 M solubilized in detergent micelles and in complex with Fab antibody fragments bound to the cytosolic domain[80]. One structure (termed short form) is similar to that reported here. The second structure (termed long form) adopts a distinct conformation that is taller and narrower in the membrane. The conformations are suggestive of short and long forms of MHV M observed in low resolution tomograms[40]. Our data suggests M in a simple lipid mixture predominantly adopts the short form. It may be that the alternate M long form, stabilized by Fab binding for structure determination, is similarly promoted in a physiological context by interactions with lipids, Spike, N, or viral RNA. Evidence for possible higher order oligomerization was also reported in some 2D classes of M solubilized in a different detergent, suggesting that M-M interactions may occur under certain biochemical conditions. Whether distinct M conformations have distinct functional roles in virus assembly, as has been suggested for MHV M, remains to be determined.

M has been shown to play a crucial role in viral assembly through protein-protein interactions with other coronavirus structural proteins such as N and S. Spike proteins are incorporated into coronavirus virions via interactions between the cytosolic tail of S and the cytosolic domain of M, however the precise details of this interaction are unknown[37]. In SARS-CoV-2, M and N or S are the minimal components required for forming VLPs when expressed heterologously in cells[43,64]. Several recent studies have

suggested that the C-terminal domain of N is the site of interaction between SARS-CoV-2 M and N, but as with S a precise binding site has not been established[81,82]. It is possible that M-N interactions are mediated by favorable electrostatic interactions between negatively charged residues of the N CTD and one or more of the basic patches identified on the surface of the cytosolic domain of M, a model supported by co-immunoprecipitation of N and M proteins expressed in 293T cells[80]. Through the sheer abundance of M dimers found in the membrane of SARS-CoV-2 virions, M and N together might facilitate VLP formation via a mechanism similar to the Gag precursor of HIV, where the high concentration of M C-terminal domains at the cytoplasmic membrane surface recruit and organize many N proteins that together physically extrude a membranous bud.

At present the World Health Organization puts the confirmed number of COVID-19 cases worldwide at nearly 530 million. Over the last two years the SARS-CoV-2 virus has undergone many mutations that have been extensively documented through global sequencing efforts[83]. Despite this, the M protein sequence has remained virtually unchanged–a testament to the critical role that M plays in viral replication and assembly[84]. Furthermore, while only 20 amino acids in length, the N-terminus of M has been found to be highly immunogenic in COVID-19 patients[53–56]. M has also been shown to modulate innate immune response and could contribute to lung injury often seen in severe cases[65,66]. Given its clear importance in the coronavirus life cycle and pathogenicity, M presents an attractive target for therapeutics or vaccines. While M is well conserved across Coronaviridae (Supplemental Fig. 2.4), it shows particularly high conservation between SARS-CoV-1 and SARS-CoV-2, with a sequence similarity of 90.54%, highlighting its potential as a therapeutic target for emergent coronaviruses in the future.

2.5 Material and Methods

2.5.1 Cloning and protein expression

The coding sequence for SARS-Cov-2 M protein (Uniprot P0DTC5) was synthesized (IDT, Newark, NJ) and cloned into a vector based on the pACEBAC1 backbone (MultiBac; Geneva Biotech, Geneva, Switzerland) with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 7xHis tag, generating a construct for expression of M-SNS-LEVLFQGP-SRGGSGAAAGSGSGS-sfGFP-GSS-7xHis[85]. MultiBac cells were used to generate a Bacmid according to manufacturer's instructions. Sf9 cells were cultured in ESF 921 medium (Expression Systems, Davis, CA) and P1 virus was generated from cells transfected with Escort IV reagent (MillaporeSigma, Burlington, MA) according to manufacturer's instructions. P2 virus was then generated by infecting cells at 2 million cells/mL with P1 virus at a MOI ~0.1, with infection monitored by fluorescence and harvested at 72 hours. 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/mL at a MOI ~2–5. At 72 hours, infected

cells containing expressed M-sfGFP protein were harvested by centrifugation at 2500 x g for 10 minutes and frozen at -80°C.

2.5.2 Protein purification

Infected Sf9 cells from 1 L of culture (~15 mL of cell pellet) were thawed in 100 mL of Lysis Buffer containing 50 mM HEPES, 150 mM KCl, 1mM EDTA pH 8. Protease inhibitors (Final Concentrations: E64 (1 μM), pepstatin A (1 μg/mL), soy trypsin inhibitor (10 μg/mL), benzamidine (1 mM), aprotinin (1 μg/mL), leupeptin (1μg/mL), AEBSF (1mM), and PMSF (1mM)) 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 x g for 45 minutes. The supernatant was discarded, and residual nucleic acid was removed from the top of the membrane pellet using DPBS. Membrane pellets were scooped into a dounce homogenizer containing extraction buffer (50 mM HEPES, 150 mM KCl, 1 mM EDTA, 1% *n*-Dodecyl-β-D-Maltopyranoside (DDM, Anatrace, Maumee, OH), 0.2% cholesteryl hemisuccinate Tris salt (CHS, Anatrace, Maumee, OH) pH 8). A stock solution of 10% DDM, 2% CHS was dissolved and clarified by bath sonication in 200 mM HEPES pH 8 prior to addition to buffer to the indicated final concentration. Membrane pellets were then homogenized in extraction buffer and this mixture (150 mL final volume) was gently stirred at 4°C for 1.5 hours. The extraction mixture was centrifuged at 33,000 x g for 45 minutes and the supernatant, containing solubilized membrane protein, was bound to 4 mL of Sepharose resin coupled to anti-GFP nanobody for 1.5 hours 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, 0.005% CHS, pH 7.4), 40 mL of buffer 2 (20 mM HEPES, 500 mM KCl, 1 mM EDTA, 0.025% DDM, 0.005% CHS, 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 hours. Timing of this step was critical as longer incubations in detergent significantly reduced yield and sample quality. Cleaved M protein was then eluted with an additional 12 mL of wash buffer, spin concentrated to ~1 mL with Amicon Ultra spin concentrator 10 kDa cutoff (Millipore) and loaded onto a Superose 6 increase column (GE Healthcare, Chicago, IL) on an NGC system (Bio-Rad, Hercules, CA) equilibrated in buffer 1. Peak fractions containing M protein were then collected and spin concentrated prior to incorporation into nanodiscs.

2.5.3 Nanodisc formation

Freshly purified M protein in Buffer 1 was reconstituted into MSP1E3D1 nanodiscs with a mixture of lipids (DOPE:POPS:POPC at a 2:1:1 mass ratio, Avanti, Alabaster, Alabama) at a final molar ratio of 1:4:400 (M:MSP1E3D1:lipid).

Twenty mM solubilized lipid in lipid dilution buffer (20 mM HEPES, 150 mM KCl, pH 7.4) was mixed with additional DDM/CHS detergent and M protein at 4°C for 30 minutes before addition of purified MSP1E3D1. This addition brought the final concentrations to approximately 10 μM M protein, 40 μM MSP1E3D1, 4 mM lipid mix,10 mM DDM, and 1.7 mM CHS. The solution with MSP1E3D1 was mixed at 4°C for 15 minutes before addition

of 150 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 final mixture was then gently tumbled at 4°C overnight (~ 12 hours). Supernatant was cleared of beads by letting large beads settle and carefully removing liquid with a pipette. Sample was spun for 10 minutes at 21,000 x g before loading onto a Superose 6 increase column in 20 mM HEPES, 150 mM KCl, pH 7.4. Peak fractions corresponding to M protein in MSP1E3D1 were collected, 10 kDa cutoff spin concentrated and used for grid preparation. MSP1E3D1 was prepared as previously described[86] without cleavage of the His-tag.

2.5.4 Cryo-EM sample preparation and data collection

M in MSP1E3D1 was prepared at a final concentration of 1.3 mg/mL. Concentrated sample was cleared by a 10-minute 21,000 x g spin at 4ºC prior to grid preparation. 3.4 µl of protein was applied to freshly glow discharged Holey Carbon, 300 mesh R 1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany) and plunge frozen in liquid ethane using a FEI Vitrobot Mark IV (ThermoFisher Scientific) was used with 4°C, 100% humidity, 1 blot force, a wait time of ~5 seconds, and a 3 second blot time.

Grids were clipped and sent to Thermo Fisher Scientific RnD division in Eindhoven, The Netherlands for data collection. Grids were loaded onto a Krios G4 microscope equipped with a Cold Field Emission gun (CFEG) and operated at 300 kV. Data were collected on a Falcon 4 detector mounted behind a Selectris X energy filter. The slit width of the energy filter was set to 10eV. 7,588 movie stacks containing 1251 frames were collected with EER (electron event representation) mode[87] of Falcon 4 detector at a magnification of 165,000 corresponding to a pixel size of 0.727 Å. Each movie stack was recorded with a total dose of 50e-/Å2 on sample and a defocus range between 0.5 to 1.2 µm.

See Supplementary Table 2.1 for detailed data collection statistics.

2.5.5 Cryo-EM data processing

Motion correction and dose weighting were performed on all 7,588 videos using RELION 4.0's[88] implementation of MotionCor2[89] at 0.727 Å per pixel. Contrast transfer function (CTF) parameters were fit with CTFFIND-4.1[90]. Template-free auto-picking of particles was performed with RELION 4.0's Laplacian-of-Gaussian filter on movies CTF fit to 5.0 Å or better, yielding an initial set of 2,379,507 particles. These particles were then extracted at a 288-pixel box size and transferred to cryoSPARC v.3.2[91] for two-dimensional classification.

Iterative rounds of 2D classification resulted in a set of 31,126 particles which were then extracted in RELION 4.0, and their coordinates were used in the Topaz particle-picking pipeline[92]. Topaz training, picking, and extraction yielded 2,376,190 particles which were then subjected to one round of 2D classification in RELION 4.0 to remove obvious noise. The resulting 2,007,561 particles were extracted and then iteratively 2D classified in cryoSPARC v3.2, resulting in a set of 54,747 'good' particles.

Both the initial auto-picked particle set and subsequent Topaz particle set were lacking in good 2D classes of side views, so a subset of 13,698 particles from the best side view classes were extracted in RELION 4.0 and used to train and pick new particles in Topaz. As before, the resulting 2,186,648 particles were subjected to one round of 2D classification in RELION 4.0 then imported into cryoSPARC v3.2 for further 2D classification.

Good particles from the initial auto-picked particle set and both Topaz particle sets were pooled and duplicates within 100 Å were removed to yield 105,535 particles. These particles were extracted and imported into cryoSPARC v3.2 for 3 rounds of 2D classification to remove remaining junk. An ab initio reconstruction of the remaining 69,182 particles was performed to provide an initial volume and a subsequent non-uniform refinement (C2, 2 extra passes, 16 Å initial resolution) produced a map with a 4.0 Å overall resolution. This map was post-processed in RELION 4.0 and used for Bayesian particle polishing.

The resulting 'shiny' particles were imported back into cryoSPARC v3.2 for one additional round of 2D classification. The final 64,966 particles were used to generate a new ab initio and a subsequent non-uniform refinement (C2, 2 extra passes, 16 Å initial resolution, 1.5 adaptive window factor) yielded the final map at 3.5 Å nominal resolution.

2.5.6 Modeling, refinement, and analysis

cryoSPARC sharpened cryo-EM maps were used to de novo model M using Coot[93]. The model was real space refined in Phenix[94] and validated using Molprobity[95]. Cavity and tunnel measurements were made with CASTp[96]. Comparisons to the structure database were performed with DALI[72]. Figures were prepared using ChimeraX[97], Prism, Adobe Photoshop, and Adobe Illustrator software.

2.5.7 Fluorescence size exclusion chromatography (FSEC)

Sf9 cells (~4 million) from the third day of infection were pelleted, frozen, and then thawed into extraction buffer (20mM 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 hour and lysate was then pelleted at 21,000 x g at 4 °C for 1 hour to clear the 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, 0.025% DDM, 0.005% CHS.

2.5.8 Molecular dynamics

The initial MD system of M-protein and lipid bilayer was built using CHARMM-GUI Membrane Builder[98–101]. A 20x20 nm lipid bilayer membrane was taken with a mixture of DOPE, POPS and POPC lipids in a 2:1:1 mass ratio. A fully hydrated bilayer was built around the M protein, centering the transmembrane region close to the lipid bilayer center. To neutralize the system, a 0.15M KCl salt concentration was used. The simulations were performed on GROMACS MD simulation package[102] with the

CHARMM36m force-field[103]. An initial minimization of the system was carried out following six-steps protocols provided on CHARMM-GUI[104]. A time step of 2fs was used with periodic boundary conditions for the simulations. A simulation temperature of 310.15 K was maintained with a Nose-Hoover thermostat[105,106] and a coupling time constant of 1.0 ps in GROMACS. The pressure was set at 1bar with a Berendsen barostat[107] during initial relaxation. For the production runs, the Parrinello-Rahman barostat was used semi-isotropically with the compressibility of 4.5×10^{-5} and a coupling time constant of 5.0 ps[108,109]. For the non-bonded interactions, a switching function between 1.0 and 1.2 nm was used. The long-range electrostatics were computed using Particle Mesh Ewald (PME)[110]. The LINCS algorithm was used to constrain hydrogen bonds[111]. We performed 1.6 μs production run for the system and used Frontera (TACC), and Midway2 (Research Computing Center at the University of Chicago) to run these simulations.

The RMSD of the protein and RMSF per residue (Fig. 5B, C) were calculated using the GROMACS module. The center-of-mass (COM) distances between two residues (Fig. 5G, H), number of Cα contacts between two monomers (Fig. 5D), and angles between transmembrane and cytosolic regions (Fig. 5F) were also calculated using the GROMACS package[102]. The analysis of the M pocket was performed using the HOLE program[112] implemented in MDAnalysis[113] (Fig. 5I, J). In Fig. 5J, each frame was taken at a 4ns time step. The lipid distribution around the M protein was calculated using the MDAnalysis Python packages[113] (Fig. S4). Visual Molecular Dynamics (VMD)[114]and PyMOL were used as visualization software.

2.6 Data and reagent availability

All data and reagents associated with this study are publicly available. The final model is in the PDB under 8CTK, the final map is in the EMDB under EMD-26993, and micrographs (original and motion corrected) and final particle stack are deposited in EMPIAR.

2.7 Acknowledgements

We thank Dan Toso, Jonathan Remis, and Paul Tobias for support collecting preliminary data at the Cal-Cryo EM facility. We thank Thermo Fisher Scientific for microscope access. We thank Diana Bautista and members in Brohawn and Bautista labs for feedback on the project. We thank Savitha Sridharan and Hillel Adesnik for the initial M DNA. SGB is a New York Stem Cell Foundation-Robertson Neuroscience Investigator. This work was funded in part by the New York Stem Cell Foundation, a Sloan Research Fellowship (to SGB), a Fast Grants Award from Emergent Ventures at the Mercatus Center, George Mason University (to SGB, Diana Bautista, and Hillel Adesnik at UC Berkeley), an NSF Graduate Research Fellowship (to KD), and an NSF RAPID grant CHE-2029092 (MD and GAV). Computer simulations were carried out on

the Frontera supercomputer at the Texas Advanced Computer Center (TACC) as funded by the National Science Foundation (OAC-1818253), as well as on the Midway2 cluster at the Research Computing Center (RCC) of the University of Chicago.

2.8 Contributions

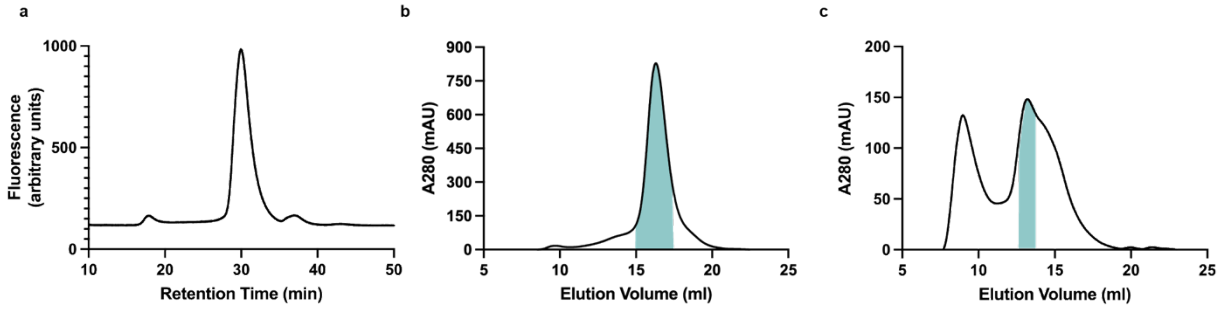
K.A.D., D.M.K., and S.G.B. conceived of the project. D.M.K. generated initial constructs and performed expression testing. K.A.D. generated final constructs and performed expression, purification, reconstitution, cryo-EM sample preparation, and cryo-EM data processing. A.K. collected cryo-EM data at Thermo Fisher Scientific. K.D. and S.G.B. built and refined the atomic model. M.D. and G.A.V. designed, performed, and analyzed molecular dynamics simulations. G.A.V. and S.G.B. secured funding and supervised research. K.A.D. and S.G.B. wrote the manuscript with input from all authors.

2.9 Supplemental tables and figures

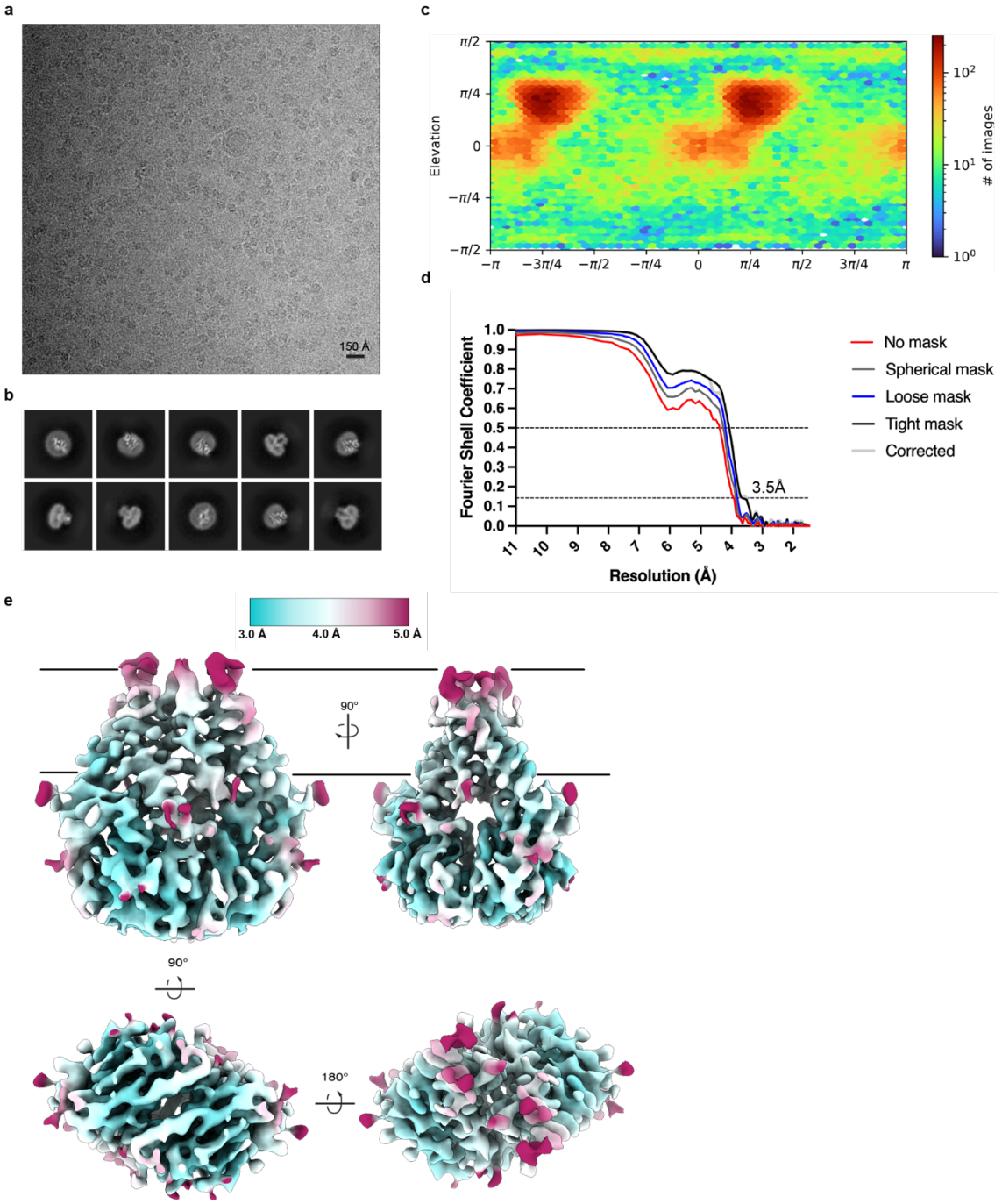
Data collection	**SARS-CoV-2 M**
PDB	8CTK
EMDB	26993
EMPIAR	XXXXX
Total movies	7588
Magnification	165,000 x
Voltage (KV)	300
Electron exposure (e^-/Å^2)	50
Defocus range (um)	-0.5 to -1.2
Super resolution pixel size (Å^2)	0.3635
Binned pixel size (Å^2)	0.727
Processing	
Initial particle images (no.)	2,007,561
Final particle images (no.)	64,966
Map resolution Masked (Å, FSC = 0.143)	3.52
Symmetry imposed	C2
Refinement	
Model resolution (Å, FSC = 0.143 / FSC = 0.5)	3.2 /3.5

