bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Crystal structure of the membrane (M) protein from a SARS-COV-2-related coronavirus

Xiaodong Wang[1], Yuwei Yang[1], Ziyi Sun[1]*, Xiaoming Zhou[1]*

[1]Department of Integrated Traditional Chinese and Western Medicine, Rare Diseases Center, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China

*To whom correspondence should be addressed:

Ziyi Sun, PhD, 17 Renmin Road South 3rd Section, Sichuan University, Chengdu, Sichuan 610041, China; Phone: +86 (28) 8540 1152; Email: ziyi.sun@scu.edu.cn

Xiaoming Zhou, PhD, 17 Renmin Road South 3rd Section, Sichuan University, Chengdu, Sichuan 610041, China; Phone: +86 (28) 8540 1152; Email: x.zhou@scu.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The membrane (M) protein is the most abundant structural protein of coronaviruses including SARS-COV-2 and plays a central role in virus assembly through its interaction with various partner proteins. However, mechanistic details about how M protein interacts with others remain elusive due to lack of high-resolution structures. Here, we present the first crystal structure of a coronavirus M protein from *Pipistrellus* bat coronavirus HKU5 (batCOV5-M), which is closely related to SARS-COV-2 M protein. Furthermore, an interaction analysis indicates that the carboxy-terminus of the batCOV5 nucleocapsid (N) protein mediates its interaction with batCOV5-M. Combined with a computational docking analysis an M-N interaction model is proposed, providing insight into the mechanism of M protein-mediated protein interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Main Text

The virion of coronaviruses including SARS-COV-2 contains four structural proteins: S (spike), M (membrane), E (envelope) and N (nucleocapsid) proteins[1-4]. Among them, M protein is the most abundant protein component and plays a central role in virus assembly mainly as a scaffolding center[1,5-7]. In coronaviruses, M protein is the only multi-spanning transmembrane structural protein and has been shown to interact with various partner proteins, including all viral structural proteins (S, M, E, N)[1,6,8-15] and some host factors involved in the interferon signaling pathway such as MAVS[16-19]. However, how M protein interacts with its partners remains largely unknown due to lack of high-resolution structures of M protein[20-23]. In this study, we aim to determine the crystal structure of a coronavirus M protein.

To obtain a suitable M protein for structural study, we screened expression of M protein genes from eight betacoronavirus strains[1,24], including SARS-COV-2 (SARS2), SARS-COV (SARS) and MERS-COV (MERS). Among them, the M protein from *Pipistrellus* bat coronavirus HKU5 (batCOV5-M) displayed the best expression and biochemical properties, and was selected for further structural study. The batCOV5-M protein is closely related to SARS2-M, SARS-M and MERS-M proteins, sharing 37% sequence identity and 68% sequence similarity among them (Supplementary Fig. 1).

After extensive crystallization trials and refinement, batCOV5-M crystallized in lipidic cubic phase (LCP) and the crystals diffracted to ~3 Å resolution. However, molecular replacement[25] using an AlphaFold[26]-predicted SARS2-M model (termed SARS2-M_{AF})[21] failed to yield a valid solution, suggesting that SARS2-M_{AF} may vary significantly from the experimental M structure. Therefore, we sought to resolve the phase problem by experimental means, such as single-wavelength anomalous diffraction (SAD) using heavy atoms[27]. Since batCOV5-M crystals grew only in the LCP glass sandwich[28], which makes soaking the crystals with heavy atoms impossible, we attempted to co-crystallize batCOV5-M with heavy atoms, but unfortunately unsuccessful. Our next strategy was to crystallize the cytosolic carboxy-terminal domain (CTD) of batCOV5-M (batCOV5-M_{CTD}, Supplementary Fig. 1) and solve its structure, which may serve as a searching template for molecular replacement of the full-length batCOV5-M. In the end, by fusing the two halves of a split superfolder GFP[29,30] to the two termini of batCOV5-M_{CTD}, the resulting batCOV5-M_{CTD}-GFP was crystallized by the vapor-diffusion sitting-drop method and its structure solved to 3.42 Å (Supplementary Fig. 2a) by molecular replacement using the GFP structure (PDB: 2B3Q) as a searching model. Interestingly, four batCOV5-M_{CTD}-GFP molecules assemble into a tetramer in the structure with M_{CTD} domains swapping β strands with each other,

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which resembles a "butterfly" shape (Supplementary Fig. 2b). Three monomer conformations of M_{CTD} were then extracted from the tetramer: conformer 1 with the β sandwich split into three parts (Supplementary Fig. 2c), conformer 2 with two parts (Supplementary Fig. 2d), and conformer 3 as an assembled M_{CTD} (termed batCOV5-$M_{CTD\text{-}xtal}$, Fig. 1a and Supplementary Fig. 2e). Though it is tempting to postulate that domain swapping of CTD may play a role in regulating the M-M protein interaction[1], further investigation is required. Meanwhile, we tried to solve the full-length batCOV5-M structure by molecular replacement using the three CTD conformations as templates, and eventually succeeded with the batCOV5-$M_{CTD\text{-}xtal}$ structure. The final batCOV5-M model was determined to 3.21 Å (termed batCOV5-M_{xtal}, Fig. 1b) and contained residues 13-199 of batCOV5-M, and the amino-terminus (12 residues) and the carboxy-terminus (21 residues) were not modeled due to weak electron densities.

The batCOV5-M_{xtal} structure is organized as a homodimer, with each protomer containing a short amino-terminus, a three-helix transmembrane domain (TMD) and a cytosolic CTD consisting of an eight-strand β-sandwich (Fig. 1b). The overall shape of the batCOV5-M_{xtal} dimer looks like a pair of pincers, with TMD as the "handle" and CTD as the "head" of the pincers (Fig. 1b). The two protomers form extensive TMD-TMD and CTD-CTD contacts, suggesting that the batCOV5-M_{xtal} dimer is a stable form (Fig. 1c). Not surprisingly, superposition of SARS2-M_{AF} onto one protomer of the batCOV5-M_{xtal} structure yielded an all-C_α RMSD of 4.57 Å with the major difference in TM1 (Fig. 1d). In the batCOV5-M_{xtal} structure, TM1 is placed almost in the sample plane as TM2/TM3, and the three helices of TMD resemble a capital letter "N", while in SARS2-M_{AF} the TMD helices form a more compact three-helix bundle. As a result, the two TM1 helices swap positions in the batCOV5-M_{xtal} dimer compared to SARS2-M_{AF}. Meanwhile, alignment between the batCOV5-$M_{CTD\text{-}xtal}$ structure and CTD of the batCOV5-M_{xtal} structure yielded an all-atom RMSD of only 0.54 Å, indicating that the individually expressed CTD of batCOV5-M protein preserves its structure (Fig. 1e). This result may explain why molecular replacement for batCOV5-M_{xtal} failed when using SARS2-M_{AF} as a template, but succeeded with the batCOV5-$M_{CTD\text{-}xtal}$ structure. Intriguingly, a structural comparison analysis using the Dali server[31] indicates the batCOV5-M_{xtal} structure shares fold with the cryo-EM structure of SARS-COV-2 ORF3a protein (termed SARS2-ORF3a_{EM}, PDB: 6XDC), which is a cation channel and also a homodimer[32]. Indeed, superposition of the two dimer structures yielded an all-C_α RMSD of 3.75 Å, suggesting that the overall folding of the two structures is alike (Fig. 1f). Nevertheless, few polar/charged residues in the batCOV5-M TMD (Supplementary Fig. 1) makes it less likely also a cation channel as ORF3a.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To validate the batCOV5-M_{xtal} structure and to study its function, we then investigated the interaction between M and N proteins (M-N interaction). The M-N interaction has been shown to facilitate virion assembly and budding of virus-like particles[6,7,13-15]. N protein is an RNA-binding protein containing two RNA-binding domains: an amino-terminal domain (N_{1b}) and a carboxy-terminal domain (N_{2b})[13-15] (Fig. 2a). An N_3 region at the end of the carboxy-terminus has been shown to interact with M protein[13-15]. Therefore, we generated four constructs for expressing N protein and its fragments: the full-length protein (residues 1-427, termed batCOV5-N_{FL}), the N_1 fragment (residues 1-190, termed batCOV5-N_1), the N_2 fragment (residues 191-390, termed batCOV5-N_2), and the N_3 fragment (residues 391-427, termed batCOV5-N_3) (Fig. 2a). In a pull-down analysis, purified batCOV5-N_{FL} and batCOV5-N_3 proteins, but not batCOV5-N_1 or batCOV5-N_2, displayed robust interactions with purified wild-type batCOV5-M protein (Fig. 2b, lane 4 and 10), consistent with previous reports[13-15]. It has also been suggested that the N_3 region may interact with M protein through electrostatic interactions[33-35]. Interestingly, batCOV5-N_3 contains six basic residues in its amino-terminal half (residues 391-410, termed batCOV5-N_{3N}) and five acidic residues in its carboxy-terminal half (residues 411-427, termed batCOV5-N_{3C}) (Fig. 2a). The pull-down analysis clearly showed that purified batCOV5-N_{3C}, but not batCOV5-N_{3N}, pulled M protein down (Fig. 2b, lane 14). Furthermore, single mutations at the five acidic residue positions (E415, D416, D419, E424 and E426) of batCOV5-N_3 decreased its binding affinity to batCOV5-M to varying extent in a microscale thermophoresis assay (Supplementary Table 1). This result is consistent with the notion that the negatively-charged batCOV5-N_{3C} interacts with the Lys/Arg/His-rich CTD of batCOV5-M (Supplementary Fig. 1), and charge interactions may play an important role in this process.

To further explore the binding mechanism of M and N_{3C} proteins, we created docking models using the batCOV5-M_{xtal} dimer structure and the batCOV5-N_{3C} sequence through the HDOCK web server[36,37], which predicts batCOV5-N_{3C} as a helix. Two plausible binding patterns between M and N_{3C} were observed (Fig. 2c) after manually excluding incorrectly placed N_{3C} helix (e.g. in TMD). In the first pattern, the N_{3C} helix binds at the interface between the CTD dimer and the TMD dimer, which form a groove nearly horizontal to the membrane plane. In the second pattern, the N_{3C} helix binds to a groove formed along the CTD dimer interface nearly vertical to the membrane plane. Both binding patterns require dimerization of batCOV5-M. Consistently, individually-expressed batCOV5-M_{CTD} does not bind batCOV5-N_{FL} (Fig. 2d), likely due to that batCOV5-M_{CTD} does not form the same dimer interface as the full-length batCOV5-M does (Fig. 2e). Nonetheless, which binding pattern may be correct? To gain more insights, the basic residues within the two putative binding grooves of batCOV5-M were selected for mutagenesis.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For the "horizontal groove", Arg42 and Lys172 were selected, whereas Arg131 and His155 were selected for both grooves as the two residues are located at the crossing of the two grooves (Fig. 2c, right panel). Although batCOV5-M mutants of R42A and R131A failed to express, the other two mutants (H155A and K172A) were expressed and showed apparently less binding to batCOV5-N_3 in pull-down assays (Fig. 2f), which favors the "horizontal groove" for N_{3C} binding. Interestingly, His155 is not charged under the pull-down assay condition (pH 7.5), suggesting that interactions other than charge interactions likely also contribute to M-N binding. To verify the H155A result, binding of batCOV5-M_{H155A} to batCOV5-N_3 was measured by microscale thermophoresis, which showed a decreased binding affinity (equilibrium dissociation constant K_d = 3.60 ± 0.17 μM) compared to the wild-type batCOV5-M protein (K_d = 0.77 ± 0.13 μM) (Supplementary Table 1), consistent with the pull-down result.