Map-sharpening B factor (Å^2)	-145
Composition	
Number of atoms	3042
Number of protein residues	376
Number of ligands	0
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (Å)	0.690
Validation	
MolProbity score	1.47
Clashscore	4.37
Ramachandran plot	
Favored (%)	96.24
Allowed (%)	3.76
Disallowed (%)	0
Rotamer outliers (%)	0
Mean B factor (Å^2)	
Protein	42.42

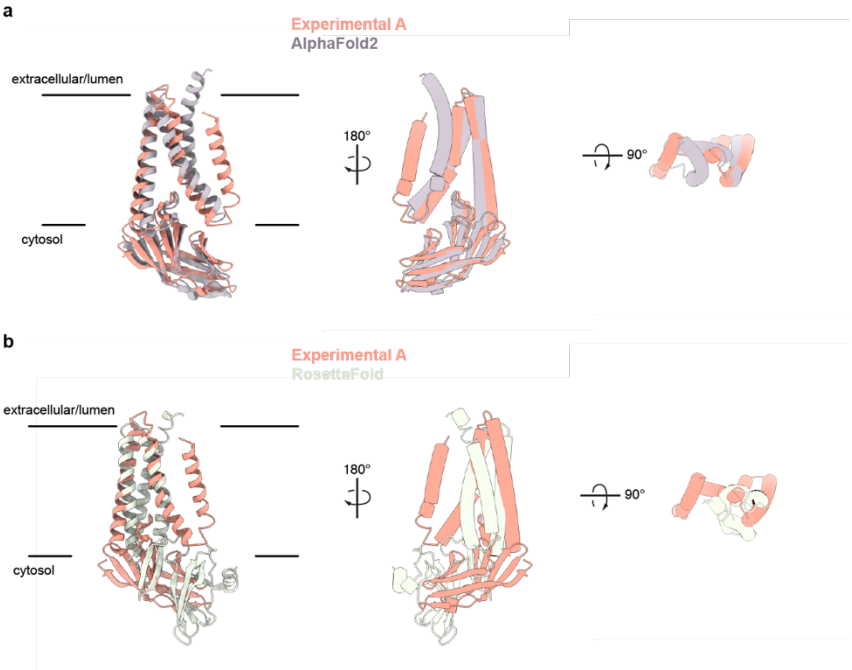
Supplemental Table 2.1 Cryo-EM data collection, refinement, and validation statistics.



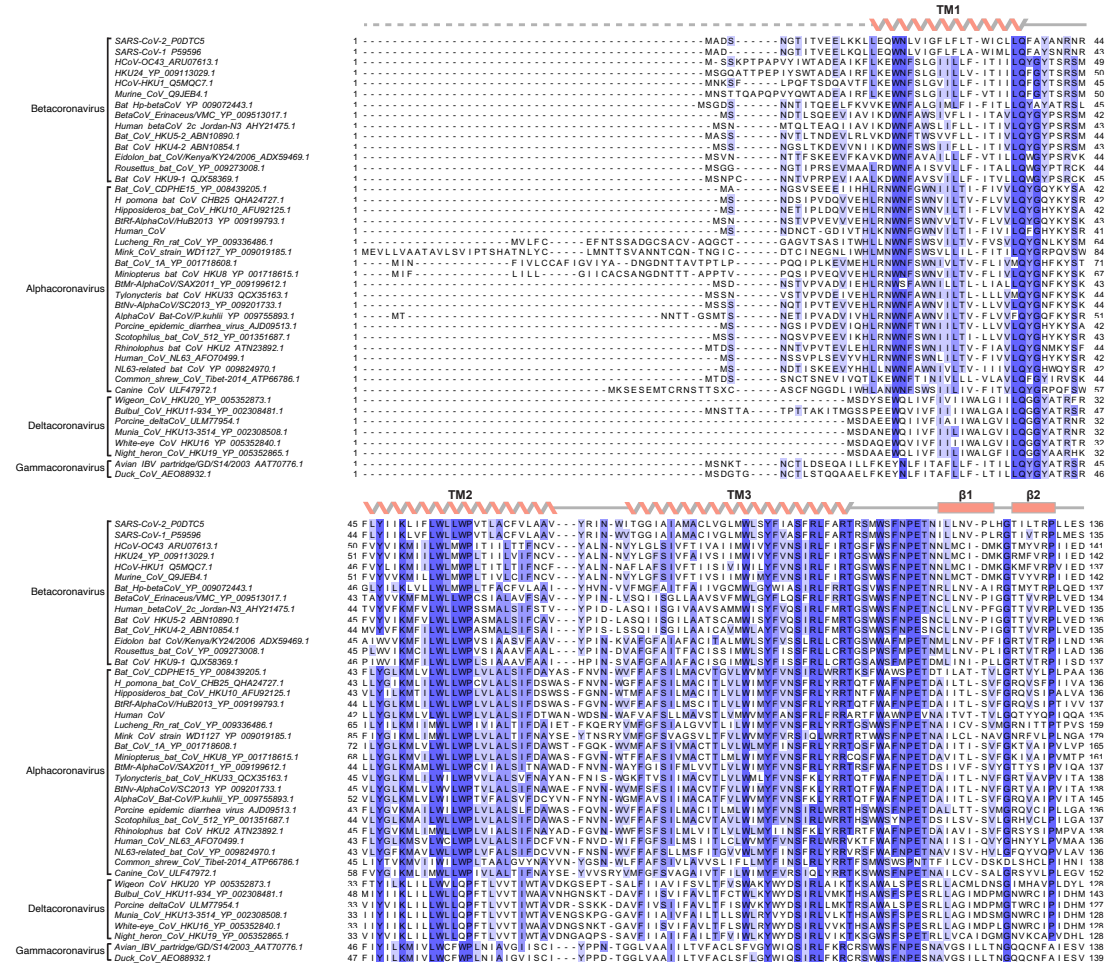
Supplemental Figure 2.1 Purification and reconstitution of membrane (M). (a) Fluorescence size exclusion chromatogram of M expressed in insect cells and extracted in DDM/CHS. **(b)** Size exclusion chromatogram of M expressed in insect cells and extracted and purified in DDM/CHS. **(c)** Size exclusion chromatogram of M reconstituted into MSP1E3D1 lipid nanodiscs. Samples were run on Superose 6 columns. Blue bars indicate pooled fractions.



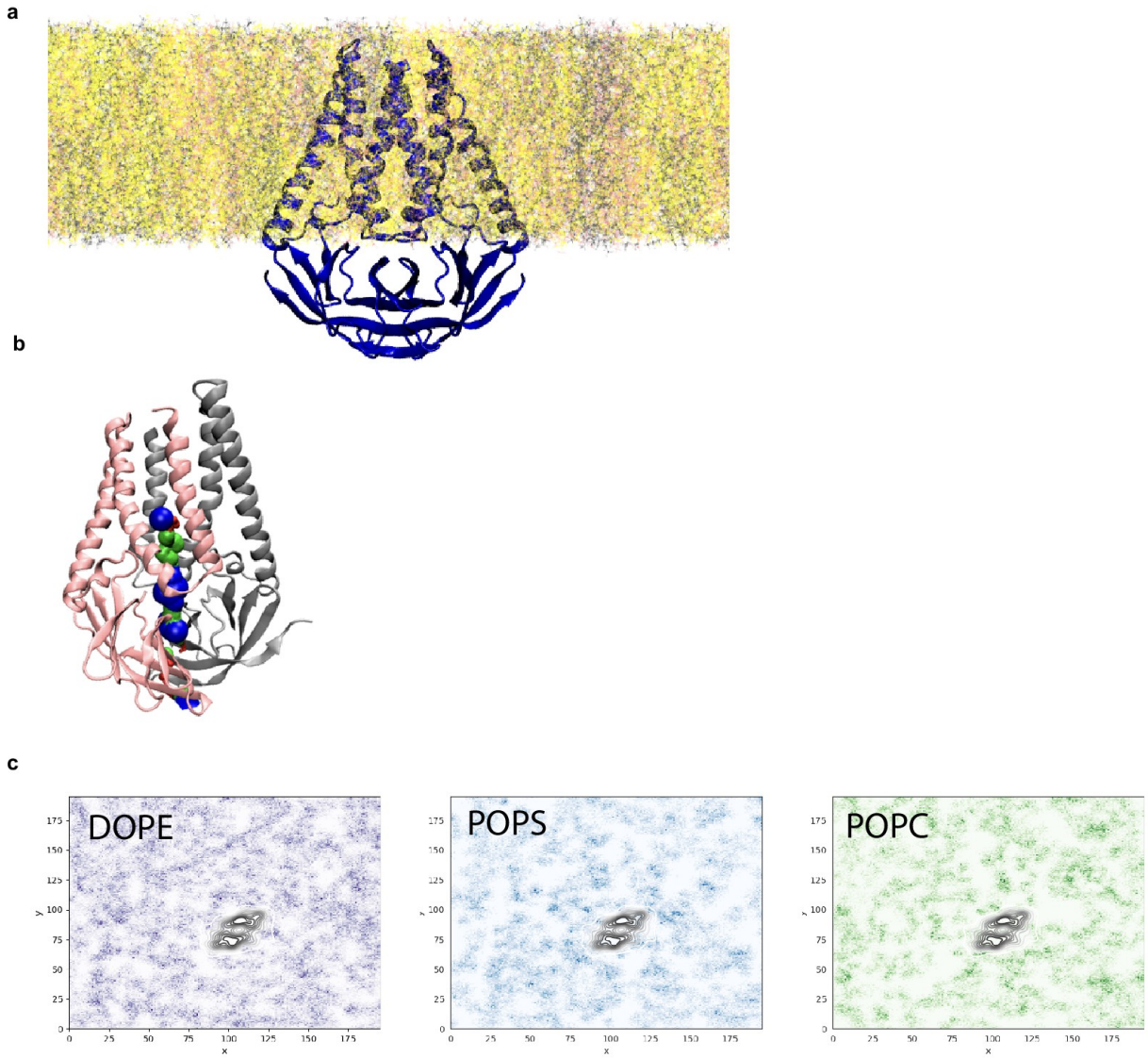
Supplemental Figure 2.2 Cryo-EM processing and validation. (a) Representative micrograph, **(b)** selected 2D class averages, **(c)** angular distribution of particles used in final refinement, **(d)** Fourier shell correlation (FSC) relationships, and **(e)** local resolution estimated in cryoSPARC colored as indicated on the final map. Side, top, and bottom views are shown.



Supplemental Figure 2.3 Comparison of experimentally determined and predicted membrane (M) protein structures. (a) Overlay of the experimental and AlphaFold2 predicted M structures[71]. **(b)** Overlay of experimental and one RoseTTAfold predicted M structure[70]. Major differences are observed in chain topology in the transmembrane region and relative orientation of transmembrane and cytosolic domains.



Supplemental Figure 2.4 Sequence alignment of membrane (M) proteins across Coronaviridae. (a) Alignment of 42 M protein sequences colored by conservation in a ramp from white (not conserved) to dark blue (highly conserved). Accession numbers are indicated. Sequences were selected from representative species from each Coronavirus subgenus. Secondary structure from SARS-CoV-2 M is drawn above the alignment.



Supplemental Figure 2.5 Analysis of pocket size and lipid interactions from molecular dynamics simulation. (a) Equilibrated membrane (M) structure in lipid membrane consisting of DOPE:POPS:POPC in a 2:1:1 mass ratio. **(b)** Final M structure of the simulation shown with the pocket colored red where the radius is too small for a water molecule, green where it can accommodate a single water molecule, and blue where it can accommodate two water molecules. **(c)** Lipid sorting patterns. Darker regions indicate higher density of the lipids. Black color is the average location of the M dimer. No obvious pattern of enrichment is observed.

Chapter Three

Structural insight into a neutralizing antibody that targets the SARS-CoV-2 M protein

This chapter presents work done in collaboration with Professor Dan Watterson and Dr. Naphak Modhiran at the University of Queensland in Brisbane, Australia. Expression, purification, and crystallographic characterization of the SARS37 neutralizing antibody were performed by Dan and Naphak and are presented here with their permission. I performed the experiments described in this chapter, interpreted the results, and wrote the chapter with input and feedback from Steve Brohawn.

3.1 Introduction

The global COVID-19 pandemic caused by the coronavirus SARS-CoV-2 was first identified in Wuhan China in December 2019 and has since caused hundreds of millions of cases and at least 7 million deaths globally[115]. COVID-19 is a highly contagious respiratory disease that is mild to moderate for most but can lead to critical illness, particularly for elderly or immunocompromised patients. Despite extensive infection, vaccine, or hybrid-induced immunity across the globe, COVID-19 poses a continuing and highly disruptive burden on global health care, classroom, and nursing home operations[116].

Therapeutic antibodies have been developed to combat severe disease in COVID-19 patients, but the evolution of variants has hampered their long-term use. Even though vaccines and monoclonal antibodies have been deployed at an unprecedented pace, they still suffer from a loss of efficacy against emerging variants[117]. A common strategy to neutralize pathogens with therapeutic antibodies is to target viral fusion proteins to block entry. Hundreds of monoclonal antibodies (mAbs) have thus been developed to target the SARS-CoV-2 spike protein, with dozens in clinical trials and a few granted emergency use authorization[118–121]. The majority of the efforts for development of anti-spike antibodies and vaccines has focused on the spike protein's receptor binding domain (RBD)[122,123]. Over time however, it has proven to be a less than ideal strategy from an evolutionary standpoint because of the surprising plasticity of this essential domain[124]. To date, most existing and approved anti-spike mAbs show weakened defense against Omicron subvariants and development of new candidates have so far failed to keep pace with the rate of mutation seen in the virus[125]. Thus, there is a critical need to develop neutralizing antibodies against less mutable targets that will cover current and emergent variants.

Of the remaining three structural proteins encoded in the viral genome, the essential membrane (M) protein is arguably the most intriguing target for developing therapeutics. M is a 50 kDa homodimer consisting of three transmembrane helices per monomer, and a C-terminal domain with a unique β-sandwich fold. M has shown to

adopt two distinct conformations – a long and short form related by a hinge-like movement of the CTDs and coupled rotation and extension of the TMs[40,80,126]. M is the most abundant structural membrane protein and is known to direct viral particle assembly through direct protein-protein interactions with other coronavirus structural proteins and is required for budding virus-like particles[43]. Importantly, the M protein is highly conserved, sharing over 90% sequence similarity with the M protein from SARS-CoV-1. In the context of an infectious viral particle, only a short, disordered 19 amino acid N-terminal peptide and a small loop between TM2-and TM3 are displayed on the virion surface. Yet despite its compact surface area, the N-terminus of M is highly immunogenic, eliciting a specific and durable response from convalescent patients[54,127].

Our collaborators at the University of Queensland, Dr. Daniel Watterson and Dr. Naphak Modhiran, recombinantly expressed a neutralizing antibody obtained from a published sequence derived from a convalescent SARS-CoV-1 patient. Initially, they attempted to determine the crystal structure of the fragment antibody (Fab) in complex with the immunogenic N-terminal peptide of the M protein. While the majority of the Fab could be modeled from their data, they observed no density for either the unusually long, 24-amino acid long CDR3 loop on the Fab heavy chain or the M peptide (Fig. 3.1). This suggested to them that the Fab is not just binding to the N-terminus and that the context of the full-length and folded protein may be required to understand the interaction.

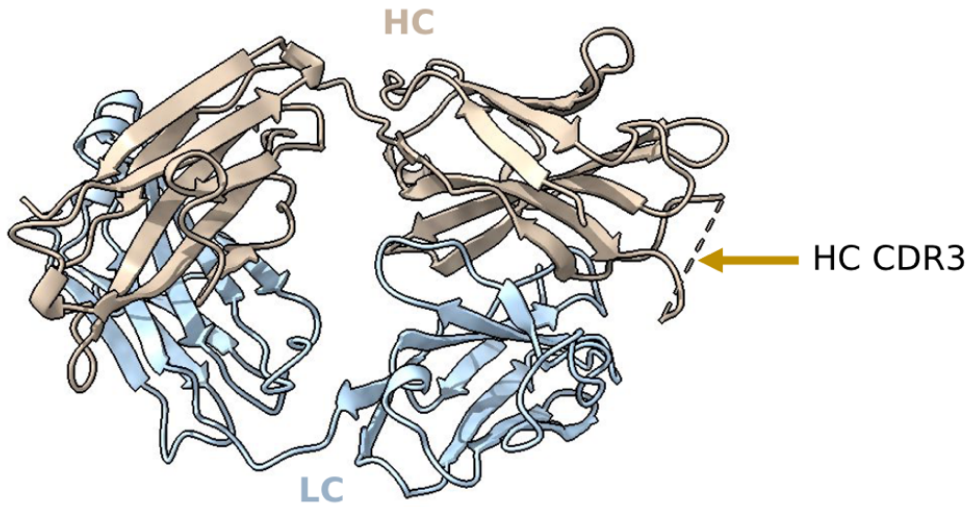


Figure 3.1 The M protein N-terminal peptide does not co-crystalize with the SARS37 Fab. (a) Crystal structure of the SARS37 Fab obtained in the presence of a peptide from the N-terminus of the M protein. The heavy chain (HC) colored in gold and light chain (LC) in blue. Missing density for the HC CDR3 loop is indicated.

Here we present our findings for the structural basis for neutralizing antibody recognition for the membrane protein of sarbecoviruses. We determined the structure of M in complex with the SARS37 neutralizing fragment antibody derived from SARS-CoV-

1 convalescent patient sera using single particle cryo-EM. We find that the Fab specifically recognizes the long conformation with high affinity. The M protein residues contacted by the Fab binding are conserved, highlighting its potential for long-term therapeutic use.

3.2 Results

Coronavirus M proteins have been shown to adopt distinct long and short conformations[40]. Previously reported cryo-EM structures of SARS-CoV-2 M were able to capture the long and short conformation using distinct Fabs that bind to M's cytosolic C-terminal domain, which we refer to as long conformation-selective Fab (lcFab) or short conformation-selective Fab (scFab)[80]. We used biolayer interferometry (BLI) to characterize the interaction between SARS37 and SARS-CoV-2 M and its dependence on M conformation (Fig. 3.2a,b). We purified M in complex with His-tagged lcFab or scFab, immobilized the complexes on Ni-NTA sensors, and probed their interaction with SARS37 Fab at different concentrations. SARS37 interacts strongly with M with over 40-fold higher affinity to the long versus short conformation (Fig. 3.1a and b, $K_{D,long}$ = 9.1 ± 1.8 nM, $K_{D,short}$ = 383 ± 51 nM). The affinity difference is due to an approximately 30-fold faster on rate and 10.2 fold slower off rate for the SARS37 Fab-long conformation interaction ($K_{on,long}$ = 4.5 ± 0.6 x 10^5 $M^{-1}s^{-1}$, $K_{off,long}$ = 4.1 ± 0.6 x 10^{-3} s^{-1}, $K_{on,short}$ = 1.4 ± 0.1 x 10^4 M-1s^{-1}, $K_{off,short}$ = 5.2 ± 0.5 x 10^{-3} s^{-1}). Whether long and short conformation M have distinct roles or are present in distinct locations within infected cells or on virions is unknown, but these data suggest that long conformation M is presented to the immune system during infection.