In this study, we report the first high-resolution crystal structure of M protein from a SARS-COV-2-related coronavirus batCOV5 and provide insight into the binding mechanism between M and N proteins. During preparation of this manuscript, Dolan and colleagues also reported a cryo-EM structure of M protein from SARS-COV-2 (termed SARS2-M_{EM}) in bioRxiv[38]. As expected, SARS2-M_{EM} is also a homodimer and shares a similar fold to the batCOV5-M_{xtal} structure (Fig. 3a). Intriguingly, superposition of SARS2-M_{EM} dimer onto batCOV5-M_{xtal} dimer yielded a large all-C_α RMSD of 8.25 Å, suggesting that the two structures are in different conformations. Alignment of the two dimer structures showed an apparent difference in shape, with batCOV5-M_{xtal} dimer long and slim (~80 Å x 45 Å) while SARS2-M_{EM} dimer short and fat (~65 Å x 55 Å) (Fig. 3a). This observation is consistent with the "long" and "compact" conformations suggested previously[9], which may be involved in M-N interaction and virus assembly. To further analyze the conformational change between SARS2-M_{EM} and batCOV5-M_{xtal}, we superimposed single protomers of the two structures, yielding an all-C_α RMSD of 5.14 Å, whereas alignment of single CTDs and single TMDs between the two structures yielded all-C_α RMSDs of 1.38 Å and 4.18 Å, respectively (Fig. 3b). Comparison of single protomers between SARS2-M_{EM} and batCOV5-M_{xtal} reveals that their CTDs are highly alike, but their TMDs differ significantly in TM1, which is much closer to TM2 in SARS2-M_{EM} than in batCOV5-M_{xtal} (Fig. 3b). As a result, using TM2/TM3 as anchors, TM1 swings ~40 degrees away from TM2 and CTD undergoes a rigid-body rotation of ~30 degrees away from TMD during a transition from the "compact" SARS2-M_{EM} to the "long" batCOV5-M_{xtal} (Fig. 3b). In the meantime, the two protomers move ~10 Å closer (Fig. 3a). In an analogy to the pincers, SARS2-M_{EM} dimer is like a pair of relaxed pincers, whereas batCOV5-M_{xtal} dimer is like the "handle" of the pincers is being clenched. It is also interesting to note that

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Trp24 of TM1 in batCOV5-M_{xtal} dimer form a π-cation-π interaction with an ammonium ion sitting between the two indole side groups (Fig. 3c), which may play a role in stabilizing the "long" conformation of M protein. Similarly, two tryptophan residues are also present in TM1 of SARS2-M (Trp20 and Trp31, Supplementary Fig. 1). It has been suggested that the "long" form, but not the "compact" form, of M protein binds N protein[9]. Therefore, it is desirable to determine an M-N complex structure in future studies to further elucidate the structure and function of M protein in coronaviruses.

References

1. Masters, P.S. The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292 (2006).
2. Gorkhali, R. et al. Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinform Biol Insights* **15**, 11779322211025876 (2021).
3. Arya, R. et al. Structural insights into SARS-CoV-2 proteins. *J Mol Biol* **433**, 166725 (2021).
4. Romano, M., Ruggiero, A., Squeglia, F., Maga, G. & Berisio, R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells* **9**(2020).
5. Sturman, L.S., Holmes, K.V. & Behnke, J. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J Virol* **33**, 449-62 (1980).
6. de Haan, C.A., Kuo, L., Masters, P.S., Vennema, H. & Rottier, P.J. Coronavirus particle assembly: primary structure requirements of the membrane protein. *J Virol* **72**, 6838-50 (1998).
7. Arndt, A.L., Larson, B.J. & Hogue, B.G. A conserved domain in the coronavirus membrane protein tail is important for virus assembly. *J Virol* **84**, 11418-28 (2010).
8. de Haan, C.A., Smeets, M., Vernooij, F., Vennema, H. & Rottier, P.J. Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. *J Virol* **73**, 7441-52 (1999).
9. Neuman, B.W. et al. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* **174**, 11-22 (2011).
10. Siu, Y.L. et al. The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J Virol* **82**, 11318-30 (2008).
11. Lim, K.P. & Liu, D.X. The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins. *J Biol Chem* **276**, 17515-23 (2001).
12. Boson, B. et al. The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *J Biol Chem* **296**, 100111 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

13. Kuo, L., Hurst-Hess, K.R., Koetzner, C.A. & Masters, P.S. Analyses of Coronavirus Assembly Interactions with Interspecies Membrane and Nucleocapsid Protein Chimeras. *J Virol* **90**, 4357-4368 (2016).
14. Kuo, L., Koetzner, C.A. & Masters, P.S. A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virology* **494**, 100-7 (2016).
15. Hurst, K.R. et al. A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. *J Virol* **79**, 13285-97 (2005).
16. Zheng, Y. et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) membrane (M) protein inhibits type I and III interferon production by targeting RIG-I/MDA-5 signaling. *Signal Transduct Target Ther* **5**, 299 (2020).
17. Fu, Y.Z. et al. SARS-CoV-2 membrane glycoprotein M antagonizes the MAVS-mediated innate antiviral response. *Cell Mol Immunol* **18**, 613-620 (2021).
18. Lei, X. et al. Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun* **11**, 3810 (2020).
19. Sui, L. et al. SARS-CoV-2 Membrane Protein Inhibits Type I Interferon Production Through Ubiquitin-Mediated Degradation of TBK1. *Front Immunol* **12**, 662989 (2021).
20. Heo, L. & Feig, M. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv* (2020).
21. Jumper, J., Tunyasuvunakool, K., Kohli, P., Hassabis, D. & Team, A. Computational predictions of protein structures associated with COVID-19. *DeepMind website* (2020).
22. Rodriguez-Enriquez, A. et al. Predicted 3D model of the M protein of Porcine Epidemic Diarrhea Virus and analysis of its immunogenic potential. *PLoS One* **17**, e0263582 (2022).
23. Thomas, S. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. *Pathog Immun* **5**, 342-363 (2020).
24. Li, X. et al. Genomic Feature Analysis of Betacoronavirus Provides Insights Into SARS and COVID-19 Pandemics. *Front Microbiol* **12**, 614494 (2021).
25. Rossmann, M.G. The molecular replacement method. *Acta Crystallogr A* **46 (Pt 2)**, 73-82 (1990).
26. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
27. Rice, L.M., Earnest, T.N. & Brunger, A.T. Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr D Biol Crystallogr* **56**, 1413-20 (2000).
28. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc* **4**, 706-31 (2009).
29. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88 (2006).
30. Liu, S., Li, S., Yang, Y. & Li, W. Termini restraining of small membrane proteins enables structure determination at near-atomic resolution. *Sci Adv* **6**(2020).
31. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29-42 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Kern, D.M. et al. Author Correction: Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat Struct Mol Biol* **28**, 702 (2021).
33. Luo, H. et al. Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. *Int J Biochem Cell Biol* **38**, 589-99 (2006).
34. Verma, S., Bednar, V., Blount, A. & Hogue, B.G. Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. *J Virol* **80**, 4344-55 (2006).
35. Verma, S., Lopez, L.A., Bednar, V. & Hogue, B.G. Importance of the penultimate positive charge in mouse hepatitis coronavirus A59 membrane protein. *J Virol* **81**, 5339-48 (2007).
36. Yan, Y., Tao, H., He, J. & Huang, S.Y. The HDOCK server for integrated protein-protein docking. *Nat Protoc* **15**, 1829-1852 (2020).
37. Remmert, M., Biegert, A., Hauser, A. & Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**, 173-5 (2011).
38. Dolan, K.A. et al. Structure of SARS-CoV-2 M protein in lipid nanodiscs. *bioRxiv* (2022).
39. Chu, R. et al. Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *J Mol Biol* **323**, 253-62 (2002).
40. Zhang, W.Z. et al. The protein complex crystallography beamline (BL19U1) at the Shanghai Synchrotron Radiation Facility. *NUCL SCI TECH* **30**, 170-81 (2019).
41. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302 (2011).
42. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
43. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).
44. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
46. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-67 (2012).
47. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Online methods

Protein expression and purification. The gene encoding M protein of *Pipistrellus* bat coronavirus HKU5 (batCOV5-M, NCBI accession YP_001039968.1) was synthesized (Genewiz, China) and cloned into a modified pPICZ plasmid (Thermo Fisher Scientific) containing an N-terminal tag of FLAG-His_{10}-TEV protease recognition site. All batCOV5-M mutations were introduced by QuikChange II system (Agilent) according to manufacturer's recommendation, and all mutations were verified by DNA sequencing. The constructs were linearized and transformed into *Pichia pastoris* strain GS115 by lithium chloride/single-strand carrier DNA/polyethylene glycol method according to manufacturer's manual (Thermo Fisher Scientific). The transformants were inoculated into YPD medium consisting of 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) D-(+)-glucose at 30 °C with shaking at 220 rpm until an OD_{600} of 3-5 was reached. To induce protein expression, yeast cells were harvested by centrifugation and resuspended to an OD_{600} of 1 in YPM medium consisting of 1% (w/v) yeast extract, 2% (w/v) peptone, 0.8% (v/v) methanol and 2.5% (v/v) dimethyl sulfoxide (DMSO) at 30 °C for 24 h. Cell pellets were resuspended in Lysis Solution (LS) containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM phenylmethanesulfonyl fluoride (PMSF) and 2 mM β-mercaptoethanol, and were lysed by an AH-1500 high-pressure homogenizer (ATS, China) at 1,300 MPa. Undisrupted cells and cell debris were separated by centrifugation at 3,000 x g, and membrane were collected by ultracentrifugation at 140,000 x g for 1 h at 4 °C. Protein was extracted by addition of 1% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace) at 4 °C for 2 h and the extraction mixture was centrifuged at 200,000 x g for 30 min at 4 °C. The supernatant was incubated with Co^{2+} resin in the presence of 20 mM imidazole pH 8.0 at 4 °C for 1 h, and the mixture was loaded in an empty chromatography column. The resin/protein was washed with 20 bed-volume of LS containing 2 mM DDM and 30 mM imidazole pH 8.0, and the protein was eluted with LS supplemented with 2 mM DDM and 250 mM imidazole pH 8.0.