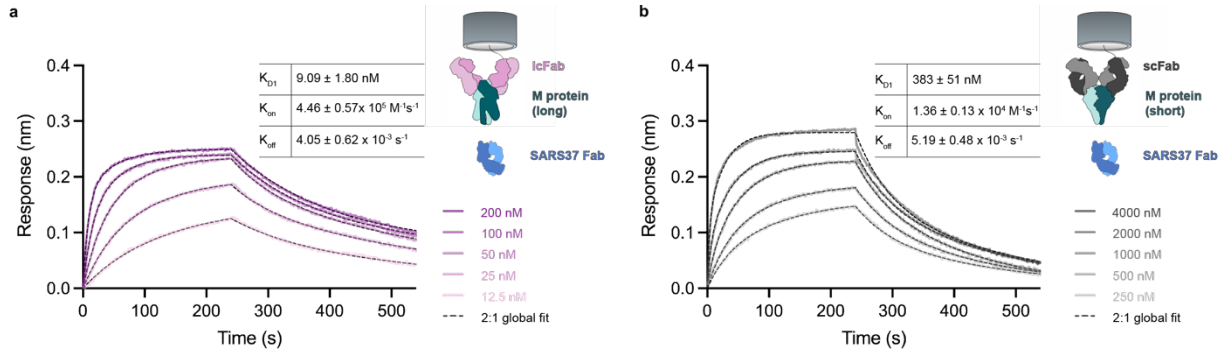


Figure 3.2 SARS37 antibody binds SARS-CoV-2 M with higher affinity to the long conformation. (a,b) Biolayer interferometry characterization of a SARS37 Fab interaction with SARS-CoV-2 M in the **(a)** long and **(b)** short conformation. Representative sensorgrams show binding response of a range of SARS37 Fab concentrations to immobilized M in complex with **(a)** long conformation-selective Fab (lcFab, pink) or **(b)** short conformation-selective Fab (scFab, gray). Raw data (shades of **(a)** pink or **(b)** gray) is overlaid with global fit to a 2:1 (SARS37 Fab:M homodimer) model (dashed lines). Dissociation constant, on rate, and off rates shown are mean ± SEM from n=4 and n=3 experiments for **(a)** and **(b)**, respectively.

Initial attempts to determine the molecular basis for SARS37 recognition of purified SARS-CoV-2 M in lipid nanodiscs using cryo-EM were hindered by significant flexibility of the complex (data not shown). We next sought to determine the structure of the SARS37 Fab bound to intact M stabilized in the long and short conformations by lcFabs and scFabs, respectively (Fig. 3.2a, Supplemental Fig. 3.1, Supplemental Fig. 3.2). The dataset collected of the SARS37 Fab:M:scFab complex showed a mixture of states with only a minority of particles appearing to have all three Fabs bound, and we were ultimately unable to resolve the SARS37 Fab binding interface for M in the short conformation (Supplemental Fig. 3.2).

Consistent with our BLI data suggesting that the SARS37 Fab preferentially recognizes M in the long conformation, our lcFab dataset was considerably more homogenous. A resulting consensus map of the complex with a 3.41 Å overall resolution showed a single SARS37 Fab bound to the extracellular surface of M in the long conformation (Fig. 3.2a). Further classification revealed subtle heterogeneity in the relative orientation of SARS37 and M related by a rocking motion. Separate refinement of two classes, both to 3.4 Å resolution, improved features of the SARS37-M interaction site for modeling (Supplemental Fig. 3.1). No substantial differences in the interaction between M and SARS37 are observed between the structures, so we focus discussion on the better resolved class 1 here.

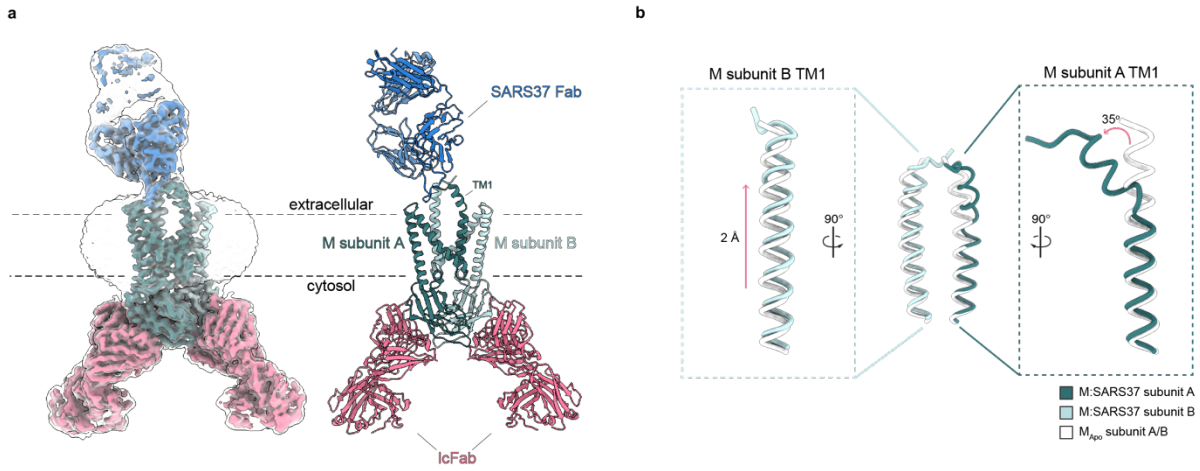


Figure 3.3 Structure of the SARS37:M:lcFab complex. (a) 3.5 Å resolution cryo-EM map of M in complex with SARS37 and lcFabs (left) and model (right) viewed from the membrane plane. Unsharpened (transparent white) is overlaid with DeepEMhancer post-processed map (colored by protein) used for modeling. Model of M-SARS37 Fab-lcFab complex colored with M subunit A dark teal, M subunit B light teal, SARS37 Fab blue, and lcFabs pink. **(b)** Close up views of distortions in TM1 helices from SARS37 bound M (dark and light teal) compared to apo M (PDB: 7VGR, in white)[80].

SARS37 Fab is perched on the extracellular surface of M, with its long CDRH3 contacting N-terminal regions at the top of TM1s and extending towards the membrane

to interact with the TM2-TM3 linker of one subunit (Fig. 3.2a). In the apo M long conformation structure the N-termini of each monomer extend from TM1 and project parallel to one another orthogonal to the membrane plane. In the SARS37 bound structure the N-termini are bent to project parallel to the membrane plane, creating a groove for the CDRH3 loop to insert into (Fig. 3.2b).

TM1 of subunit A is significantly more distorted, with the helix unwound from residues 10-20 and bending 35° towards the membrane. A short β-strand like structure forms between amino acids Thr7, Ile8, and Thr9 of the M subunit A N-terminus and Val113, Ile114, and Asp115 of SARS37 (Fig. 3.3b). The remaining interactions between SARS37 and subunit A are hydrophobic and driven by shape complementarity provided by the space between TM1 and TM2 that allows for the CDR3 loop to insert into the membrane (Fig. 3.3a-b).

Slightly more than half of the interface is formed between SARS37 and subunit B (457.7 $Å^2$ of the 806.3 $Å^2$ total buried surface area). The TM1 helix of M subunit B is moderately distorted compared to apo M structures, stretching to enable interactions critical for SARS37 binding specificity (Fig. 3.4c). Key contact residues on M include Val10, Glu11, and Glu12, which form extensive interactions with nearby amino acids of SARS37 CDRH3. Glu11 forms an electrostatic interaction with Arg118 of CDRH3 (Fig. 3.4a-c). The amine nitrogen of M residue Val10 is positioned to hydrogen bond with the backbone carbonyl oxygen of CDRH3 Pro117 while its side chain forms hydrophobic interactions with Ile114 and Ile107. Tyr119 of CDRH3 is positioned to form additional hydrogen bonds with both Glu11 and Glu12 of M.

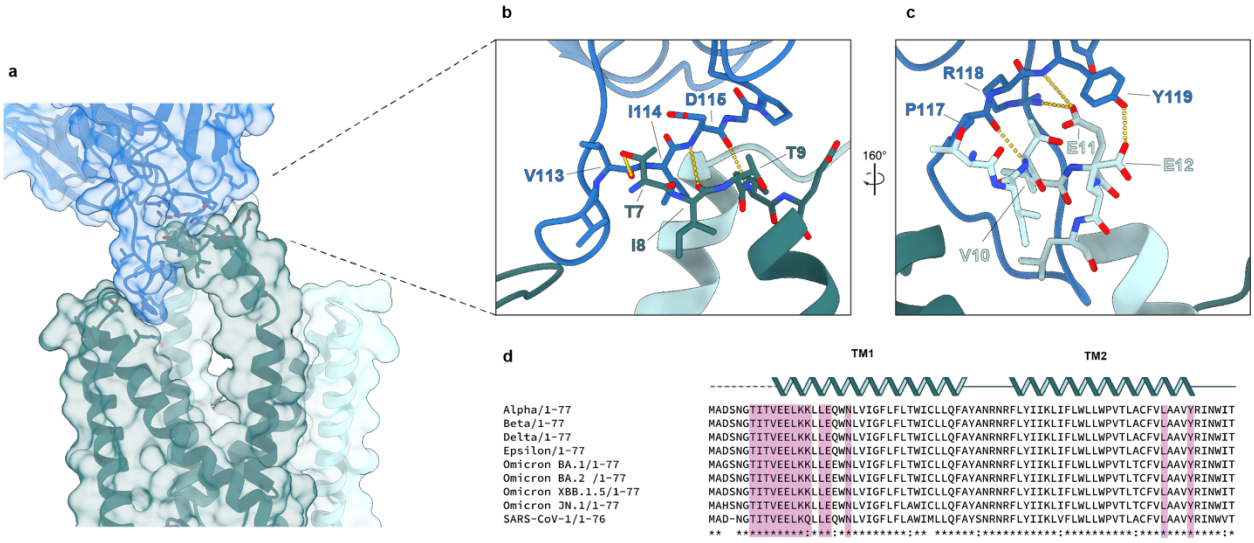


Figure 3.4 The SARS37 binding interface is highly conserved. (a) Close up views of the SARS37 Fab-M binding interface with surfaces transparent and interacting residues shown as sticks. **(b,c)** Apparent hydrogen bonding interaction (dashed yellow lines) between SARS37 Fab and M N-termini. **(d)** Multiple sequence alignment of M protein from SARS-CoV-2 variants and other betacoronaviruses. Secondary structure above is shown above, with residues in the SARS37 Fab binding epitope highlighted, and degree of conservation below (conserved residues are marked with an asterisk and conservative substitutions are marked with a colon).

The extracellular M epitope, consisting of the N-terminal segment, top of TM1, and TM2-TM3 linker, has remained remarkably conserved throughout the COVID-19 pandemic and maintains over 90% sequence identity with the M protein from SARS-CoV-1. Residues involved in the SARS37 binding interaction have been particularly conserved across SARS-CoV-2 variants – remaining completely unchanged to present (Fig. 3.4d).

3.3 Discussion

The development of new therapeutic antibodies to treat coronavirus infections is a critical component of the global effort to combat an ongoing SARS-CoV-2 pandemic. Crucially, the most effective of these neutralizing antibodies will target mutation resistant sites to avoid immune escape as the virus evolves. The highly conserved and immunogenic coronavirus M protein is known to play an essential role in viral particle assembly and egress but has never been utilized as an antibody target for active infections. Our collaborators Drs. Naphak Modhiran and Daniel Watterson, identified a potently neutralizing antibody from convalescent patient sera that targets the M protein, and we sought to structurally characterize the interaction.

We first measured binding affinity of a neutralizing antibody for the SARS-CoV-2 M protein stabilized in either one of two distinct conformations M is known to adopt using biolayer-interferometry (BLI). We determined that the SARS37 neutralizing Fab binds to M in the long conformation with over 40 times greater affinity than for the short conformation (9 nM and 383 nM, respectively). This discrepancy between the two forms is primarily a product of faster association for binding to long form M. This measured preference for long form M provides concrete evidence that this conformation is present in infectious viral particles circulating during infection, and that it can be effectively targeted therapeutically.

We then used single-particle cryo-EM to investigate the molecular basis of neutralizing antibody binding to the SARS-CoV-2 M protein. These efforts produced a 3.4 Å resolution map of the SARS37 Fab in complex with M stabilized in the long conformation from which we were able to model the entirety of the SARS37-M interaction. We found that SARS37 Fab binds on the extracellular surface of M with its long CDRH3 loop inserted into the membrane in a cleft formed between TM1 helices from both M subunits and TM2 of one protomer. The specificity of the binding interaction is provided by short β-sheet-like structures formed between the CDRH3 loop and residues on the N-termini of both subunits.

The work presented here provides insight into a novel mode of action for a SARS-CoV-2 neutralizing antibody and potential new avenue for treating infections. Importantly, the M residues of the binding interface have remained completely

unchanged throughout the pandemic, demonstrating the great potential of this antibody in our arsenal against an ever-evolving virus.

3.4 Material and Methods

3.4.1 Cloning, expression and purification of SARS-CoV-2 M protein

The coding sequence for SARS-Cov-2 M protein (Uniprot P0DTC5) was synthesized (IDT, Newark, NJ) and cloned into a vector based on the pACEBAC1 backbone (MultiBac; Geneva Biotech, Geneva, Switzerland) with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 7xHis tag, generating a construct for expression of M-SNS-LEVLFQGP-SRGGSGAAAGSGSGS-sfGFP-GSS-7xHis[85]. Bacmid was generated by transposition in transformed DH10MultiBac cells according to the manufacturer's instructions. Sf9 cells were cultured in ESF 921 medium (Expression Systems, Davis, CA) and P1 virus was generated from cells transfected with Escort IV reagent (MillaporeSigma, Burlington, MA) according to manufacturer's instructions. P2 virus was then generated by infecting cells at 2 million cells/mL with P1 virus at a MOI ~0.1, with infection monitored by fluorescence and harvested at 72 hours. 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/mL at a MOI ~2–5. At 72 hours, infected cells containing expressed M-sfGFP protein were harvested by centrifugation at 2500 x *g* for 10 minutes and frozen at -80 °C.

Infected Sf9 cells from 1 L of culture (~15 mL of cell pellet) were thawed in 100 mL of Lysis Buffer containing 50 mM HEPES pH 8.0, 150 mM KCl, 0.5 mM EDTA, 2 mM $MgCl_2$, 2 mM DTT. Protease inhibitors (Final Concentrations: E64 (1 μM), pepstatin A (1 μg/mL), soy trypsin inhibitor (10 μg/mL), benzamidine (1 mM), aprotinin (1 μg/mL), leupeptin (1μg/mL), AEBSF (1 mM), and PMSF (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 x *g* for 45 minutes. The supernatant was discarded, and residual nucleic acid was removed from the top of the membrane pellet using DPBS. Membrane pellets were resuspended in High Salt Lysis buffer (50 mM HEPES pH 8.0, 1 M KCl, 1 mM EDTA, 2 mM DTT) plus protease inhibitors, followed by ultracentrifugation at 150,000 x *g* for 45 minutes. The supernatant was discarded, membrane pellets were washed again with DPBS, then transferred into a Dounce homogenizer containing extraction buffer (50 mM HEPES, 150 mM KCl, 1 mM EDTA, 1% lauryl maltose neopentyl glycol (LMNG, Anatrace, Maumee, OH), 0.1% cholesteryl hemisuccinate Tris salt (CHS, Anatrace, Maumee, OH) pH 8). A stock solution of 10% LMNG, 1% CHS was dissolved and clarified by bath sonication in 200 mM HEPES pH 8 prior to addition to buffer to the indicated final concentration. Membrane pellets were homogenized in extraction buffer and this mixture (150 mL final volume) was gently stirred at 4°C for 1.5 hours. The extraction mixture was centrifuged at 33,000 x *g* for 45 minutes and the supernatant, containing solubilized membrane protein, was bound to 7.5 mL of Sepharose resin coupled to anti-GFP nanobody for 1.5 hours at 4°C. The resin was then collected in a column and washed with 100 mL of buffer 1 (20 mM HEPES pH

8.0, 150 mM KCl, 1 mM EDTA, 0.005% LMNG, 0.0005% CHS), 150 mL of buffer 2 (20 mM HEPES pH 8.0, 500 mM KCl, 1 mM EDTA, 0.005% LMNG, 0.0005% CHS), and 25 mL of buffer 1. The resin was then resuspended in 7.5 mL of buffer 1 with 0.5 mg of PPX protease and rocked gently in the capped column for 2 hours. Cleaved M protein was then eluted with an additional 15 mL of buffer 1, spin concentrated to ~1 mL with Amicon Ultra spin concentrator 30 kDa cutoff (Millipore, Tullagreen, Ireland), and loaded onto a Superose 6 increase column (GE Healthcare, Chicago, IL) on an NGC system (Bio-Rad, Hercules, CA) equilibrated in buffer 1. Peak fractions containing M protein were stored overnight at 4 °C prior to preparation of Fab complexes. Excess M was flash frozen and stored at -80 °C for biolayer interferometry experiments.

3.4.2 Cloning, expression and purification of conformationally selective Fabs

Coding sequences for conformationally selective antibody fragments (lcFab and scFab)[80] were synthesized and cloned in a pTwist CMV BetaGlobin WPRE Neo vector with C-terminal 7X-His tags on heavy chain (Twist Biosciences). Fabs were cotransfected in suspension HEK293T GNTI- cells for expression using PEI according to manufacturers instructions.

For cryo-EM experiments, concentrated supernatants were obtained from Twist Biosciences (South San Francisco, CA). The 8 mL supernatants were thawed on ice, buffered with 900 uL of 1 M HEPES pH 8.0, and centrifuged at 21,000 x *g* for 10 minutes to pellet any precipitated material. Clarified supernatant was then incubated with 3 mL His-Pure Ni NTA resin (ThermoFisher Scientific, Rockford, IL) for 1.5 hours at 4 °C. The Ni NTA resin was collected in a column and washed with 30 mL each of Cholate buffer (40 mM HEPES pH 8.0, 300 mM KCl, 50 mM Na-cholate, 10 mM Imidazole) and Imidazole Wash Buffer (40 mM HEPES pH 8.0, 300 mM KCl, 20 mM Imidazole). The Fabs were eluted in 4 column volumes of Ni NTA elution buffer (40 mM HEPES pH 8.0, 300 mM KCl, 200 mM Imidazole). The eluate was spin concentrated to ~1 mL with Amicon Ultra spin concentrator 30 kDa cutoff (Millipore, Tullagreen, Ireland) and loaded onto a Superdex 200 increase (S200) column (GE Healthcare, Chicago, IL) on an NGC system (Bio-Rad, Hercules, CA) equilibrated in SEC buffer (20 mM HEPES pH 8.0, 150 mM KCl, 1 mM EDTA). Fractions containing complete Fabs were pooled, concentrated, and stored at 4 oC for future use.

For Biolayer interferometry experiments, 1L HEK293T GNTI- cells cultured in Freestyle media (ThermoFisher Scientific, Rockford, IL) were cotransfected at 3 million cells/mL with 1 mg total DNA (1:2 (w/w) heavy chain to light chain DNA ratio) and 2 mg polyethylenimine (Sigma-Aldrich, St. Louis, MO) for 4 days according to manufacturer's instructions. Supernatants of lcFab/scFab transfected cells were harvested by centrifugation at 200 x g for 5 minutes followed by an additional centrifugation at 4,000 x g for 10 minutes, then 45 µM sterile filtered into bottles and stored at 4 °C. Supernatants were then buffered with Tris pH 8.0 up to a final concentration of 40 mM, and 12.5 mL of Ni NTA resin was added to incubate overnight at 4 °C. The Ni NTA resin was collected in a gravity column and washed with 100 mL Tris Cholate buffer (40 mM Tris pH 8.0, 300

mM KCl, 50 mM Na-cholate, 10 mM Imidazole), 100 mL Tris Imidazole wash buffer (40 mM Tris pH 8.0, 300 mM KCl, 20 mM Imidazole), and eluded in 3 column volumes of Tris Ni NTA elution buffer (40 mM Tris pH 8.0, 300 mM KCl, 200 mM Imidazole). The eluate was spin concentrated with a 30 kDa MW cutoff concentrator to 1 mL and loaded on a S200 column equilibrated in SEC buffer (20 mM HEPES pH 8.0, 150 mM KCl, 1 mM EDTA) for storage. Fractions containing complete Fabs were pooled, concentrated, and stored at 4 °C for future use.