To generate the batCOV5-M_{CTD}-GFP construct, a superfolder GFP[29] was split into two halves[30] and was fused to the N- and C-termini of batCOV5-M_{CTD} (residues 115-203 of batCOV5-M) by gene synthesis (Genewiz, China). The fusion protein-encoding DNA was cloned into a modified pPICZ plasmid (Thermo Fisher Scientific) containing a C-terminal TEV site and a His_{10} tag. Transformation and expression of batCOV5-M_{CTD}-GFP followed the same protocol as batCOV5-M except that the expression was induced at 25 °C. Cell pellets were resuspended in LS containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 20% (v/v) glycerol and 1 mM PMSF, and were lysed similarly to batCOV5-M. Undisrupted cells and cell debris were separated by

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

centrifugation at 140,000 x g at 4 °C for 1 h. The supernatant was supplemented with 20 mM imidazole pH 8.0 and was immediately loaded onto a pre-washed Co^{2+} affinity column. The column was then washed with 20 bed-volume of LS containing 30 mM imidazole pH 8.0, and the protein was eluted with LS containing 250 mM imidazole pH 8.0.

The gene encoding N protein of batCOV5 (batCOV5-N, NCBI accession YP_001039969.1) was synthesized (Genewiz, China) and cloned into a modified pPICZ plasmid (Thermo Fisher Scientific) containing an N-terminal tag of FLAG-His_{10}-TEV site, followed by the bacterial cytochrome b562RIL (BRIL)[39] to improve expression of the protein of interest. All batCOV5-N fragments were generated by polymerase chain reaction (PCR) and cloned into the same vector as batCOV5-N. For microscale thermophoresis (MST) analysis, batCOV5-N_3-enconding DNA was sub-cloned into another modified pPICZ plasmid (Thermo Fisher Scientific) containing a C-terminal tag of TEV site-GFP-His_{10}. Single mutations of batCOV5-N_3 were introduced by QuikChange II system (Agilent). Transformation, expression and purification of batCOV5-N and fragments followed the same protocol as batCOV5-M_{CTD}-GFP, except that LS contained 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol and 1 mM PMSF.

Crystallization. Affinity-purified batCOV5-M protein was concentrated to 5 mg/ml, and treated with trypsin (TPCK-treated, Sigma-Aldrich) at a 1:50 ratio (trypsin:batCOV5-M, w/w) for 20 min at 18 °C to generate a stable core. The digestion was stopped by 10 mM PMSF and the protein was further purified by size-exclusion chromatography (SEC) in a buffer consisting of 150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.5 mM DDM and 5 mM β-mercaptoethanol. The peak fractions were pooled, concentrated to ~30 mg/ml, and mixed with 1-oleoyl-rac-glycerol (monoolein, Sigma-Aldrich) at a 2:3 ratio (protein:lipid, w/w) using the twin-syringe mixing method[28]. The protein-lipid mixture was dispensed manually in ~50-nl drops onto 96-well glass sandwich plates and overlaid with 0.8 µl of precipitant solution per drop. The batCOV5-M crystals were grown in 300 mM ammonium formate, 50 mM Tris-HCl pH 8.8, 35% (v/v) PEG 500 monomethyl ether and 7.1 mM pentaethylene glycol monooctyl ether. The crystals usually appear in one week and grow to full size in two weeks, and were flash-frozen directly in liquid nitrogen without additional cryoprotection.

For batCOV5-M_{CTD}-GFP, affinity-purified protein was used directly for crystallization by vapor diffusion sitting-drop method without further SEC purification or concentration step. The batCOV5-M_{CTD}-GFP crystals were grown in 0.4 M ammonium sulfate, 0.1 M Bis-Tris pH 5.3, 27% (w/v) PEG 3350 and 0.5% (v/v) ethyl acetate. The crystals were either cryoprotected by 15-20% (v/v) glycerol or directly flash-frozen in liquid nitrogen without additional cryoprotection.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data collection and structure solution. Diffraction data were collected on beamlines BL18U1 and BL19U1[40] of National Facility for Protein Science in Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility (SSRF). The data were indexed, integrated and scaled using the autoPROC pipeline package (Global Phasing Limited)[41], which includes XDS[42] and AIMLESS (CCP4 package)[43]. The batCOV5-M_{CTD}-GFP structure was solved by molecular replacement with Phaser[44] using a published superfolder GFP structure[29] (PDB: 2B3Q) as a template. The full-length batCOV5-M structure was then solved by molecular replacement using the assembled batCOV5-M_{CTD} structure (Supplementary Fig. 2e) as a searching model. Manual model building and refinement was carried out using Coot[45] and phenix.refine[46], and Molprobity[47] was used to monitor and improve protein geometry. Non-crystallographic symmetry (NCS) restraints were applied throughout the refinement to improve maps. The data collection and refinement statistics were generated using phenix.table_one[46] and the values are listed in Supplementary Table 2. All structural figures, RMSD calculations and length/angle measurements were performed in PyMOL (Schr□dinger, LLC).

Pull-down assay. For pull-down assays, affinity-purified batCOV5-M and mutants were treated with TEV protease (to remove His-tag) and Endoglycosidase H (New England Biolabs), and purified batCOV5-M_{CTD}-GFP was treated with TEV protease, then all were further purified by SEC. In a pull-down experiment, affinity-purified batCOV5-N and fragments were first incubated with Co^{2+} beads for 30 min at 4 °C, followed by addition of His-tag-free batCOV5-M or variants to continue incubation for another 30 min at 4 °C in the presence of 30 mM imidazole pH 8.0. The Co^{2+} beads were collected by centrifugation at 5,000 x g for 1 min and washed twice with the assay buffer before being analyzed by SDS-PAGE gels.