3.4.3 M protein-Fab complex purification

For cryo-EM experiments, the peak fraction from size exclusion chromatography containing pure homodimeric M protein was collected, incubated with SARS37 Fab at a 1:2.2 M monomer:SARS37 Fab molar ratio for 45 minutes on ice, and run on a Superose 6 increase column in Buffer 1 to separate excess Fabs. Pooled peak fractions containing M-SARS37 Fab complex were spin concentrated with an Amicon Ultra 50 kDa cutoff spin concentrator to 1.8 mg/mL. Concentrated sample was cleared by a 10-minute 21,000 x g spin at 4ºC and then incubated with either lcFab or scFab at a molar ratio of 1:2.2 M protein dimer:lc/scFab and incubated on ice for 30 minutes prior to grid preparation.

For biolayer interferometry experiments, purified and frozen M protein was thawed on ice and centrifuged at 21,000 x g for 10 minutes to pellet aggregates. Clarified M was then incubated on ice with 1:1 molar ratio M monomer to lcFab or scFab for 30 minutes then loaded on a Superose 6 increase column in Buffer 3 (20 mM HEPES pH 8.0, 150 mM KCl, 0.005% LMNG, 0.0005% CHS) to separate excess Fabs. Fractions containing lcFab/M and scFab/M were collected for use as immobilized bait protein on Octet Ni-NTA biosensors (ForteBio, Dallas, TX).

3.4.4 Cryo-EM sample preparation and data collection

3.4 µl of SARS37 Fab:M:lcFab or SARS37 Fab:M:scFab sample was applied to freshly glow discharged Holey Carbon, 300 mesh R 1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany) and plunge frozen in liquid ethane using a FEI Vitrobot Mark IV (ThermoFisher Scientific) with a ~5 second wait time set to 4 °C, 100% humidity, blot force 1, and 3 second blot time.

Grids were clipped and cryo-EM micrographs were collected on a Titan Krios G2 microscope operated at 300 kV. Fifty frame videos with a total electron dose of 50 e-/Å2 were recorded on a Gatan K3 Summit direct electron detector in counting mode with a pixel size of 0.7311 Å. 242 videos were recorded around a central hole position via image shift (two targets per hole, 11x11 holes). Targets were collected with a nominal defocus range of -0.5 to -1.8 µm. See Supplemental Table 3.1 for detailed data collection statistics.

3.4.5 Cryo-EM data processing

Motion correction and dose weighting were performed on all 9,320 videos using RELION 4.0's[88] implementation of MotionCor2[89] at 0.7311 Å per pixel. Contrast transfer function (CTF) parameters were fitted with CTFFIND-4.1[90]. Template-free auto-picking of particles

was performed with RELION 4.0's Laplacian-of-Gaussian filter on 7,472 movies CTF fit to 8.0 Å or better, yielding an initial set of 1,375,627 particles. These particles were then extracted at a 500-pixel box size and transferred to cryoSPARC v.4.3[91] for iterative rounds of two-dimensional classification, resulting in a set of 86,978 particles which were then extracted in RELION 4.0 and their coordinates used in the Topaz particle-picking pipeline[92].

Topaz training, picking and extraction yielded 1,934,074 particles which were extracted and then iteratively 2D classified in cryoSPARC, resulting in a set of 473,423 'good' particles. A one class ab initio reconstruction was performed to obtain an initial volume and a subsequent non-uniform refinement (C1, 2 extra passes, 16 Å initial resolution) produced a map with an overall resolution of 3.4 Å. This map was post-processed in RELION and used for Bayesian particle polishing.

The resulting 'shiny' particles were imported back into cryoSPARC for iterative rounds of multi-class ab initio and heterogeneous refinement. Subsequent non-uniform (NU) refinements of outputs from these jobs yielded similar maps with overall resolutions of 3.1-3.4 Å, however the density at interface between M and the SARS37 Fab was consistently weak and heterogeneous suggesting complex flexibility. Three NU refined maps (and their 266,191 associated particles) where this region was most clearly defined were used as inputs for 3D classification (3 classes, 3 Å target resolution, 10 O-EM epochs, RMS density change convergence check on, 14 Å initial resolution). Non-uniform refinements (C1, 1 extra final pass, 14 Å initial resolution) of two of the resultant 3D class maps generated the final full maps for separate SARS37 Fab conformations related by a rocking motion. Separate masked Local Refinements in cryoSPARC generated maps that were clearer at the Fab binding interface and were used to build models.

3.4.5 Modeling, refinement and analysis

Modeling was performed in Coot[93] using cryoSPARC sharpened and DeepEMhancer[128] post-processed (using the highRes deep learning model) cryo-EM maps with the M-lcFab cryo-EM structure (PDB ID: 7VGR) and SARS37 Fab crystal structure as initial models. The model was real space refined in Phenix[94] and validated using Molprobity[95]. Figures were prepared using ChimeraX[97], Prism, Adobe Photoshop, and Adobe Illustrator software.

3.4.6 Biolayer interferometry

Binding interactions between M protein and SARS37 Fab were measured by Biolayer Interferometry (BLI) using an Octet Red384 (ForteBio, Dallas, TX). Ni-NTA sensors were pre-incubated for at least 10 minutes in bait sample buffer (20 mM HEPES pH 8.0, 500 mM KCl, 0.005% LMNG, 0.0005% CHS) prior to running experiments. Experiments for both M:lcFab and M:scFab bait complexes were performed with the following sequence: loading bait for 240 s, buffer wash for 30 s, baseline recording for 60 s, SARS37 Fab prey association 240 s, dissociation 300 s, followed by Ni-NTA sensor

regeneration in sample buffer plus 200 mM Imidazole. A reference sensor was for background signal subtraction then data were normalized to baseline step and aligned to the dissociated step. Global curve fitting with a 2:1 heterogeneous ligand model in the ForteBio Data Analysis software most closely matched the experimental data and was used to obtain kinetic values. Results are averages of four replicates for lcFab-M experiments and three replicates for M:scFab. Data plots were generated in Prism.

3.5 Acknowledgements

We thank Dan Toso, Jonathan Remis, and Paul Tobias for support in data collection at the Cal-Cryo EM facility. We also thank Dave Savage and Luke Oltrogge for the use of their Octet Red384 BLI instrument and for their time and guidance for experimental setup and data analysis. This work was funded in part by NIH AI169896, the New York Stem Cell Foundation, a Sloan Research Fellowship (to SGB), a Fast Grants Award from Emergent Ventures at the Mercatus Center, and an NSF Graduate Research Fellowship (to KD).

3.6 Contributions

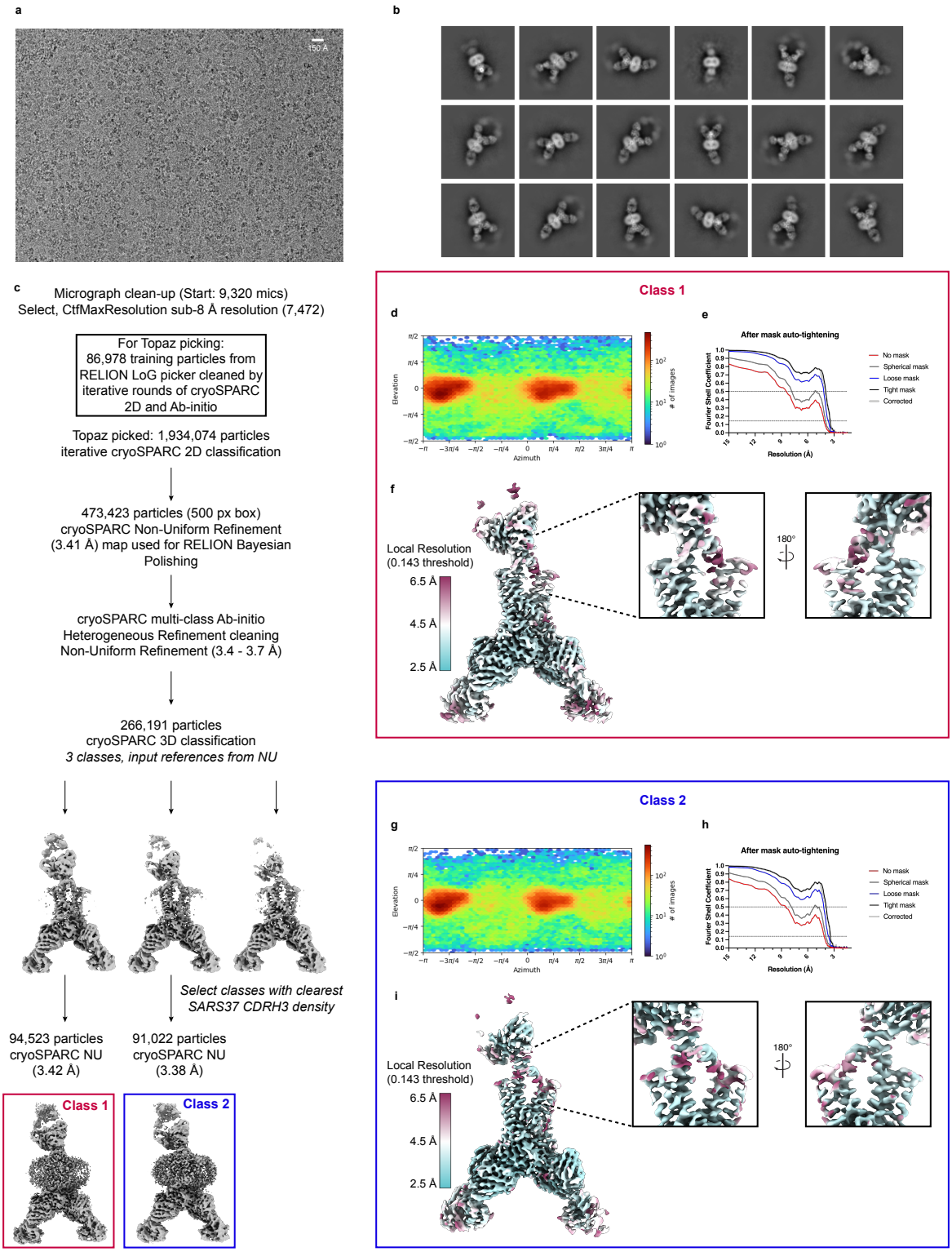
K.A.D., N.M., D.W. and S.G.B. conceived of the project. D.W. and N. M. provided the SARS37 Fab sample. K.A.D. generated final constructs and performed expression, purification, and complex formation, and bio-layer interferometry. K.A.D. performed the cryo-EM sample preparation, data collection, and data processing. K.A.D. and S.G.B. built and refined the atomic models. D.W. and S.G.B. secured funding and supervised research. K.A.D. and S.G.B. wrote the manuscript with input from all authors.

3.7 Supplemental tables and figures

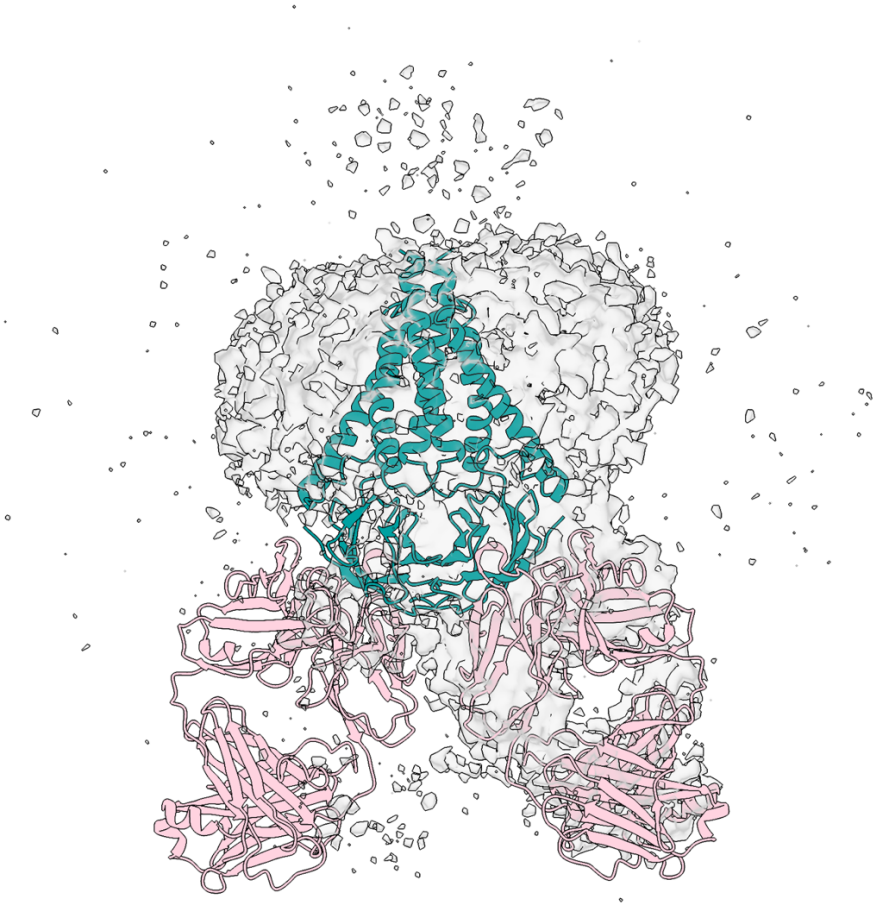
Data collection	**SAR37-M-lcFab 1**	**SAR37-M-lcFab 2**
Total movies	9,320	
Magnification	105,000x	
Voltage (KV)	300	
Electron exposure ($e^-/Å^2$)	50	
Defocus range (um)	-0.5 to -1.8	
Pixel size ($Å^2$)	0.7311	
Processing		

Initial particle images (no.)	1,934,074	1,934,074
Final particle images (no.)	94,523	91,022
Map resolution Masked (Å, FSC = 0.143)	3.42	3.38
Symmetry imposed	C1	C1
Refinement		
Model resolution (Å, FSC = 0.143 / FSC = 0.5)	3.3/3.7	3.5/3.8
Map-sharpening B factor ($Å^2$)	-95.1	-87.5
Composition		
Number of atoms	8638	8600
Number of protein residues	1100	1095
Number of ligands	0	0
RMS deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (Å)	0.496	0.534
Validation		
MolProbity score	1.52	1.64
Clashscore	3.84	4.97
Ramachandran plot		
Favored (%)	95.02	94.53
Allowed (%)	4.98	5.47
Disallowed (%)	0	0
Rotamer outliers (%)	0.54	0.11
Mean B factor ($Å^2$)		
Protein	81.50	85.02

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



Supplemental Figure 3.1 Cryo-EM data, processing pipeline, and validation for SARS37:M:IcFab complex in LMNG/CHS. (a) Representative micrograph. **(b)** Selected 2D class averages and **(c)** Cryo-EM data processing pipeline. Validation for conformations 1 **(d-f)** and 2 **(g-i)**.



Supplemental Figure 3.2 Cryo-EM map for SARS37:M:scFab complex in LMNG/CHS. (a) Representative surface map with PDB: 7VGS model docked in. M is colored in dark teal, scFabs in pink, SARS37 Fab density is missing and not modeled.

Chapter Four

Direct SARS-CoV-2 M-lipid interaction controls M conformational dynamics and virus assembly

This chapter presents work done in collaboration with the laboratory of Prof. Robert Stahelin at Purdue University and Greg Voth at the University of Chicago. The preliminary lipid binding and sedimentation assays were performed in the Stahelin lab and presented here with their generous permission. I performed all other experiments described in this chapter, interpreted the results, and wrote the chapter with input and feedback from Steve Brohawn.

4.1 Introduction

Coronaviruses are lipid envelope encapsulated single-stranded RNA viruses with a genome encoding four structural proteins – spike (S), membrane (M), envelope (E), and nucleocapsid (N)[40]. The lipids that comprise the envelope are derived directly from host membranes at the site where viral budding occurs[129]. Viruses often utilize specific lipids to drive every step of the infection cycle – from viral entry[130–133], to biosynthesis[134,135], assembly and egress[136,137] – and have been shown to modulate the composition of cellular membranes in order to facilitate these processes[138].

The involvement of lipids in betacoronavirus infection has been documented for decades. Studies have shown that mouse hepatitis virus and SARS-CoV-2 viral particles target cholesterol and ceramide rich microdomains for viral entry[133] and that depletion of cholesterol with cyclodextrins leads to a significant decrease in S receptor binding in cells[139]. Genome amplification and biosynthesis in coronaviruses is marked by the formation of double-membrane vesicles (DMV) driven by non-structural proteins[134], found to be enriched with triacylglycerols and cholesterol[134,140]. Coronavirus nucleocapsid proteins have long been known to associate with lipid membranes, whether free or bound to viral RNA[141,142]. More recently a preprint reported specific interactions between N protein and anionic lipids which could drive localization to sites of viral transcription and virion assembly[143].

The SARS-CoV-2 M protein is an abundant and highly conserved structural membrane protein known to facilitate viral particle assembly through direct protein-protein interactions with each of the other three structural proteins S, E, and N[40]. Additionally, M has been shown to be essential for budding minimal virus-like particles [43]. M is a 50 kDa homodimer with three transmembrane helices per monomer and a C-terminal domain (CTD) with a unique β-sandwich fold featuring a strikingly electropositive surface[80,126]. The CTD is initially exposed to the cytoplasm at the budding site of the ER-Golgi intermediate compartment where viral assembly occurs, and ultimately faces the interior of the virion where the electropositive surface is believed to facilitate interactions with the dense core of viral RNA and nucleocapsid protein[40].

Despite the M protein's critical roles in infectious viral particle assembly and release, and the growing evidence of the critical importance of lipids in these processes, and the potential involvement of specific M-lipid interactions remains unstudied. Here we demonstrate that the SARS-CoV-2 M protein specifically binds to ceramide-1-phosphate (C1P) and investigate the molecular basis for C1P modulation of M protein conformational dynamics using single-particle cryo-EM. We show that C1P binds to and stabilizes M in the short conformation, but not the long conformation. Analyzing differences between our C1P-bound M model and previously determined structures of short and long form M, we characterize residues across the C-terminal domain that contribute to the interchange between the two conformers. Based on these observations, we design point mutants with the intention of stabilizing M in the long conformation and validate with AlphaFold[71] structural predictions.

4.2 Results and discussion

Preliminary lipid binding and sedimentation assay data performed by our collaborators indicated that the SARS-CoV-2 M protein can efficiently bind to the anionic sphingolipid ceramide-1-phosphate (C1P). They also saw binding to dihydro C1P and sphingosine-1-phosphate (S1P), as well as phosphatidic acid (PA), albeit to a significantly lesser extent (Supplemental Fig. 4.1). These data suggested selectivity for the phosphate headgroup as well as the sphingoid backbone. Investigating specific protein-ligand interactions by cryo-EM requires a homogenous sample and our previous work with apo M in nanodiscs suggests it is too small and flexible to achieve the resolution necessary for this analysis[144]. To overcome this barrier, we used Fabs generated by Zhang et al. in a previous study[80] that bind to the CTD of M and stabilize two distinct conformations that M can adopt – a long and short form – which we will refer to as long-conformationally selective Fab (lcFab) or short-conformationally selective (scFab). These Fabs allow us to investigate differences in lipid binding behavior between the short and long conformation of M, as well as aid in cryo-EM alignments for high resolution structure determination. In order to be able to directly compare our results with the previously published models, we prepared our M:Fab complexes in LMNG/CHS detergent and incubated each complex in 100 μM C1P prior to freezing.

The cryo-EM structure of M stabilized in the short conformation by scFabs in the presence of C1P shows several lipid-like densities that were not present in the published maps (EMD-31978). One density on the inter-virion leaflet is complete and can clearly be modeled as C1P (Fig. 4.1a). The C1P phosphate head group is coordinated by nearby positively charged residues forming potential electrostatic interactions with Arg44 on TM2, Arg131 on the C-terminal domain, and Trp110 on the hinge domain (Fig. 4.1b). Overall, these residues contribute to a cationic pocket that the headgroup fits into (Fig. 4.1c). While portions of this C1P density were clearly visible throughout processing this dataset, the entire molecule was only able to be visualized with a subset of particles. This suggests that there is some flexibility allowed in this binding pocket, likely between

the headgroup position and Arg44 which was the least clearly resolved density in the binding site.