Docking of batCOV5-N_{3C} in batCOV5-M. Computational docking was performed using the HDOCK web server[36] (http://hdock.phys.hust.edu.cn/). For receptor, the batCOV5-M structure was prepared as a .pdb file containing only one protein dimer (chain A and chain B) by removing other chains and all “HETATM” records. For ligand, non-polar residues at both ends of the batCOV5-N_{3C} sequence (VPSFEDVVDAIFPDSEA, 17 residues) were removed to generate a core sequence (SFEDVVDAIFPDSE, 14 residues), which was submitted directly to the system for template-free docking. Docked models were manually examined to remove incorrectly placed batCOV5-N_{3C} (e.g. in TMD).

Microscale thermophoresis (MST). MST analysis was performed using Monolith NT.115 (NanoTemper, Germany). All affinity-purified proteins were further purified by SEC in a buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.5 mM DDM. The peak fractions were

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pooled and diluted to 100 nM using the SEC buffer. The GFP moiety fused to batCOV5-N_3 and its mutants provided the fluorescence signal required by MST. Meanwhile, purified batCOV5-M and batCOV5-M_{H155A} served as ligands and were prepared according to the MST manual with the highest concentration of 32 μM for both proteins. The batCOV5-N_3 and mutant samples were mixed with serial-diluted ligands and were incubated for 10 min at room temperature. Then the samples were loaded into capillaries and MST measurements were performed according to the Monolith manual. The equilibrium dissociation constant (K_d) was determined in the MO.Affinity Analysis software (NanoTemper, Germany) with the K_d fit function. All MST measurements were performed in three biologically independent experiments (N=3), and K_d values are expressed as mean ± SD in the text and Supplementary Table 1. Two-tailed Student's t-test was performed for statistical analysis in Supplementary Table 1.

Acknowledgements

Diffraction data used in this study were collected on beamlines BL18U1 and BL19U1 of National Facility for Protein Science in Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility (SSRF). The authors thank the staff from these beamlines for assistance during data collection. This work was supported in part by the National Natural Science Foundation of China (NSFC) grant 31770783, and the 1.3.5 Project for Disciplines of Excellence grant by West China Hospital of Sichuan University to XZ.

Data availability

The atomic coordinates and structure factors generated in this study have been deposited in the Protein Data Bank under the following accession codes: 7Y96 for batCOV5-M_{CTD}-GFP and 7Y9B for batCOV5-M. Three previously reported structures used in this study are available in the Protein Data Bank under the following accession codes: 2B3Q for a superfolder GFP, 6XDC for ORF3a of SARS-COV-2 and 8CTK for M protein of SARS-COV-2. Source data of K_d values are provided in Supplementary Table 1.

Author contributions

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

XZ and ZS conceived and supervised the project, and wrote the manuscript with input from all authors. XW performed structural and functional studies; YY performed functional studies; and all authors analyzed the data.