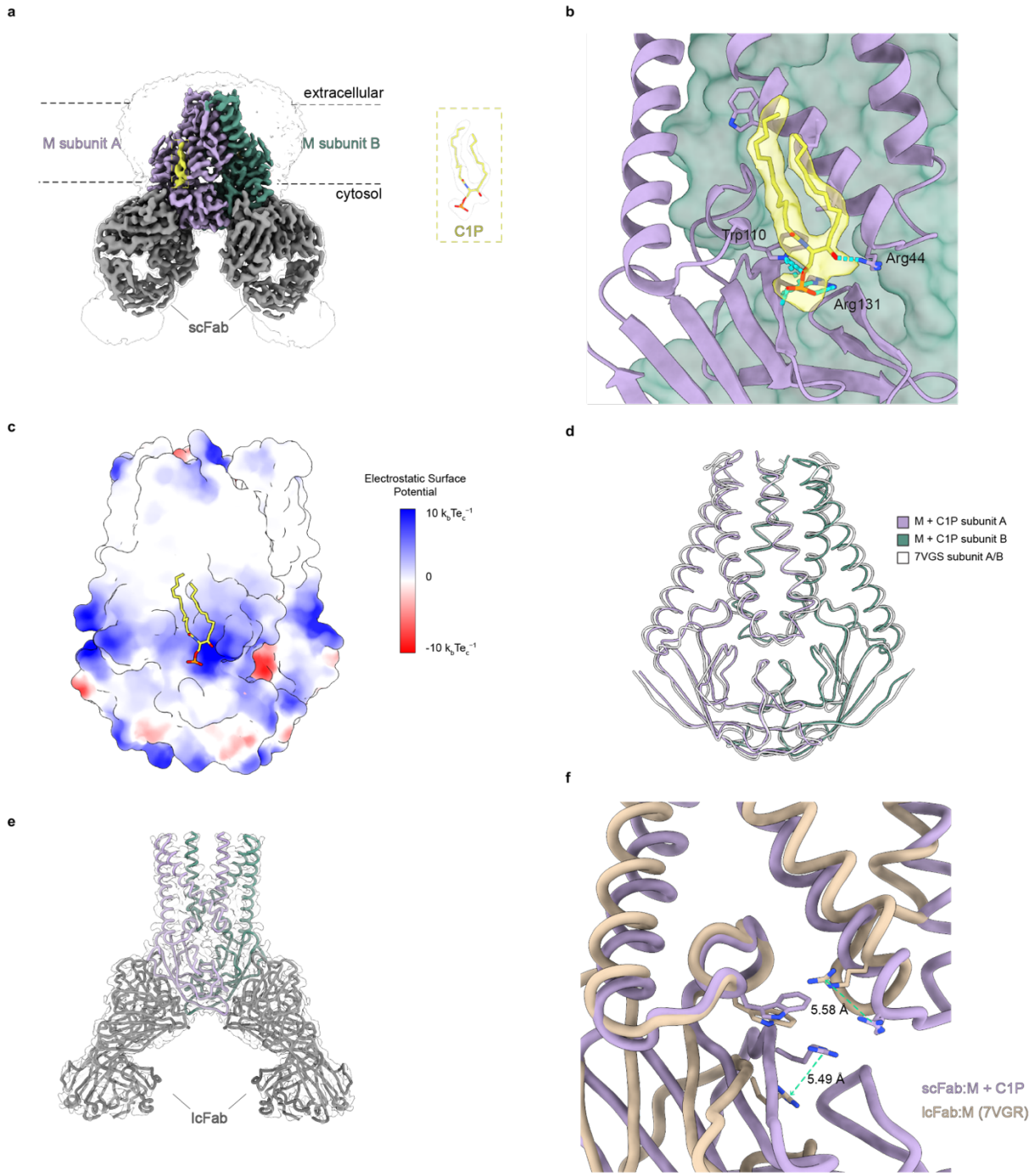


Figure 4.1 M in the short conformation is stabilized by C1P binding. (a) 3.03 Å resolution cryo-EM map of short conformation M in complex with scFabs. M subunits A and B are colored purple and green, respectively, scFabs are in gray and C1P is in yellow. Unsharpened map (transparent white) is overlaid with cryoSPARC post-processed used for modeling. C1P model and density fit are shown in the right panel. **(b)** Close-up of the C1P binding site for subunit

A in purple. Subunit B is shown as a transparent green surface. Cyan dashed lines represent potential electrostatic interactions. **(c)** Surface of M + C1P colored according to electrostatic surface potential from red (electronegative) to blue (electropositive). The C1P molecule is in yellow. **(d)** Overlay of M:scFab + C1P model (in purple and green) with published M:scFab model (PDB: 7VGS) in white. **(e)** Published long conformation model (PDB: 7VGR) docked into M:lcFab + C1P map (transparent). **(f)** Close-up of C1P binding site for subunit A with M:scFab + C1P model in purple and aligned published long conformation model (7VGR) in tan. Movement of C1P headgroup coordinating residues between conformations are indicated in green.

In our map we see one of the acyl chains packed into the cleft between TM2 and TM3, against hydrophobic residues I48, I52, and Trp55 on TM2. This acyl chain density is very strong and could be seen early in data processing and was also weakly visible in the map associated with the published model (EMD-31978), suggesting that a proportion of those associated particles contain bound lipids that co-purified from HEK293 cell membranes. The second acyl chain, which is oriented away from M's transmembrane domain into the lipid bilayer, is likely more flexible and is thus less well resolved.

Overall, the model of the M:scFab complex in the presence of C1P is very similar to the published structure (PDB 7VGS), with a few notable differences. While TM1 and TM2 overlay well between our model and 7VGS, we see that TM3 of both C1P bound subunits are compacted slightly towards the center of the molecule, with an overall shift of 1.5-2 Å for residues on TM3 (Fig. 4.1d). This shift continues into the hinge domain but the majority of the CTD is the same. While our reconstruction and EMD-31978 are to similar overall resolutions (3.0 Å and 2.8 Å, respectively), we observe a pronounced stabilizing effect with C1P binding. This effect can most clearly be seen in flexible regions of M, such as the hinge domain and loops between TMs, where the M:scFab + C1P density is significantly better resolved. C1P-induced stabilization is further supported by reduced variability of refined B-Factors and lower local resolution despite similar particle numbers and similar processing routines[80] (Supplemental Fig. 4.2, Supplemental Fig. 4.4).

In contrast to our short conformation data, no evidence for specific C1P binding was observed from a cryo-EM structure of M in the long conformation in presence of C1P. An M:lcFab + C1P reconstruction showed neither additional density attributable to specific C1P binding or clear protein conformational changes. In the published map for the M:lcFab complex, short acyl-chain-like densities can be seen surrounding M on the inter-virion leaflet (EMD-31977). We see corresponding densities in our map at similar resolution (3.0 Å and 2.7 Å, respectively) (Fig. 4.1e, Supplemental Fig. 4.3), with size and location matching between datasets. These densities could be lipids that co-purify from their respective expression system or LMNG hydrocarbon tails. However, we do not see any evidence of a specific C1P binding site and the published model 7VGR fits the M:lcFab + C1P map very well (Fig. 4.1e). Consistently, we do not see evidence of overall stabilization of the long conformation by C1P when compared to the previously published apo structure. Comparison of short and long conformations shows reorganization of C1P-interacting residues that explains the absence of specific C1P binding in the long conformation: in the M_{long} model, two key arginines (Arg44 and Arg131) move > 5 Å away from the C1P headgroup observed in M_{short} (Fig. 4.1f). Taken

together, these data suggest C1P specifically binds and stabilizes the short conformation.

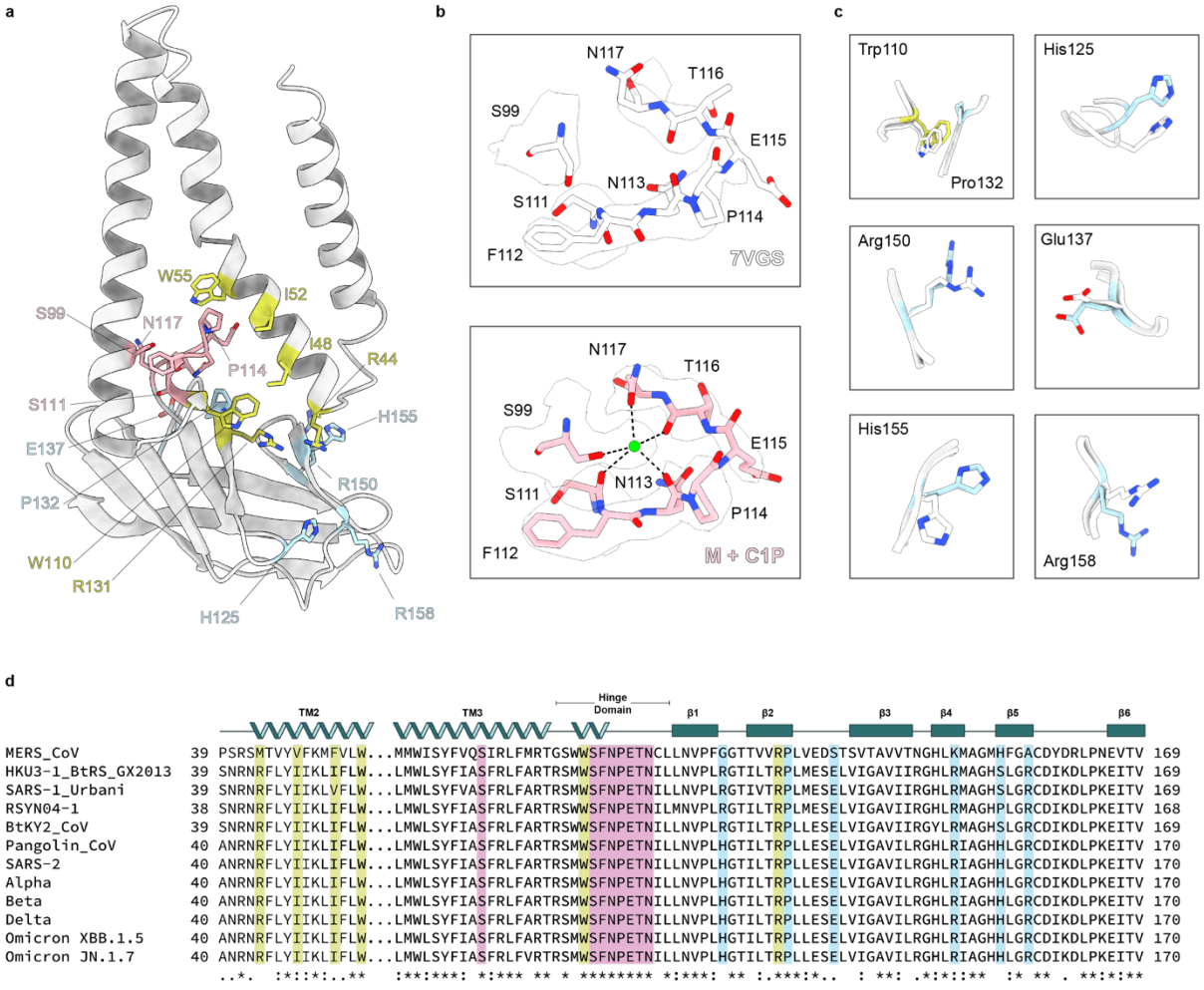


Figure 4.2 Differences between our M:scFab + C1P model to 7VGS are distributed throughout the transmembrane and C-terminal domains. (a) Subunit A of scFab:M + C1P model colored in light gray. Residues with differences between the +/- C1P models are shown and colored by region. C1P binding site residues are in yellow, transmembrane domain ion binding site is in pink, and C-terminal domain residues are light blue. **(b)** Close-up views of the scFab:M + C1P ion binding site in pink with K^+ ion modeled in green (bottom), and missing ion density of 7VGS in white (top). **(c)** Highlighted changes in rotamers between the scFab:M + C1P model (colored as indicated in **(a)**), and 7VGS in white. **(d)** Multiple sequence alignment of M protein from SARS-CoV-1, SARS-CoV-2 variants, and other sarbecoviruses, with degree of conservation below (conserved residues are marked with an asterisk and conservative substitutions are marked with a colon). Residues are highlighted as indicated in **(a)**.

There are additional differences that we observe in our M:scFab + C1P map that extend away from the C1P binding site and may give insight into M's conformational equilibrium. In addition to the compaction of TM3 towards TM2 that we see in our M:scFab + C1P structure, we see small changes in rotamers of S99 and N117. A consequence of this small rearrangement is the presence of a density that appears to

be an ion that is not seen in the Zhang et al. map[80] (Fig. 4.2a-b). The site is surrounded by side chain oxygens on Ser99 and Asn117, and backbone carbonyls from Ser111, Asn113, and Thr116, with distances from the center of the density ranging from 2.0-3.4 Å. We prepared our sample in the presence of 150 mM KCl and reasoned that that likely indicated this to be a potassium binding site. Our M:scFab + C1P buffers contain 1 mM EDTA, however it has been shown that cations, calcium in particular, can be introduced at high concentrations during the blotting step of vitrification[145], and thus cannot be ruled out on that basis alone. Multiple lines of evidence support metal binding to this site. First AlphaFold3[71] predictions of M dimer structures with two cation (potassium, sodium, calcium, and magnesium) (Supplemental Fig. 4.5) all show ion occupancy of the putative binding site with minimal distortions to the overall fold. Second, Metric Ion Classification (a recently developed deep learning tool for assigning ions in cryo-EM maps[146]) and CheckMyMetal (CMM, an ion validation tool that accounts for ligand chemistry, geometry, and refined B-factor statistics[147]) both predict metal over water binding to this site. We therefore separately modeled each ion in the structure, refined the model in Phenix[94] and compared validation statistics in CMM. The site was predicted in all cases to be most likely occupied by potassium and K+ is included in the final models.

Interestingly, one of the regions that best matched the published model (PDB: 7VGS) were the residues that we observed to specifically interact with the C1P headgroup. One exception is a very slight shift in the position of Trp110 on the hinge domain towards the CTD, which in turn appears to induce a small change in the position of Pro132 on β2 strand (Fig. 4.2c). Looking over the rest of the model we see other clear differences in rotamers distributed over nearly the entire CTD. Aside from Pro132, these differences involve charged residues. His125 on β1 of the CTD rotates towards Arg158 on β5, which rotates away accordingly. A trio of residues, Glu137, Arg150 and His155, all reside on the dimer interface of the CTD. Comparing the C1P bound short conformation to 7VGS, we see a notable switch in the relative positions of His155 and Arg150 that occur in concert with each other. On the opposing subunit, Glu137 shifts as well to avoid a steric clash with Arg150. Finally, in the transition from the published short (7VGS) to long (7VGR), the entire loop containing His155 and Arg150 slides 7 Å away from the opposing Glu137, while retaining the same relative rotamer positions (Fig 4.2c, Fig. 4.3). Overall, these changes result in subtle differences in the electrostatic charge character of the CTD (Supplemental Fig. 4.7) and could play a role in M recognizing specific binding partners during virion particle assembly.

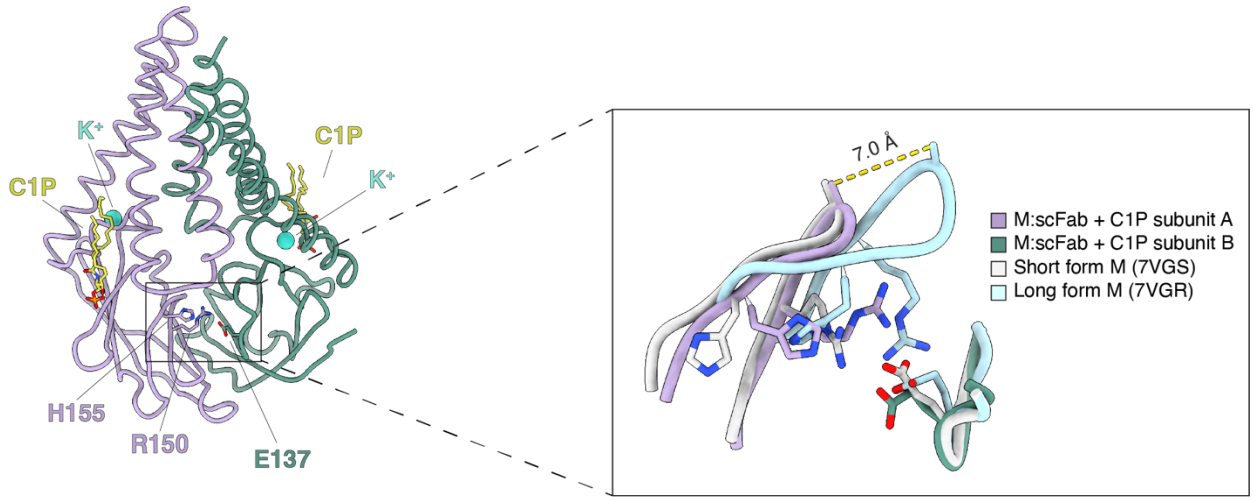


Figure 4.3 Conformational change between subunits is directed by a trio of residues at the CTD dimer interface. Model of M:scFab + C1P with subunit A residues Arg150 and His155 in purple, and Glu137 on subunit B in green shown. C1P and K^+ ligands are colored in yellow and turquoise, respectively. Inset is an overlay of the 3 CTD residues from M:scFab + C1P (subunit A in purple, subunit B in green), short form M (7VGS in light grey), and long form M (7VGR in light blue) aligned on the CTD of subunit B. The dashed yellow line indicated the distance between Cβ atoms of Ala152 of M:scFab + C1P and 7VGR.

A recent preprint characterized a small molecule, JNJ-9676, that binds to M in the short conformation and sits in the pocket between TM2 and TM3 above the hinge domain – a binding site that overlaps with the potassium ion binding site[42]. The authors found that the drug prevents the production of infectious viral particles, likely a consequence of blocking the ability for M to adopt the long conformation, suggesting that the short to long transition is critical for M function during the process of viral assembly or escape. They characterized escape mutants that had reduced potency compared to wild type M, which included residues such as S99, N117, and P132 near the drug binding site that we have already discussed, as well as others distributed over the CTD that are more difficult to rationalize. Given that the drug does not bind to M in the long conformation, we speculated mutations at these residues might disrupt drug binding by shifting M's conformational equilibrium to promote long over short. To investigate this question, we modeled structures of a series of point mutants based on our findings using AlphaFold3. While wild-type M is predicted to adopt a short-like conformation, consistent with our previous structural and MD results, several mutants instead are predicted to adopt an intermediate or long-like conformation (Fig. 4.4).

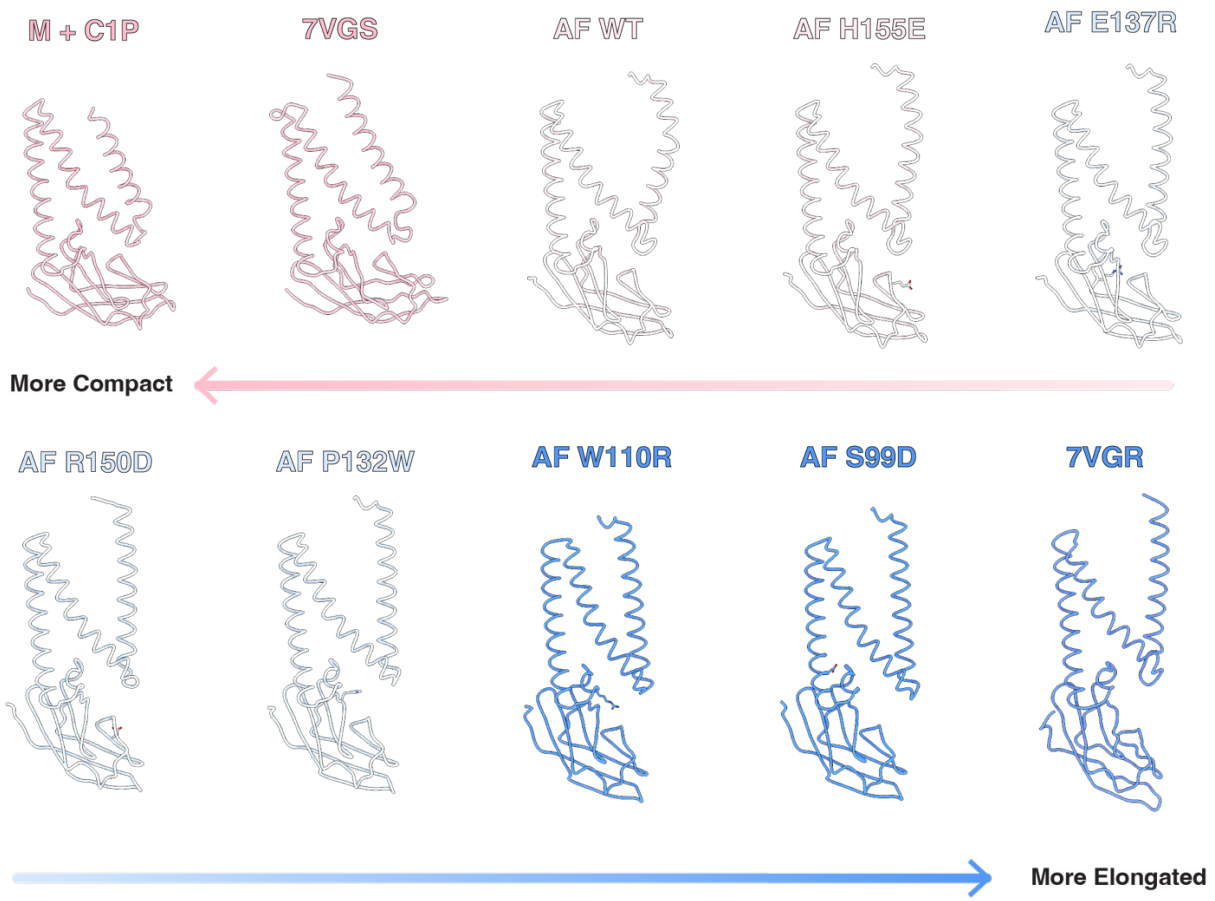


Figure 4.4 AlphaFold predicted models of CTD mutants show a distribution of conformations between long and short. Models are arranged in order from most compact (M:scFab + C1P) in pink to most elongated (7VGR) in blue and colored in a gradient according to their position.

Starting with the potassium ion binding site, we analyzed mutants of the major ion coordinating residues, which sit right above the hinge domain. In the transition from the short to long conformation (Supplemental Fig. 4.6) the hinge is shifted down which in turn pushes the CTD down, away from the plane of the membrane. We reasoned that adding bulkier or charged residues to disrupt the tightly packed potassium binding site might also induce this effect. Consistent with this theory we found AlphaFold prediction of S99D and S99Q mutants to closely resemble the long conformation structure (Fig. 4.4). Any mutations of N117 we analyzed did not produce this effect and instead more closely resembled the AlphaFold prediction for wild-type M and the experimentally determined short conformation (7VGS). The orientation of this residue allows it to project into the membrane, without disrupting the packing of the TM2-3 and hinge domain interface. The P132W on β2 of the CTD was predicted to result in a similar effect to the S99 mutants, despite the hinge remaining closer to its short form “up” position. This appears to be due to a steric clash with the C1P binding residue T110 on the hinge, resulting in the same downward rotation of the CTD seen in the long conformation model. Surprisingly, charge swap mutants of both CTD dimer interface residues (E137R and

R150D/E) were also predicted to be in the long conformation – suggesting the cooperativity of the short to long transition between subunits (Fig. 4.3).

Taken together, these findings suggest that shifting from the short to long conformation involves changes that propagate from TM3 through the hinge domain, into the CTD, and finally to the CTD dimer interface. Our C1P model represents the 'most compact' on the continuum from fully short to most elongated (Fig. 4.4). The residues involved are very well conserved amongst SARS-CoV-2 variants and related sarbecoviruses, suggesting a conserved principle underlying M conformational dynamics (Fig. 4.2c). The broad distribution of residues involved in this conformational change would naturally explain the high degree of conservation observed for the entire CTD and highlight its great potential as a target for novel therapeutics.

4.3 Conclusion

The M protein is the most abundant coronavirus structural membrane protein and plays an essential role in viral assembly and budding[40]. M is a 50 kDa dimer and is known to undergo conformational changes from a compact to elongated form[80]. These conformational changes are believed to be functionally important and associated with specific timing in the viral life cycle, although the precise molecular determinants that influence the M protein's conformational dynamics are not well understood. We used single-particle cryo-EM to investigate specific interactions between M and the anionic sphingolipid C1P in both the short and long conformations, in order to lend insight into these dynamics. Our structure of M stabilized in the short conformation by Fabs revealed a C1P binding site residing on the cleft between transmembrane helices 2 and 3 on the inter-virion surface – coordinated by a trio of positively charged residues. This binding site is disrupted in the long conformation, and we see no evidence of C1P binding to M in our long form data, suggesting that M:C1P interactions occur with specific timing in the viral life cycle.

C1P binding to short form M induces a further compaction of the transmembrane domain compared to previously determined structures of M in the short conformation in detergent and nanodiscs. One consequence of this compaction is the appearance of an ion binding site between TM2 and TM3 above the hinge domain. We also found a series of side chain rotamer changes in residues distributed across hinge and C-terminal domains, distal to the C1P binding site. Based on these observations, we wondered if they might be due to the compaction we observe with C1P bound inducing small, propagated movements that extend from TM3. We used AlphaFold3[71] to investigate these residues by predicting the structure of single point mutants. We found several regions where mutations produced predictions that closely match the cryo-EM structure of the long conformation and contribute to a picture of how the change from short to long form occurs.

In the long conformation the hinge domain is shifted down, pushing β1 of the CTD away from the plane of the membrane. Adding bulk or charge swapping S99 at the ion binding site can induce this same effect. Interestingly, even while the hinge domain remains up, a P132W mutant on β2 is predicted to produce this same rotation in the CTD seemingly due to steric avoidance of T110 on the hinge. Farther away on the CTD we found a pair of opposing residues at the CTD dimer interface, E137 and R150, that rotate around each other in the transition from the most compact (C1P bound), to short, to the long conformation, demonstrating how movements may be translated across subunits. Taken together, our findings produce a picture of conformational dynamics and stability, suggest a conserved role for specific M-lipid interactions in assembling viral particles, and present a potential avenue for novel therapeutics against SARS-CoV-2 and related sarbecoviruses.

4.4 Material and Methods

4.4.1 Cloning, expression, and purification of SARS-CoV-2 M protein and Fabs

Full-length SARS-CoV-2 M protein was expressed and purified in LMNG/CHS detergent from Sf9 insect cells as described in section 3.4.1 of this document. Conformationally selective Fabs (scFab and lcFab)[80] were expressed and purified in HEK293T GNTI- cells as described in section 3.4.2.

4.4.2 M protein-Fab complex formation and purification

The peak fraction from size exclusion chromatography containing pure homodimeric M protein was collected, incubated with purified lcFab or scFab at a 1:2.2 M monomer:Fab molar ratio for 45 minutes on ice, and run on a Superose 6 increase column in Buffer 1 to separate excess Fabs. Pooled peak fractions containing M:Fab complexes were spin concentrated with an Amicon Ultra 50 kDa cutoff spin concentrator to 2.2 mg/mL (scFab) and 2.7 mg/mL (lcFab). Concentrated complex samples were cleared by a 10-minute 21,000 x g spin at 4 ºC, then C16 ceramide-1-phoshate d18:1/16:0 (Avanti Polar Lipids, Alabaster, AL) (C1P) was added to each to a final concentration of 100 µM. Samples were then incubated on ice for 30 minutes prior to grid preparation.

4.4.3 Cryo-EM sample preparation and data collection

3.4 µl of M:lcFab or M:scFab in LMNG/CHS/C1P samples were applied to freshly glow discharged Holey Carbon, 300 mesh R 1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany) (M:scFab) or C-Flat Au 300 mesh 1.2/1.3 (Electron Microscopy Sciences, Hatfield, PA) (M:lcFab) then plunge frozen in liquid ethane using a FEI Vitrobot Mark IV (ThermoFisher Scientific) with a ~5 second wait time set to 4°C, 100% humidity, blot force 1, and 3 second blot time.

Grids were clipped and cryo-EM micrographs were collected on a Titan Krios G3i microscope operated at 300 kV. Fifty frame videos with a total electron dose of 50 e-/Å2 were recorded on a Gatan K3 Summit direct electron detector. The M:scFab dataset was collected in super resolution mode with a pixel size of 0.4055 Å (0.8110 Å physical pixel

size), and the M:IcFab dataset was collected in counting mode with a pixel size of 0.848 Å. Eleven videos were recorded around a central hole position via image shift. Targets were collected with a target defocus range of -0.5 to -1.8 µm. See Supplemental Table 4.1 for detailed data collection statistics.

4.4.4 Cryo-EM data processing

Data processing for both samples were done entirely in cryoSPARC v4.4.1[91]. For M:scFab in LMNG/CHS/C1P 6,741 movie stacks were collected, motion-corrected, binned to 0.8110 Å $pixel^{-1}$ using Patch Motion Correction, and CTF-corrected using Patch CTF Estimation (Supplemental Fig. 4.2). Micrographs with an estimated CTF resolution below 7 Å were discarded, leaving 5,105 for further processing. Initial particle picking was performed using cryoSPARC's reference free blob picker and curated from 404,953 particles down to 29,857 'good' particles by 2D classification.

Topaz training, picking and extraction yielded 351,588 particles which were then iteratively 2D classified, resulting in a set of 163,911 'good' particles. A two-class ab-initio reconstruction was performed to further sort out 44,905 particles that did not contribute to a high-resolution 3D map. Subsequent non-uniform refinement (C1, 2 extra passes, 14 Å initial resolution) produced a map with an overall 2.94 Å resolution showing clear lipid densities on both leaflets. This map and corresponding 119,006 particles were subjected to local CTF refinement, then global CTF refinement, followed by polishing via Reference Based Motion Correction which further improved the resolution to 2.7 Å with improved lipid densities. A round of 3D classification (2 classes, 3 Å target resolution, 10 O-EM epochs, 0.95 class similarity) and subsequent C2 non-uniform refinement on the class with the clearest lipid densities produced a final 3.03 Å resolution map (44,726 particles) where a C1P molecule could be unambiguously modeled (Supplemental Fig. 4.2, Supplemental Table 4.1). The same non-uniform refinement with C1 symmetry produced a 3.26 Å resolution map with no apparent differences between subunits or lipid binding, so the C2 map was chosen for building an atomic model as the C1P densities were more complete and clearly defined.

The M:IcFab LMNG/CHS/C1P dataset was processed in a similar manner, with 9,910 initial micrographs culled to 8,432 using a 5 Å cutoff. Topaz was trained on 136,478 particles yielding a stack of 561,632 particles, which was cleaned, refined and polished as described for the scFab dataset, yielding 301,406 'shiny' particles. A round of 3D classification yielded a class of 190,251 particles with 3.01 Å final resolution after non-uniform refinement with C2 symmetry (Supplemental Fig. 4.3, Supplemental Table 4.1). Further attempts at 3D classification and heterogenous refinement failed to separate these particles into subsets with unique characteristics or improved map quality.

4.4.5 Modeling, refinement, and analysis

Modeling of the M:scFab complex in LMNG/CHS/C1P was performed in Coot[93] using cryoSPARC sharpened cryo-EM maps with the SARS-CoV-2 M protein short conformation structure (PDB ID: 8CTK) and an AlphaFold3[71] model of the scFab

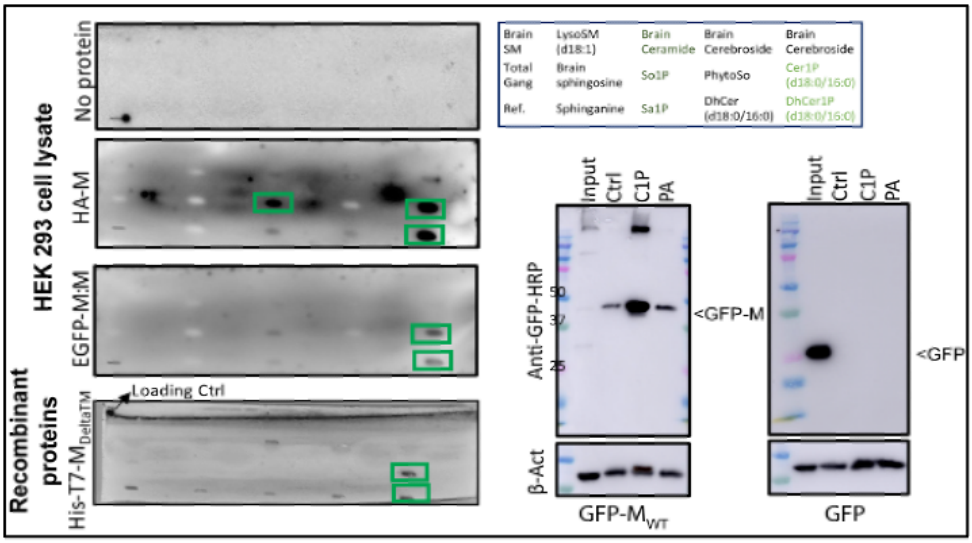
variable domains rigid body fitted into the map in Coot as initial models. The C1P ligands (3 letter code 1PX) were added in coot. The complex with ligands was real space refined in Phenix[94] and validated using Molprobity[95]. Mutant models were generated with AlphaFold3. Figures were prepared using ChimeraX[97], Prism, Adobe Photoshop, and Adobe Illustrator software.

4.5 Supplemental Tables and Figures

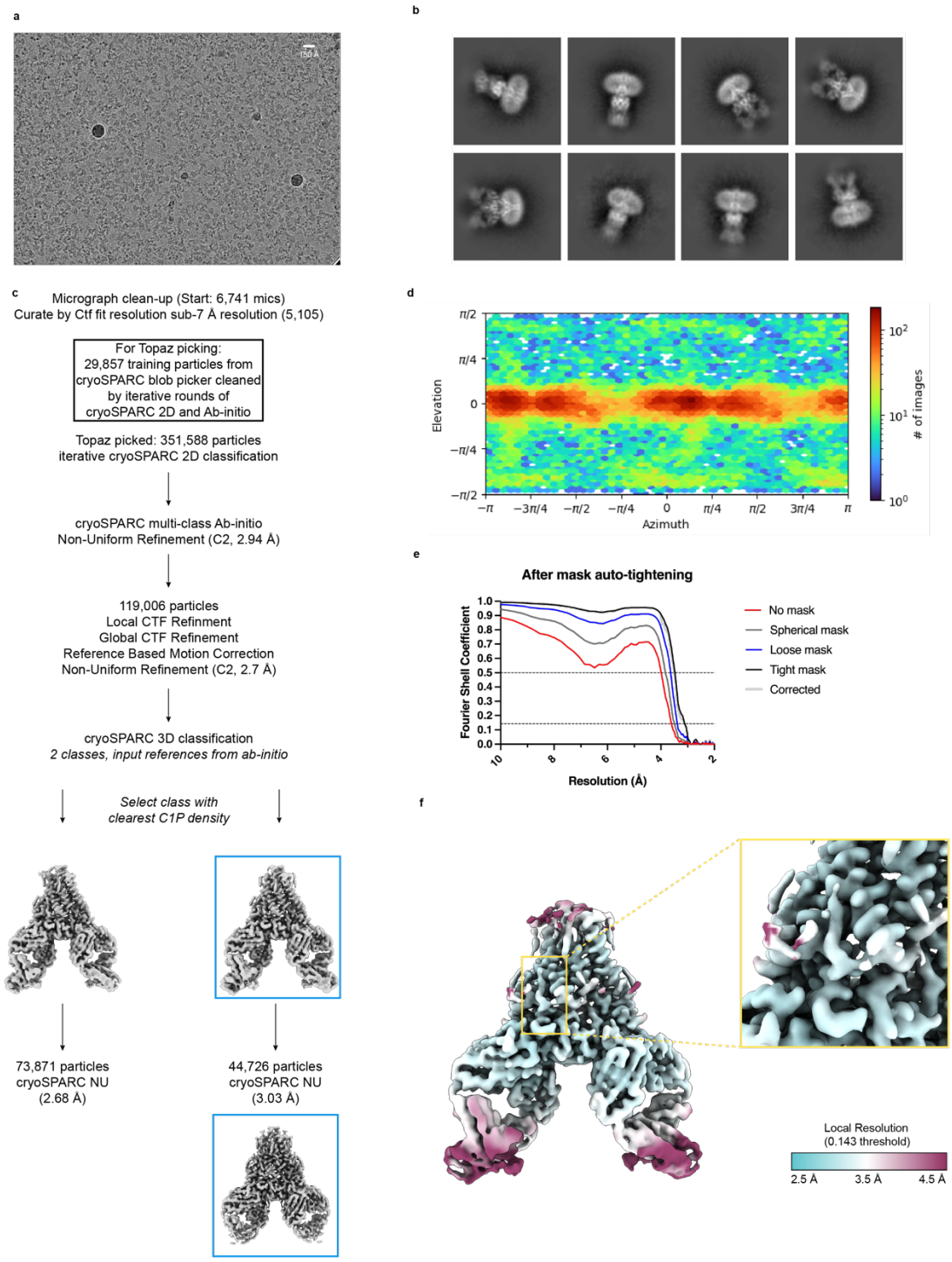
Data collection	**M:scFab + C1P**	**M:lcFab + C1P**
Total movies	6,741	8,432
Magnification	105,000x	
Voltage (kV)	300	
Electron exposure ($e^-/Å^2$)	50	
Defocus range (um)	-0.5 to -1.8	
Pixel size ($Å^2$)	0.4055 (super res)	0.848
Processing		
Initial particle images (no.)	1,934,074	1,934,074
Final particle images (no.)	94,523	91,022
Map resolution Masked (Å, FSC = 0.143)	3.03	3.01
Symmetry imposed	C2	C2
Refinement		
Model resolution (Å, FSC = 0.143 / FSC = 0.5)	3.0/3.3	-
Map-sharpening B factor ($Å^2$)	-108.4	-119.0
Composition		
Number of atoms	6584	-
Number of protein residues	828	-
Number of ligands	4	-
RMS deviations		
Bond lengths (Å)	0.003	-
Bond angles (Å)	0.465	-
Validation		
MolProbity score	1.74	-
Clashscore	6.54	-
Ramachandran plot		
Favored (%)	96.45	-

Allowed (%)	3.55	-
Disallowed (%)	0	-
Rotamer outliers (%)	1.57	-
Mean B factor (Å^2)		
Protein	58.43	-
Ligand	65.28	-

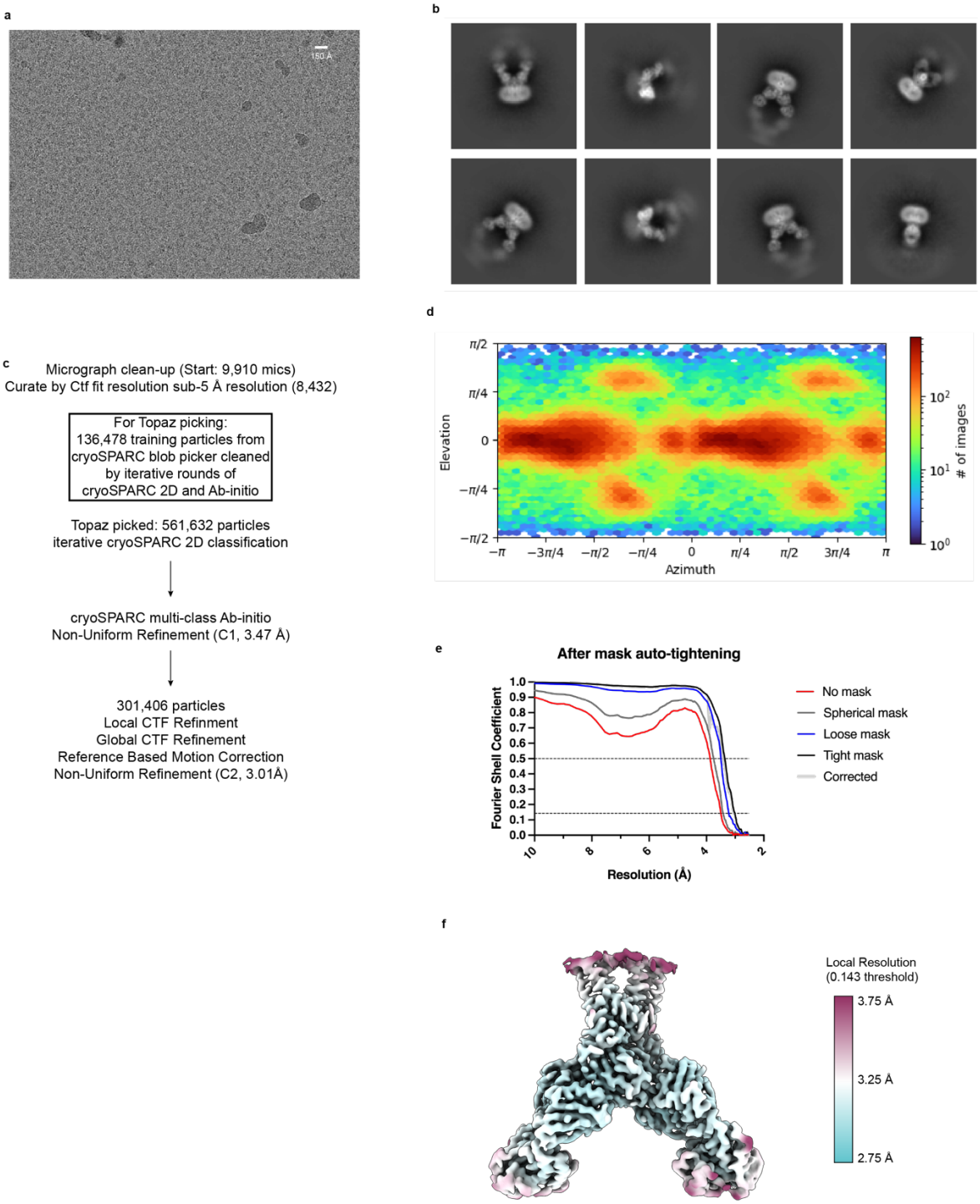
Supplemental Table 4.1 Cryo-EM data collection, refinement, and validation statistics.