Competing interests

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

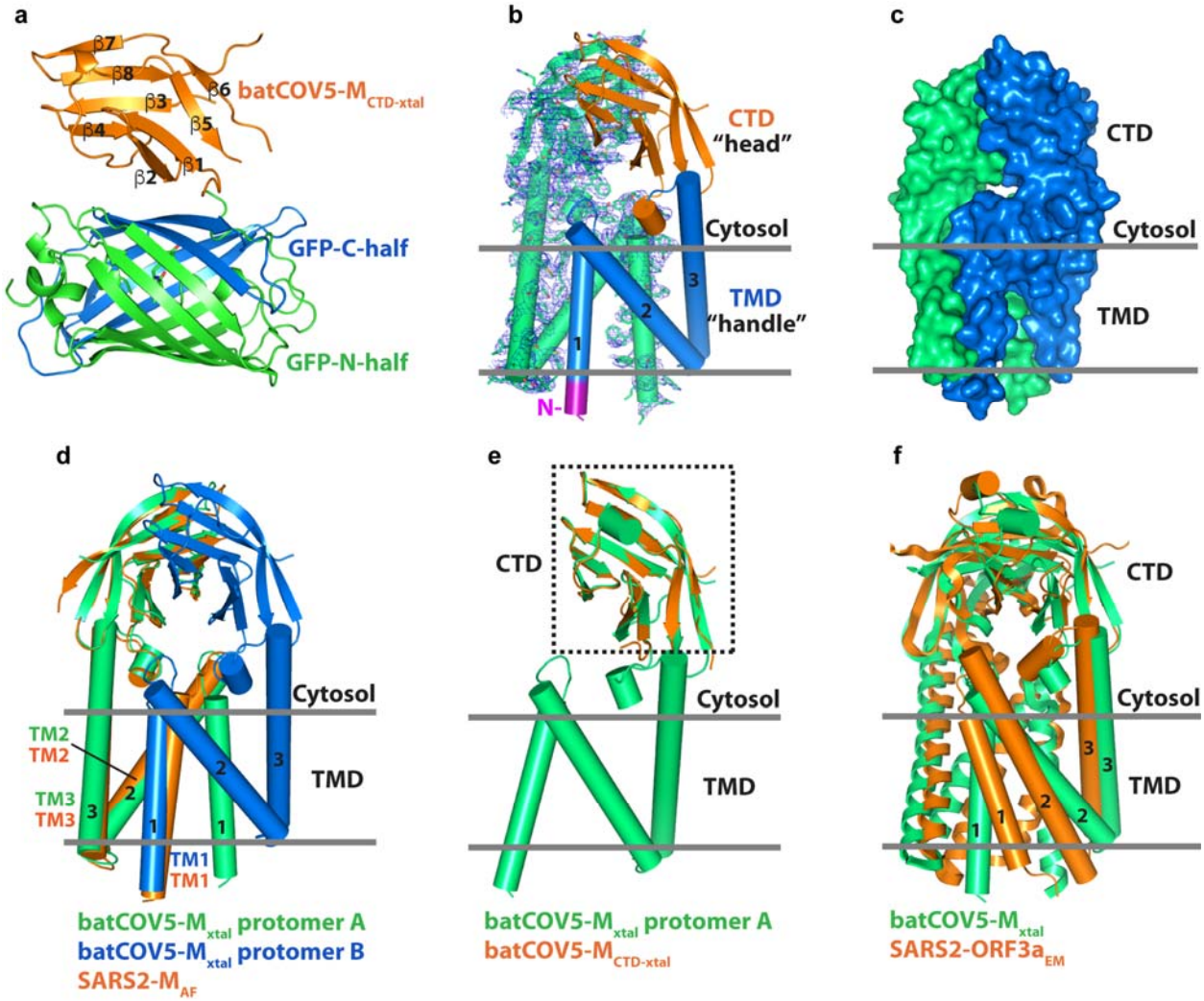

Figure 1. Crystal structure of M protein of the betacoronavirus batCOV5. (a) Crystal structure of batCOV5-M_{CTD} (in orange) fused with a split superfolder GFP (N-terminal half in green and C-terminal half in blue) rendered in the cartoon mode. The β strands of CTD are labeled from β1 to β8. (b) Crystal structure of batCOV5-M homodimer. For one protomer (in green), the $2F_o$-F_c map contoured at 1.2 σ level is shown as blue mesh. In the other protomer, the N-terminus, TMD and CTD are displayed in purple, blue and orange, respectively. The TM helices are labeled from 1 to 3 throughout this manuscript. The "head" and "handle" labels refer to an analogy to a pair of pincers in the main text. Relative membrane position is indicated by two grey lines throughout this manuscript. (c) The batCOV5-M_{xtal} dimer is rendered as surfaces with one protomer in green and the other in blue. (d) Superposition of the AlphaFold-predicted SARS2-M_{AF} monomer model (in orange) onto protomer A (in green) of batCOV5-M_{xtal} dimer. The batCOV5-M_{xtal} protomer B is displayed in blue. (e) Superposition of batCOV5-$M_{CTD\text{-}xtal}$ (in

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

orange) onto one protomer of batCOV5-M_{xtal} (in green). (f) Superposition of the cryo-EM structure of SARS2-ORF3a (in orange) onto batCOV5-M_{xtal} dimer (in green). For better viewing, each structure has one protomer rendered as cylinders.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