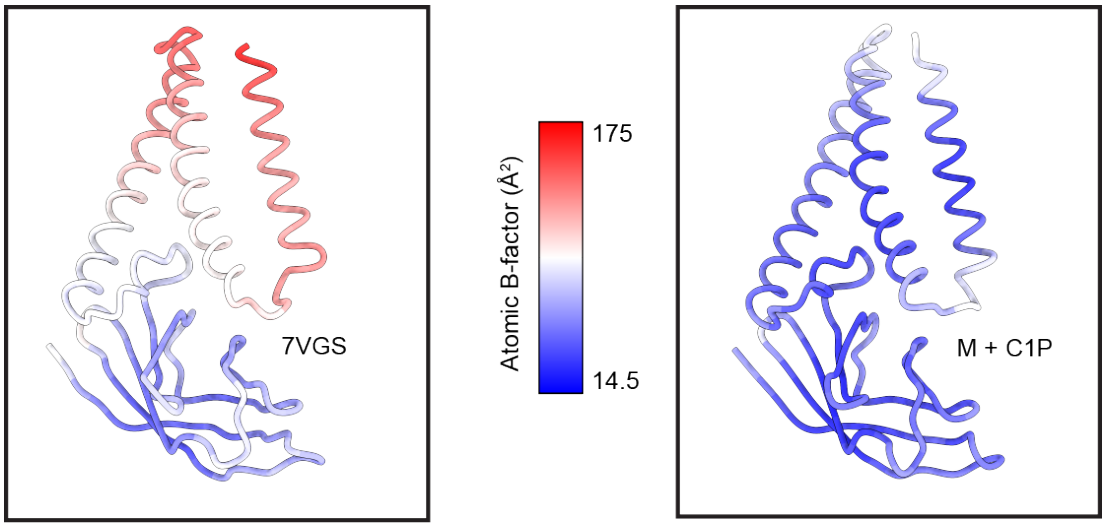
Supplemental Figure 4.1 The SARS-CoV-2 M protein binds ceramide-1-phosphate (C1P). (a) Avanti sphingolipid snoopers (lipid-overlay assay) were used to detect binding of SARS-CoV-2 M with an HA-tag (HA-M), SARS-CoV-2 M with a EGFP tag (EGFP-M) and a mock transfection control. The top 3 panels in A were done with cell lysates from HEK293 cells. M bound C1P, dihydroC1P, and sphingosine-1-phosphate (S1P, So1P in legend). See legend in upper right. The M cytoplasmic domain (lacking the transmembrane domain (residues 101-222)) was then expressed and shown to bind C1P and DHC1P (bottom panel). **(b)** C1P, PA and control beads were then used for pull down assays of M protein. Note these beads have been verified by Echelon Biosciences and verified by the Stahelin lab in previous studies of C1P and PA-binding proteins. M protein oligomer (left panel) binds to C1P but not control or PA beads whereas GFP control (right panel) does not bind lipid beads. The M protein monomer with GFP tag bound to C1P and did exhibit some binding to PA (lower bands on left). Beta-actin was used as a loading control in all experiments (done in triplicate).



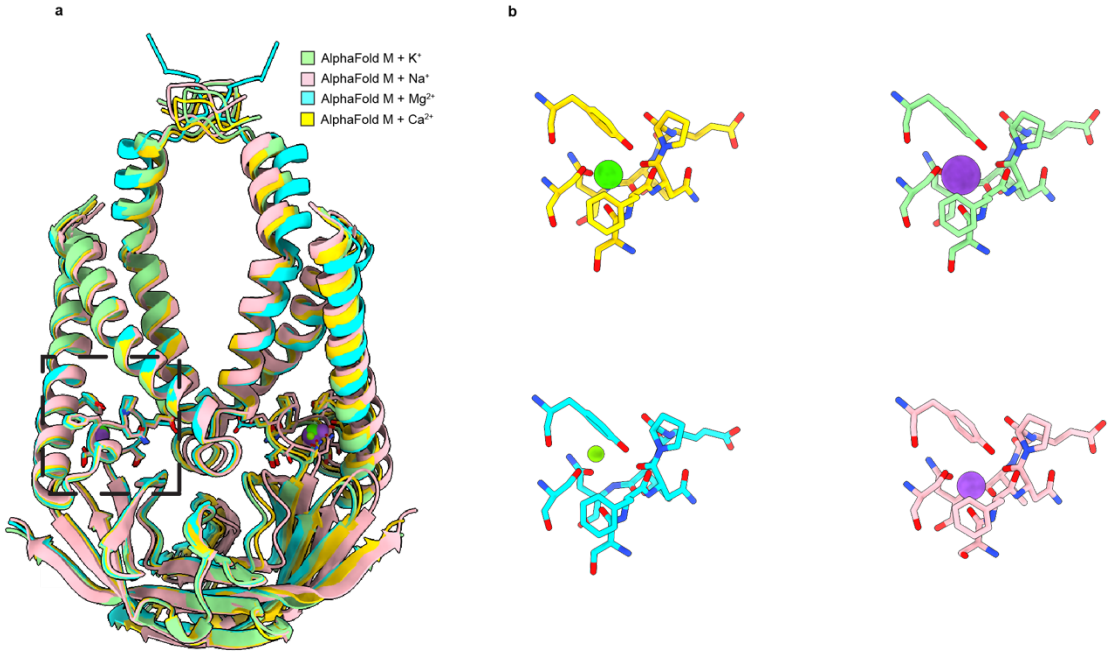
Supplemental Figure 4.2 Cryo-EM data, processing pipeline, and validation for M:scFab complex in LMNG/CHS/C1P. (a) Representative micrograph. **(b)** Selected 2D class averages and **(c)** Cryo-EM data processing pipeline. **(d)** Angular viewing distribution of refined particles and **(e)** Fourier shell correlation curves for final map with resolution cut-off at FSC = 0.5 and 0.143 indicated. **(f)** Local resolution for final unsharpened map with the region where C1P ligands were modeled inset.



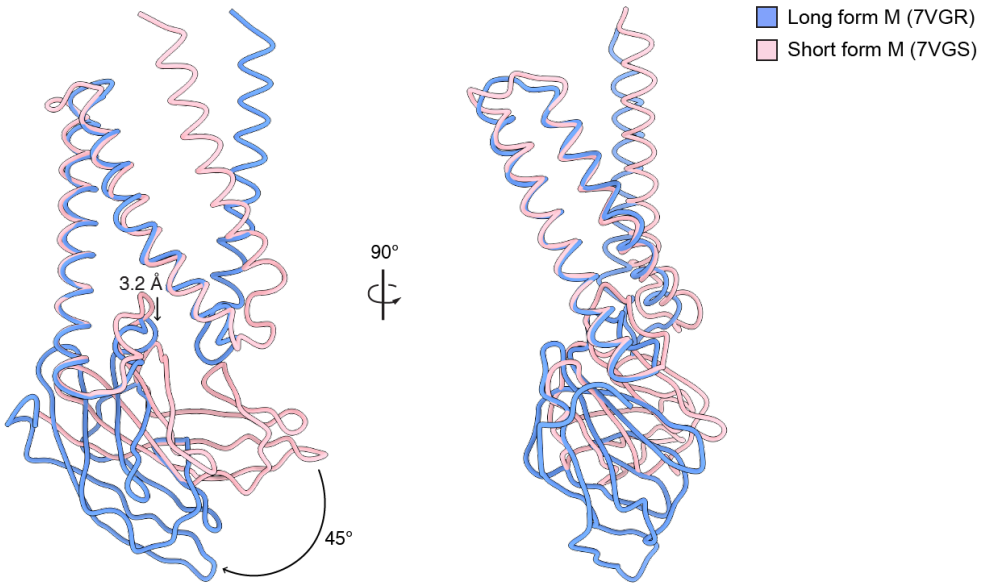
Supplemental Figure 4.3 Cryo-EM data, processing pipeline, and validation for M:IcFab complex in LMNG/CHS/C1P. (a) Representative micrograph. **(b)** Selected 2D class averages and **(c)** Cryo-EM data processing pipeline. **(d)** Angular viewing distribution of refined particles and **(e)** Fourier shell correlation curves for final map with resolution cut-off at FSC = 0.5 and 0.143 indicated. **(f)** Local resolution for final unsharpened map.



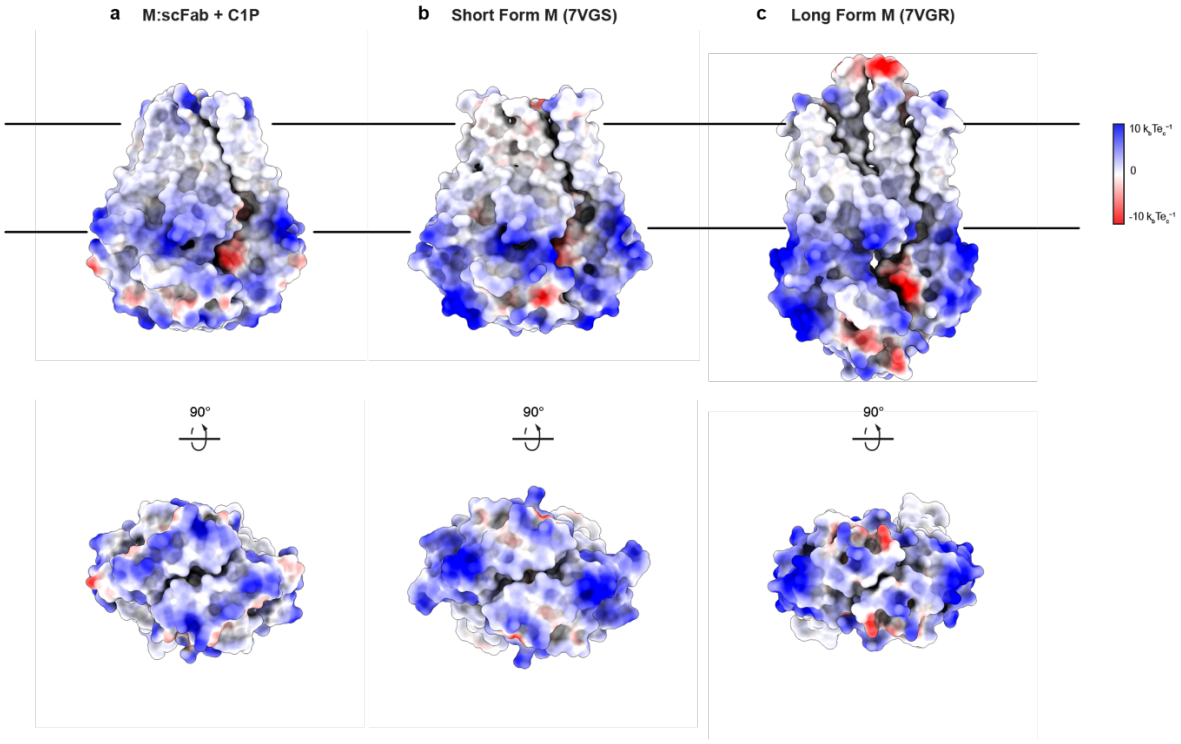
Supplemental Figure 4.4 Comparison of atomic displacement parameters (B-factors) for +/- C1P short conformation models. Models are colored from low B-factor in blue to high B-factor in red.



Supplemental Figure 4.5 AlphaFold predicted models of wild-type M dimers plus ions. (a) Overlay of M dimer models generated by AlphaFold3 with 2 K^+, Na^+, Mg^{2+}, or Ca^{2+} ion ligands (colored in green, pink, cyan, and gold, respectively) **(b)** Close-up of ion binding site from one subunit for each ion colored as in **(a)**.



Supplemental Figure 4.6 Overlay of short form M and long form M monomer. Long Form is in blue and short form in pink. Downward translation of the hinge domain and tilt of the CTD in transition from short to long are indicated.



Supplemental Figure 4.7 Differences in electrostatic surface potential between conformations. Side and bottom views of **(a)** M:scFab + C1P, **(b)** 7VGS, and **(c)** 7VGR colored according to electrostatic surface potential from electronegative (red) to electropositive (blue).

Chapter Five

Final conclusions and outlook

My work presented in this thesis began in the summer of 2020 with the overarching goal of understanding the structure of SARS-CoV-2 M and its function in the viral life cycle. Towards this end, I developed the first published routine for the expression and purification of a coronavirus M protein and determined the structure of M in lipid nanodiscs using single-particle cryo-EM (Chapter 2). Within months of submitting this work for review, a complementary study was published by Zhang et al. that included structures of SARS-CoV-2 M captured in two distinct conformations by fragment antibodies (Fabs) bound to the C-terminal domain (CTD), one more compact (M_{short}) and a more elongated, narrower form (M_{long})[80]. While I saw no evidence of M conformational heterogeneity in my purification process or cryo-EM data, the fact that the Fabs were able to recognize both conformations in detergent suggests that the two states exist in dynamic equilibrium. Our apo M structure in nanodiscs closely matched their M_{short} conformation, and molecular dynamic simulations performed by our collaborator suggests that this structure is stable in a simple lipid bilayer. Evidence shows M is expressed in infected cells before S or E, and that M directs incorporation of all structural proteins into newly forming virions through protein-protein interactions[59,148]. Our apo M cryo-EM and molecular dynamics data suggest that M_{short} is likely the conformation present as transcription is occurring, and thus is likely the conformation that initially recognizes S, E, and N.

In chapter 3, I utilized the conformationally selective Fabs reported in Zhang et al. to determine the structure of M_{long} in complex with a neutralizing Fab (SARS37) isolated from convalescent patient sera. Using biolayer-interferometry, I found that the SARS37 Fab specifically recognizes M_{long}, with over 40-fold greater affinity than measured for M_{short}. Our structure of SARS37:M_{long} showed that SARS37 has a long, flexible CDR3 loop that inserts into the membrane and fits in a cleft between transmembrane helices (TM) 1-2 of one M_{long} subunit. The specificity of the binding interface is provided through electrostatic and hydrophobic interactions between membrane proximal CDR3 residues, and the N-termini from both M_{long} subunits. In agreement with the weaker SARS37 binding to M_{short} observed in our BLI data, the cryo-EM data I collected on the SARS37:M_{short} complex showed a comparatively high degree of flexibility and heterogeneity.

The N-terminus of M has been shown to be highly immunogenic for a variety of coronavirus strains and host species, however this is the first ever recorded example of a coronavirus neutralizing antibody (nAb) that targets the M protein. These data also provide the first direct evidence that M_{long} is present in infectious virions and can specifically be targeted to neutralize infection. As M has never been implicated as playing a role in viral entry, it is very much an open question as to exactly how this antibody can be neutralizing. The following mechanisms could provide an explanation and warrant

further investigation[149]: 1.) In enveloped viruses, nAbs often bind to a target, such as those that bind to the receptor binding domain of the coronavirus S protein, that physically block attachment to host cell receptors. As the M protein is highly abundant in the viral envelope, this effect could potentially be achieved by many nAbs binding on the virion surface and physically blocking S protein attachment to its host receptor. 2.) nAb binding leading to virion aggregation that disrupts attachment to host cells would be a potential simple explanation for anti-M neutralization. 3.) nAb binding can prevent a conformational change that is required as part of the infection cycle or promote conformational changes that render the target nonfunctional. Since SARS37 Fab specifically recognizes M_{long}, this mechanism is possibly only if the conformational change from M_{long} to M_{short} is required for viral entry.

In the final story described in chapter 4, I sought to investigate specific M protein-lipid interactions with ceramide-1-phosphate (C1P) using cryo-EM. C1P is produced through the activity of a ceramide kinase in the Golgi, where coronavirus budding occurs. Lipid binding assays performed by our collaborator provided evidence for a specific M-C1P interaction. I found that C1P binds to M_{short} in a cleft between TM2-3, with the phosphate headgroup coordinated in an electropositive pocket of the CTD. Comparing our C1P-bound M_{short} map and model to the previously published M_{short} data from Zhang et al., I found that C1P binding appears to induce an overall stabilizing effect, compacting the transmembrane domain and revealing a novel K^+ binding site bridging TMs and the CTD. I also observe a series of side-chain rotamer differences between the two maps that are distributed across the hinge and CTD, including the CTD dimer interface. I used AlphaFold3[71] to model mutations at these sites and found that their structures were predicted to fit in a continuum between the published M_{long} and M_{short} conformations. All the residues are highly conserved amongst sarbecoviruses, suggesting that they are under purifying selection, perhaps due to their importance in M conformational dynamics.

In contrast, I neither saw evidence of C1P binding to M_{long}, nor did I observe any stabilizing effect from the addition of C1P compared to published M_{long} data. Furthermore, the C1P coordinating site I observed in the Mshort:C1P dataset is disrupted by the conformational change to M_{long}. Implying that C1P binds to and stabilizes M_{short} but not M_{long}. Recently, a preprint was posted that characterized the binding of an antiviral small molecule to the transmembrane domain of M[42]. The drug binding site overlaps significantly with the M_{short} C1P and K^+ binding sites I have characterized here. They found that the drug prevents viral egress and based on their structure propose that it works by preventing the conformational switch from M_{short} to M_{long}. Taken into consideration with my M_{short} C1P data, this suggests that C1P may regulate the timing of this conversion by stabilizing the short conformation until a certain point in virion assembly.

The work presented here enhances our understanding of coronavirus M protein structure and conformational dynamics, but many open questions about M function

remain unanswered. As a protein biochemist, I am more than anything interested in understanding how proteins work. As such I would love to move beyond the single-particle methods I have used thus far into more complex systems where I can look at the M in VLPs and virions to understand how M interacts with its many binding partners. The M protein is highly conserved and indispensable for coronavirus replication and understanding the entire picture of its role in virion assembly and budding will only expand our ability to target it for preventing current and future pandemics.

References

1. Schalk, A. F. & Hawn, M. C. An apparently new respiratory disease of baby chicks. *J. Am. Med. Assoc.* **78**, 413–422 (1931).

2. Beach, J. R. & Schalm, O. W. A Filterable Virus, Distinct from that of Laryngotracheitis, the Cause of a Respiratory Disease of Chicks. *Poult. Sci.* **15**, 199–206 (1936).

3. Vlasova, A. N., Kenney, S. P., Jung, K., Wang, Q. & Saif, L. J. Deltacoronavirus Evolution and Transmission: Current Scenario and Evolutionary Perspectives. *Front. Vet. Sci.* **7**, (2021).

4. Fan, Y., Zhao, K., Shi, Z.-L. & Zhou, P. Bat Coronaviruses in China. *Viruses* **11**, 210 (2019).

5. Miller, A. K. *et al.* Slippery when wet: cross-species transmission of divergent coronaviruses in bony and jawless fish and the evolutionary history of the Coronaviridae. *Virus Evol.* **7**, (2021).

6. Wood, E. An apparently new syndrome of porcine epidemic diarrhoea. *Vet. Rec.* **100**, 243–244 (1977).

7. Cheever, F. S., Daniels, J. B., Pappenheimer, A. M. & Bailey, O. T. A MURINE VIRUS (JHM) CAUSING DISSEMINATED ENCEPHALOMYELITIS WITH EXTENSIVE DESTRUCTION OF MYELIN. *J. Exp. Med.* **90**, 181–194 (1949).

8. Woo, P. C. Y. *et al.* Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *J. Virol.* **86**, 3995–4008 (2012).

9. Rexhepaj, M. *et al.* Broadly neutralizing antibodies against emerging delta-coronaviruses. *bioRxiv* 2024.03.27.586411 (2024) doi:10.1101/2024.03.27.586411.

10. Lednicky, J. A. *et al.* Independent infections of porcine deltacoronavirus among Haitian children. *Nature* **600**, 133–137 (2021).

11. Wang, L., Byrum, B. & Zhang, Y. New Variant of Porcine Epidemic Diarrhea Virus, United States, 2014. *Emerg. Infect. Dis.* **20**, 917–919 (2014).

12. Tyrrell, D. A. J. & Bynoe, M. L. Cultivation of a Novel Type of Common-cold Virus

in Organ Cultures. *BMJ* **1**, 1467–1470 (1965).

13. Tang, G., Liu, Z. & Chen, D. Human coronaviruses: Origin, host and receptor. *J. Clin. Virol.* **155**, 105246 (2022).

14. Zhong, N. *et al.* Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353–1358 (2003).

15. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).

16. SØRENSEN, M. D. *et al.* Severe Acute Respiratory Syndrome (SARS). *Ann. N. Y. Acad. Sci.* **1067**, 500–505 (2006).

17. Donnelly, C. A. *et al.* Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet (London, England)* **361**, 1761–6 (2003).

18. Anderson, R. M. *et al.* Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **359**, 1091–1105 (2004).

19. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).

20. Alsaad, K. O. *et al.* Histopathology of Middle East respiratory syndrome coronovirus infection – clinicopathological and ultrastructural study. *Histopathology* **72**, 516–524 (2018).

21. Pustake, M., Tambolkar, I., Giri, P. & Gandhi, C. SARS, MERS and CoVID-19: An overview and comparison of clinical, laboratory and radiological features. *J. Fam. Med. Prim. Care* **11**, 10 (2022).

22. Abdelazim, M. *et al.* A longitudinal study of Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels. *BMC Vet. Res.* **19**, 228 (2023).

23. Dudas, G., Carvalho, L. M., Rambaut, A. & Bedford, T. MERS-CoV spillover at the camel-human interface. *Elife* **7**, (2018).

24. Msemburi, W. *et al.* The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* **613**, 130–137 (2023).

25. Coronaviridae Study Group of the International Committee on Taxonomy of

Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).

26. Tsai, P.-H. *et al.* Clinical manifestation and disease progression in COVID-19 infection. *J. Chinese Med. Assoc.* **84**, 3–8 (2021).

27. Ledford, H. How common is long COVID? Why studies give different answers. *Nature* **606**, 852–853 (2022).

28. Omotayo, A. O. & Ogunniyi, A. I. COVID-19 induced shocks and its implications for human capital development. *Int. J. Equity Health* **23**, 30 (2024).

29. Kakaei, H., Nourmoradi, H., Bakhtiyari, S., Jalilian, M. & Mirzaei, A. Effect of COVID-19 on food security, hunger, and food crisis. in *COVID-19 and the Sustainable Development Goals* 3–29 (Elsevier, 2022). doi:10.1016/B978-0-323-91307-2.00005-5.

30. Masters, P. S. The Molecular Biology of Coronaviruses. in 193–292 (2006). doi:10.1016/S0065-3527(06)66005-3.

31. Neuman, B. W. *et al.* Supramolecular Architecture of Severe Acute Respiratory Syndrome Coronavirus Revealed by Electron Cryomicroscopy. *J. Virol.* **80**, 7918–7928 (2006).

32. Kanso, M. A. *et al.* Coronavirus pleomorphism. *Phys. Fluids* **34**, (2022).

33. Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174**, 11–22 (2011).

34. Kuo, L., Hurst-Hess, K. R., Koetzner, C. A. & Masters, P. S. Analyses of Coronavirus Assembly Interactions with Interspecies Membraneand Nucleocapsid Protein Chimeras. *J. Virol.* **90**, 4357–4368 (2016).

35. de Haan, C. A. M., Vennema, H. & Rottier, P. J. M. Coronavirus Envelope Assembly is Sensitive to Changes in the Terminal Regions of the Viral M Protein. in 367–375 (1998). doi:10.1007/978-1-4615-5331-1_48.

36. Kuo, L., Koetzner, C. A. & Masters, P. S. A key role for the carboxy-terminal tail of the murine coronavirusnucleocapsid protein in coordination of genome packaging. *Virology* **494**, 100–107 (2016).

37. Boson, B. *et al.* The SARS-CoV-2 envelope and membrane proteins modulate maturation andretention of the spike protein, allowing assembly of virus-likeparticles. *J. Biol. Chem.* **296**, (2021).

38. Xiang, Q. *et al.* SARS-CoV-2 Membrane protein regulates the function of Spike by inhibiting its plasma membrane localization and enzymatic activity of Furin. *Microbiol. Res.* **282**, 127659 (2024).

39. Cattin-Ortolá, J. *et al.* Sequences in the cytoplasmic tail of SARS-CoV-2 Spike facilitate expression at the cell surface and syncytia formation. *Nat. Commun.* **12**, 5333 (2021).

40. Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174**, 11–22 (2011).

41. Chua, E. Y. D. *et al.* Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. *Annu. Rev. Biochem.* **91**, 1–32 (2022).

42. Loock, M. Van *et al.* A small-molecule SARS-CoV-2 inhibitor targeting the membrane protein. *Res. Sq.* (2024) doi:10.21203/rs.3.rs-3975125/v1.

43. Plescia, C. B. *et al.* SARS-CoV-2 viral budding and entry can be modeled using BSL-2 levelvirus-like particles. *J. Biol. Chem.* **296**, (2021).

44. Siu, Y. L. *et al.* The M, E, and N Structural Proteins of the Severe Acute RespiratorySyndrome Coronavirus Are Required for Efficient Assembly, Trafficking,and Release of Virus-Like Particles. *J. Virol.* **82**, 11318–11330 (2008).

45. Wang, C. *et al.* MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques. *Oncotarget* **8**, 12686–12694 (2017).

46. Boscarino, J. A., Logan, H. L., Lacny, J. J. & Gallagher, T. M. Envelope Protein Palmitoylations Are Crucial for Murine Coronavirus Assembly. *J. Virol.* **82**, 2989–2999 (2008).

47. Sturman, L. S., Holmes, K. V & Behnke, J. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J. Virol.* **33**, 449–462 (1980).

48. Armstrong, J., Niemann, H., Smeekens, S., Rottier, P. & Warren, G. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature* **308**, 751–752 (1984).

49. Yu, A. *et al.* A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.* **120**, 1097–1104 (2021).

50. Finkel, Y. *et al.* The coding capacity of SARS-CoV-2. *Nature* **589**, 125-U254 (2021).

51. Godet, M., L'Haridon, R., Vautherot, J.-F. & Laude, H. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* **188**, 666–675 (1992).

52. He, Y., Zhou, Y., Siddiqui, P., Niu, J. & Jiang, S. Identification of Immunodominant Epitopes on the Membrane Protein of the Severe Acute Respiratory Syndrome-Associated Coronavirus. *J. Clin. Microbiol.* **43**, 3718–3726 (2005).

53. Hotop, S.-K. *et al.* Peptide microarrays coupled to machine learning reveal individualepitopes from human antibody responses with neutralizing capabilitiesagainst SARS-CoV-2. *Emerg. Microbes Infect.* **11**, 1037–1048 (2022).

54. Heffron, A. S. *et al.* The landscape of antibody binding in SARS-CoV-2 infection. *PLOS Biol.* **19**, (2021).

55. Martin, S. *et al.* SARS-CoV-2 integral membrane proteins shape the serological responses ofpatients with COVID-19. *ISCIENCE* **24**, (2021).

56. Joerrissen, P. *et al.* Antibody Response to SARS-CoV-2 Membrane Protein in Patients of theAcute and Convalescent Phase of COVID-19. *Front. Immunol.* **12**, (2021).

57. de Haan, C. A. M., Smeets, M., Vernooij, F., Vennema, H. & Rottier, P. J. M. Mapping of the Coronavirus Membrane Protein Domains Involved in Interaction with the Spike Protein. *J. Virol.* **73**, 7441–7452 (1999).

58. Schmidt, A., Wolff, M. H. & Weber, O. *Coronaviruses with Special Emphasis on First Insights Concerning SARS*. (Birkhäuser Basel, 2005). doi:10.1007/b137625.

59. Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly andmorphology. *J. Struct. Biol.* **174**, 11–22 (2011).

60. Lim, K. P. & Liu, D. X. The Missing Link in Coronavirus Assembly. *J. Biol. Chem.* **276**, 17515–17523 (2001).

61. Perrier, A. *et al.* The C-terminal domain of the MERS coronavirus M protein contains a*trans*-Golgi network localization signal. *J. Biol. Chem.* **294**, 14406–14421 (2019).

62. Vennema, H. *et al.* Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15**, 2020–8 (1996).

63. Hsieh, P.-K. *et al.* Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. *J.*

Virol. **79**, 13848–13855 (2005).

64. Xu, R., Shi, M., Li, J., Song, P. & Li, N. Construction of SARS-CoV-2 Virus-Like Particles by Mammalian ExpressionSystem. *Front. Bioeng. Biotechnol.* **8**, (2020).

65. Fu, Y.-Z. *et al.* SARS-CoV-2 membrane glycoprotein M antagonizes the MAVS-mediated innateantiviral response. *Cell. Mol. Immunol.* **18**, 613–620 (2021).

66. Sui, L. *et al.* SARS-CoV-2 Membrane Protein Inhibits Type I Interferon ProductionThrough Ubiquitin-Mediated Degradation of TBK1. *Front. Immunol.* **12**, (2021).

67. Tan, Y. *et al.* Unification and extensive diversification of M/Orf3-related ion channel proteins in coronaviruses and other nidoviruses. *Virus Evol.* **7**, 14 (2021).

68. Pezeshkian, W. *et al.* Molecular architecture and dynamics of SARS-CoV-2 envelope byintegrative modeling. *bioRxiv* (2022) doi:10.1101/2021.09.15.459697.

69. Kern, D. M. *et al.* Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat. Struct. Mol. Biol.* **28**, 573+ (2021).

70. Heo, L. & Feig, M. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.03.25.008904.

71. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583+ (2021).

72. Holm, L. Using Dali for Protein Structure Comparison. in *STRUCTURAL BIOINFORMATICS: METHODS AND PROTOCOLS* (ed. Gaspari, Z.) vol. 2112 29–42 (HUMANA PRESS INC, 2020).

73. Xue, Y., Ward, J. M., Yuwen, T. & Podkorytov Ivan S.and Skrynnikov, N. R. Microsecond Time-Scale Conformational Exchange in Proteins: Using LongMolecular Dynamics Trajectory To Simulate NMR Relaxation Dispersion Data. *J. Am. Chem. Soc.* **134**, 2555–2562 (2012).

74. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure andfunction. *Curr. Opin. Struct. Biol.* **19**, 120–127 (2009).

75. Don, C. G. & Smiesko, M. Microsecond MD simulations of human CYP2D6 wild-type and five allelicvariants reveal mechanistic insights on the function. *PLoS One* **13**, (2018).

76. Monje-Galvan, V. & Voth, G. A. Molecular interactions of the M and E integral

membrane proteins ofSARS-CoV-2. *FARADAY Discuss.* **232**, 49–67 (2021).

77. Brandner, A., De Vecchis, D., Baaden, M., Cohen, M. M. & Taly, A. Structural dataset from microsecond-long simulations of yeast mitofusinFzo1 in the context of membrane docking. *DATA Br.* **26**, (2019).

78. Nury, H. *et al.* One-microsecond molecular dynamics simulation of channel gating in anicotinic receptor homologue. *Proc. Natl. Acad. Sci. UNITED STATES OFAMERICA* **107**, 6275–6280 (2010).

79. Monje-Galvan, V. & Voth, G. A. Binding mechanism of the matrix domain of HIV-1 gag on lipid membranes. *Elife* **9**, (2020).

80. Zhang, Z. *et al.* Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nat. Commun.* **13**, (2022).

81. Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phaseseparates with RNA. *Nat. Commun.* **12**, (2021).

82. Lu, S. *et al.* The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusivecondensates with RNA and the membrane-associated M protein. *Nat. Commun.* **12**, (2021).

83. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *BIOINFORMATICS* **34**, 4121–4123 (2018).

84. Cagliani, R., Forni, D., Clerici, M. & Sironi, M. Computational Inference of Selection Underlying the Evolution of theNovel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. *J. Virol.* **94**, (2020).

85. Kern, D. M. & Brohawn, S. G. SARS-CoV-2 3a expression, purification, and reconstitution into lipidnanodiscs. in *ION CHANNELS: CHANNEL PRODUCTION AND OPTICAL METHODS* (eds. Minor, D. L. & Colecraft, H. M.) vol. 653 207–235 (ELSEVIER ACADEMIC PRESS INC, 2021).

86. Ritchie, T. K. *et al.* RECONSTITUTION OF MEMBRANE PROTEINS IN PHOSPHOLIPID BILAYER NANODISCS. in *METHODS IN ENZYMOLOGY; LIPOSOMES, PT F* (ed. Duzgunes, N.) vol. 464 211–231 (ELSEVIER ACADEMIC PRESS INC, 2009).

87. Guo, H. *et al.* Electron-event representation data enable efficient cryoEM file storagewith full preservation of spatial and temporal resolution. *IUCRJ* **7**, 860–869 (2020).

88. Zivanov, J. *et al.* A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11**, (2022).

89. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

90. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

91. 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).

92. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking incryo-electron micrographs. *Nat. Methods* **16**, 1153+ (2019).

93. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *ACTA Crystallogr. Sect. D-BIOLOGICAL Crystallogr.* **66**, 486–501 (2010).

94. Afonine, P. V *et al.* Real-space refinement in *PHENIX* for cryo-EM and crystallography. *ACTA Crystallogr. Sect. D-STRUCTURAL Biol.* **74**, 531–544 (2018).

95. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atomstructure validation. *PROTEIN Sci.* **27**, 293–315 (2018).

96. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).

97. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *PROTEIN Sci.* **27**, 14–25 (2018).

98. Jo, S., Kim, T., Iyer, V. G. & Im, W. Software news and updates - CHARNIM-GUI: A web-based grraphical userinterface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

99. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

100. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, andCHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).

101. Jo, S., Kim, T. & Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2**, e880 (2007).

102. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

103. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsicallydisordered proteins. *Nat. Methods* **14**, 71–73 (2017).

104. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application toYeast Membranes. *Biophys. J.* **97**, 50–58 (2009).

105. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

106. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).

107. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

108. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).

109. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

110. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

111. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

112. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).

113. Michaud-Agrawal, N., Denning, E. J. & Woolf Thomas B.and Beckstein, O. Software News and Updates MDAnalysis: A Toolkit for the Analysis ofMolecular Dynamics Simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

114. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

115. World Health Organization (WHO) Coronavirus (COVD-19) Dashboard. https://covid19.who.int/ (2024).

116. Wünsch, K. *et al.* COVID-19 in Elderly, Immunocompromised or Diabetic Patients—From Immune Monitoring to Clinical Management in the Hospital.

Viruses **14**, 746 (2022).

117. Zhao, X. *et al.* Potent and broadly neutralizing antibodies against sarbecoviruses induced by sequential COVID-19 vaccination. *Cell Discov.* **10**, 14 (2024).

118. EUA of Bamlanivimab and Etesevimab. https://www.fda.gov/media/145802/download (2021).

119. EUA of Sotrovimab. https://www.fda.gov/media/149534/download (2022).

120. EUA of Bebtelovimab. https://www.fda.gov/media/156152/download (2022).

121. EUA of Evusheld. https://www.fda.gov/media/154701/download (2022).

122. Xia, S. *et al.* A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci. Adv.* **5**, (2019).

123. Jiang, S., Lu, L., Liu, Q., Xu, W. & Du, L. Receptor-binding domains of spike proteins of emerging or re-emerging viruses as targets for development of antiviral vaccines. *Emerg. Microbes Infect.* **1**, e13 (2012).

124. Olukitibi, T. A. *et al.* Significance of Conserved Regions in Coronavirus Spike Protein for Developing a Novel Vaccine against SARS-CoV-2 Infection. *Vaccines* **11**, (2023).

125. Li, G., Hilgenfeld, R., Whitley, R. & De Clercq, E. Therapeutic strategies for COVID-19: progress and lessons learned. *Nat. Rev. Drug Discov.* **22**, 449–475 (2023).

126. Dolan, K. A. *et al.* Structure of SARS-CoV-2 M protein in lipid nanodiscs. *Elife* **11**, (2022).

127. Williams, D. M. *et al.* Establishing SARS-CoV-2 membrane protein-specific antibodies as a valuable serological target via high-content microscopy. *iScience* **26**, 107056 (2023).

128. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).

129. Navaratnarajah, C. K., Warrier, R. & Kuhn, R. J. Assembly of Viruses: Enveloped Particles. in *Encyclopedia of Virology* 193–200 (Elsevier, 2008). doi:10.1016/B978-012374410-4.00667-1.

130. Mazzon, M. & Mercer, J. Lipid interactions during virus entry and infection. *Cell. Microbiol.* **16**, 1493–1502 (2014).

131. Meisen, I. *et al.* The human H3N2 influenza viruses A/Victoria/3/75 and A/Hiroshima/52/2005 preferentially bind to α2-3-sialylated monosialogangliosides with fucosylated poly-N-acetyllactosaminyl chains. *Glycobiology* **22**, 1055–1076 (2012).

132. Roth, S. L. & Whittaker, G. R. Promotion of vesicular stomatitis virus fusion by the endosome-specific phospholipid bis(monoacylglycero)phosphate (BMP). *FEBS Lett.* **585**, 865–869 (2011).

133. Niort, K. *et al.* Cholesterol and Ceramide Facilitate Membrane Fusion Mediated by the Fusion Peptide of the SARS-CoV-2 Spike Protein. *ACS Omega* **8**, 32729–32739 (2023).

134. Roingeard, P. *et al.* The double-membrane vesicle (DMV): a virus-induced organelle dedicated to the replication of SARS-CoV-2 and other positive-sense single-stranded RNA viruses. *Cell. Mol. Life Sci.* **79**, 425 (2022).

135. Tabata, K. *et al.* Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2 replication organelle formation. *Nat. Commun.* **12**, 7276 (2021).

136. Motsa, B. B. & Stahelin, R. V. Lipid–protein interactions in virus assembly and budding from the host cell plasma membrane. *Biochem. Soc. Trans.* **49**, 1633–1641 (2021).

137. Beckmann, N. & Becker, K. A. Ceramide and Related Molecules in Viral Infections. *Int. J. Mol. Sci.* **22**, 5676 (2021).

138. van der Meer-Janssen, Y. P. M., van Galen, J., Batenburg, J. J. & Helms, J. B. Lipids in host–pathogen interactions: Pathogens exploit the complexity of the host cell lipidome. *Prog. Lipid Res.* **49**, 1–26 (2010).

139. Baglivo, M. *et al.* Natural small molecules as inhibitors of coronavirus lipid-dependent attachment to host cells: a possible strategy for reducing SARS-COV-2 infectivity? *Acta Biomed.* **91**, 161–164 (2020).

140. Ding, C., Chen, Y., Miao, G. & Qi, Z. Research Advances on the Role of Lipids in the Life Cycle of Human Coronaviruses. *Microorganisms* **12**, 63 (2023).

141. Wong, F. & Anderson, R. Involvement of Lipids in Membrane Binding of Mouse Hepatitis Virus Nucleocapsid Protein. in 203–208 (1994). doi:10.1007/978-1-4615-2996-5_32.

142. Anderson, R. & Wong, F. Membrane and Phospholipid Binding by Murine Coronaviral Nucleocapsid N Protein. *Virology* **194**, 224–232 (1993).

143. Dutta, M., Su, Y., Voth, G. A. & Stahelin, R. V. The SARS-CoV-2 nucleoprotein

associates with anionic lipid membranes. *bioRxiv Prepr. Serv. Biol.* (2023) doi:10.1101/2023.09.15.557899.

144. Ansell, T. B. *et al.* LipIDens: simulation assisted interpretation of lipid densities in cryo-EM structures of membrane proteins. *Nat. Commun.* **14**, 7774 (2023).

145. 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).

146. Shub, L., Liu, W., Skiniotis, G., Keiser, M. J. & Robertson, M. J. Metric Ion Classification (MIC): A deep learning tool for assigning ions and waters in cryo-EM and x-ray crystallography structures. *bioRxiv* 2024.03.18.585639 (2024) doi:10.1101/2024.03.18.585639.

147. Gucwa, M. *et al.* CMM-An enhanced platform for interactive validation of metal binding sites. *Protein Sci.* **32**, e4525 (2023).

148. Scherer, K. M. *et al.* SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. *Sci. Adv.* **8**, eabl4895 (2022).

149. Burton, D. R. Antiviral neutralizing antibodies: from in vitro to in vivo activity. *Nat. Rev. Immunol.* **23**, 720–734 (2023).