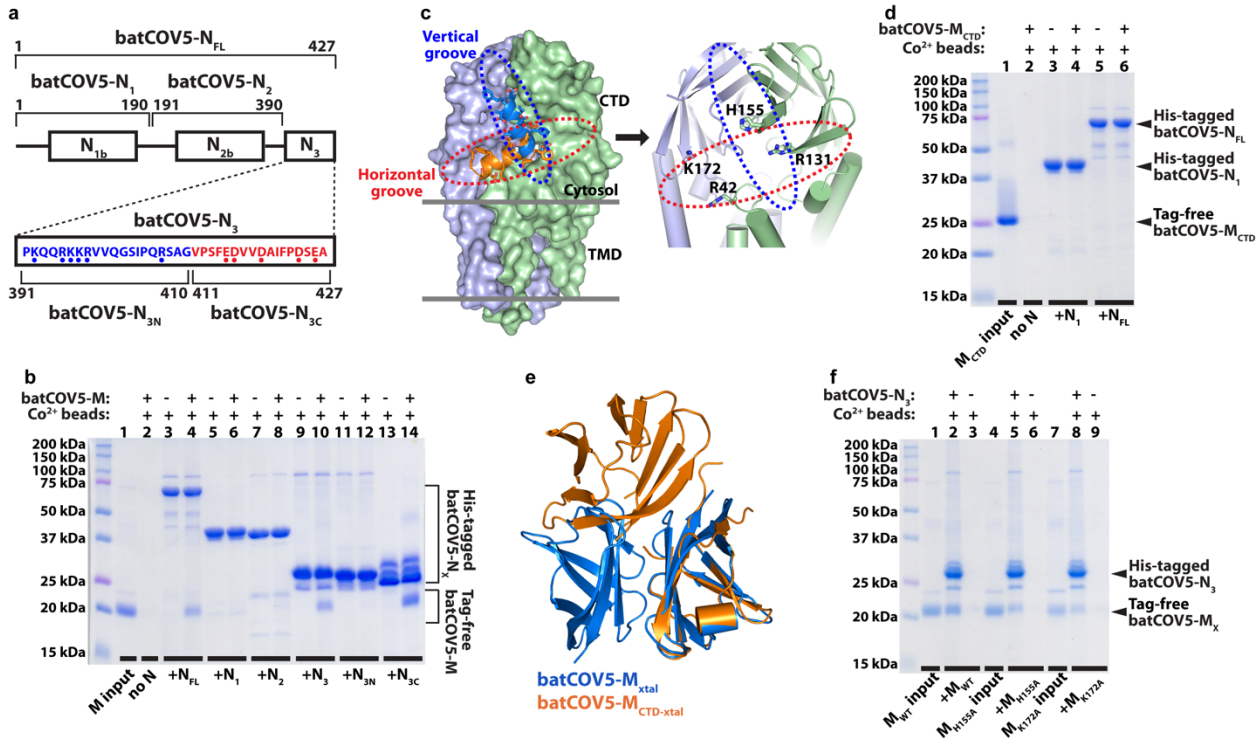

Figure 2. The interaction mechanism between M and N proteins of batCOV5. (a) Scheme of batCOV5-N fragments used in this study. The N_{3N} and N_{3C} sequences are shown with blue or red dots indicating basic or acidic residues, respectively. (b) A pull-down experiment using purified His-tagged batCOV5-N fragments to pull down tag-free batCOV5-M protein is examined by SDS-PAGE gel analysis. Signs and labels above and below the gel indicate the components present in the pull-down experiment. The label "batCOV5-N_X" refers to all batCOV5-N fragments in this panel. Lane 1 is the batCOV5-M protein input for pull-down. Molecular weight of protein standards is labeled on the left of gel. (c) Left: Two docked models of batCOV5-N_{3C} (as an orange helix and a blue helix, respectively) into the batCOV5-M_{xtal} dimer structure. The two protomers are displayed in light blue and light green surfaces. The "horizontal groove" and the "vertical groove" are indicated by a red and a blue dashed oval, respectively. Right: An enlarged view of the "horizontal groove" and the "vertical groove". Four basic residues within the two grooves are rendered as sticks. (d) A pull-down experiment using purified His-tagged batCOV5-N fragments to pull down tag-free batCOV5-M_{CTD}-GFP protein, similar to panel b. To simplify text labels, "batCOV5-M_{CTD}-GFP" is shortened as "batCOV5-M_{CTD}" in this panel. (e) Superposition of batCOV5-$M_{CTD\text{-}xtal}$ dimer (in orange) onto CTD of the batCOV5-M_{xtal} dimer (in blue). (f) A pull-down experiment using purified His-tagged batCOV5-N_3 to pull down tag-free wild-type or mutant batCOV5-M proteins, similar to panel b. The label "batCOV5-M_X" refers to all

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

batCOV5-M variants in this panel. All pull-down experiments were repeated three times using biologically independent samples with similar results.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.28.497981; this version posted July 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

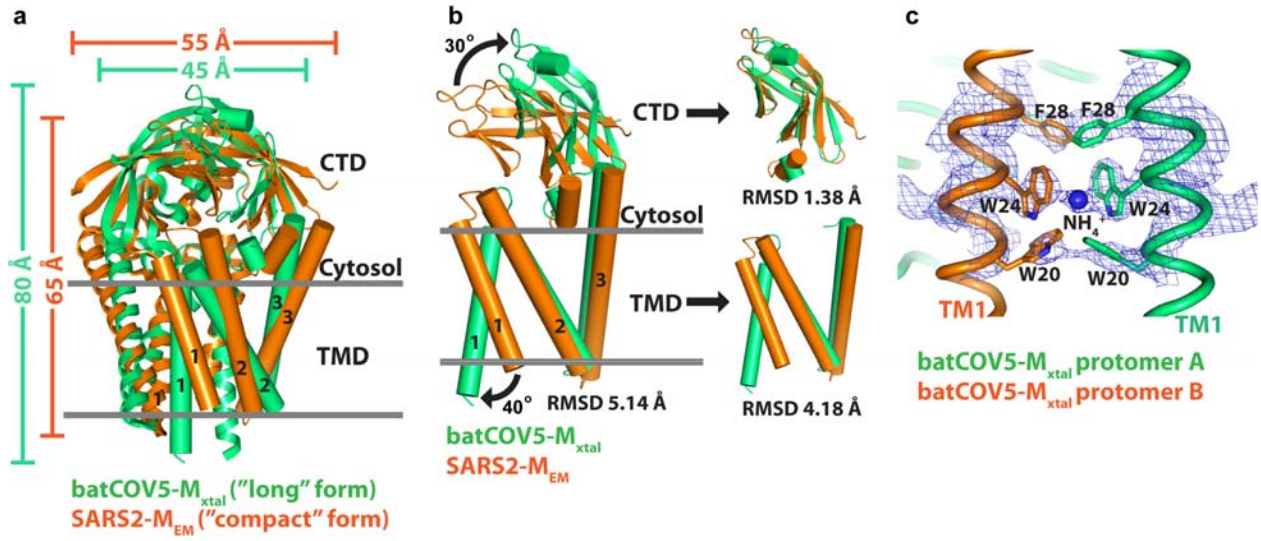

Figure 3. Structural comparison between batCOV5-M$_{xtal}$ and SARS-M$_{EM}$. (a) Superposition of the cryo-EM structure of SARS2-M (in orange) onto batCOV5-M$_{xtal}$ dimer (in green). For better viewing, each structure has one protomer rendered as cylinders. Dimensions of the two dimers are indicated. (b) Left: Superposition of one SARS2-M$_{EM}$ protomer (in orange) onto one batCOV5-M$_{xtal}$ protomer (in green) using TM2/TM3 as anchors. Right: Individual CTDs and TMDs of the two structures are aligned using C_α atoms. (c) The simulated-annealing $2F_o$-F_c map contoured at 1.2 σ level is shown as blue mesh around an ammonium ion between the two Trp24 residues of TM1 in batCOV5-M$_{xtal}$ dimer. The two protomers are displayed as tubes in orange and in green, respectively. The ammonium ion is rendered as a blue sphere and Trp24 as sticks.