Elucidation of SARS-CoV-2 budding mechanisms through molecular dynamics simulations of M and E protein complexes

Logan T. Collins[1,2,*], Tamer Elkholy[1,3], Shafat Mubin[1,4], David Hill[1,5], Ricky Williams[1,6], Kayode Ezike[1,7], Ankush Singhal[1,8]

[*]Corresponding author; [1]Conduit Computing; [2]Washington University in St. Louis, Department of Biomedical Engineering; [3]Zapata Computing; [4]Valdosta State University, Department of Physics; [5]Whiteboard Federal Technologies; [6]Harvard University, Department of Electrical Engineering; [7]Attune; [8]Leiden University, Department of Chemistry

Abstract:
SARS-CoV-2 and other coronaviruses pose major threats to global health, yet computational efforts to understand them have largely overlooked the process of budding, a key part of the coronavirus life cycle. When expressed together, coronavirus M and E proteins are sufficient to facilitate budding into the ER-Golgi intermediate compartment (ERGIC). To help elucidate budding, we ran atomistic molecular dynamics (MD) simulations using the Feig laboratory's refined structural models of the SARS-CoV-2 M protein dimer and E protein pentamer. Our MD simulations consisted of M protein dimers and E protein pentamers in patches of membrane. By examining where these proteins induced membrane curvature *in silico*, we obtained insights around how the budding process may occur. The M protein dimers acted cooperatively to induce membrane curvature while E protein pentamers kept the membrane planar. These results could eventually help guide the development of antiviral therapeutics which inhibit coronavirus budding.

Introduction:

The global COVID-19 pandemic has highlighted the need for diverse ways of responding to infectious disease. Though vaccines, diagnostics, and treatments have all helped combat SARS-CoV-2, the pandemic continues to wreak havoc across our world. While much has been learned about the biology of SARS-CoV-2, some aspects of its life cycle still represent areas of mystery. It is of paramount importance to develop new tools for fighting coronaviruses, so gaining improved understanding of understudied parts of the SARS-CoV-2 life cycle may open valuable doors.

In order to form infectious particles, coronaviruses must bud into the ERGIC.[1] When expressed together without the help of any other coronavirus proteins, the M and E proteins are sufficient to allow budding of virus-like particles (VLPs) which resemble those produced by wild-type coronaviruses.[2–4] The exact mechanisms by which the M and E proteins contribute to budding remain unclear. Some have proposed that M proteins oligomerize into a matrix layer to induce membrane curvature,[5,6] though more recent data on SARS-CoV-2 has indicated that its M proteins do not form such a matrix.[7] The role of the E protein in budding is also poorly understood, though it is thought to somehow coordinate envelope assembly.[6,8,9] It should be noted that the M protein is roughly 300 times more abundant in the ERGIC than the E protein.[10] Expression of the nucleocapsid N protein has also been shown to greatly enhance the yield of budding VLPs compared to when only the M and E protein are present.[11] By contrast, the famous S protein is not strictly required for coronavirus budding, though it is incorporated into the VLPs when expressed alongside M and E.[3] Because of their vital role in budding, the M and E proteins represent key components of the coronavirus life cycle.

Molecular dynamics (MD) simulations can help to better understand biological phenomena, yet there has not been much work involving MD and coronavirus budding. Monje-Galvan and Voth recently performed MD simulations which characterized the movements of individual M protein dimers and individual E protein pentamers in virtual ERGIC membrane.[19] This revealed some new insights, including that the M protein dimer can introduce local deformations in the membrane. However, their study did not investigate how multiple interacting M dimers or multiple interacting E pentamers might influence membrane curvature, which is important for understanding budding. Yu et al. reported a coarse-grained MD investigation of the completed SARS-CoV-2 virion, which included numerous M, E, and S proteins.[20] Though the study did involve all of the three structural proteins, it focused on the completed spherical virus rather than on budding. There remains a need for MD simulations of the budding process which interrogate how multiple SARS-CoV-2 structural protein complexes may facilitate budding.

We utilized atomistic MD simulations via GROMACS to investigate the roles of M and E protein complexes in budding. Because of the lack of complete crystal structures of the M and E proteins, we used the Feig laboratory's predicted structural models of the M protein dimer and E protein pentamer.[21] We constructed planar membrane patches with lipid composition mimicking that of the ERGIC and inserted transmembrane M and E protein complexes. We ran 800 ns simulations on five systems: a membrane-only system (mem), a system with a single E protein pentamer (1E), a system with four E protein pentamers (4E), a system with a single M protein dimer (1M), and a system with four M protein dimers (4M). Though the focus of our study was on the interactions between complexes of the same type, we also ran a 400 ns simulation on a system with three M protein dimers and one E protein pentamer (3M1E). From our simulations, one of the most notable outcomes was that the 4M system gained a substantial degree of curvature over time (Fig. 1A), while other systems such as mem had much less curvature (Fig. 1B). Our simulations revealed various other insights around the mechanisms of SARS-CoV-2 budding as well.

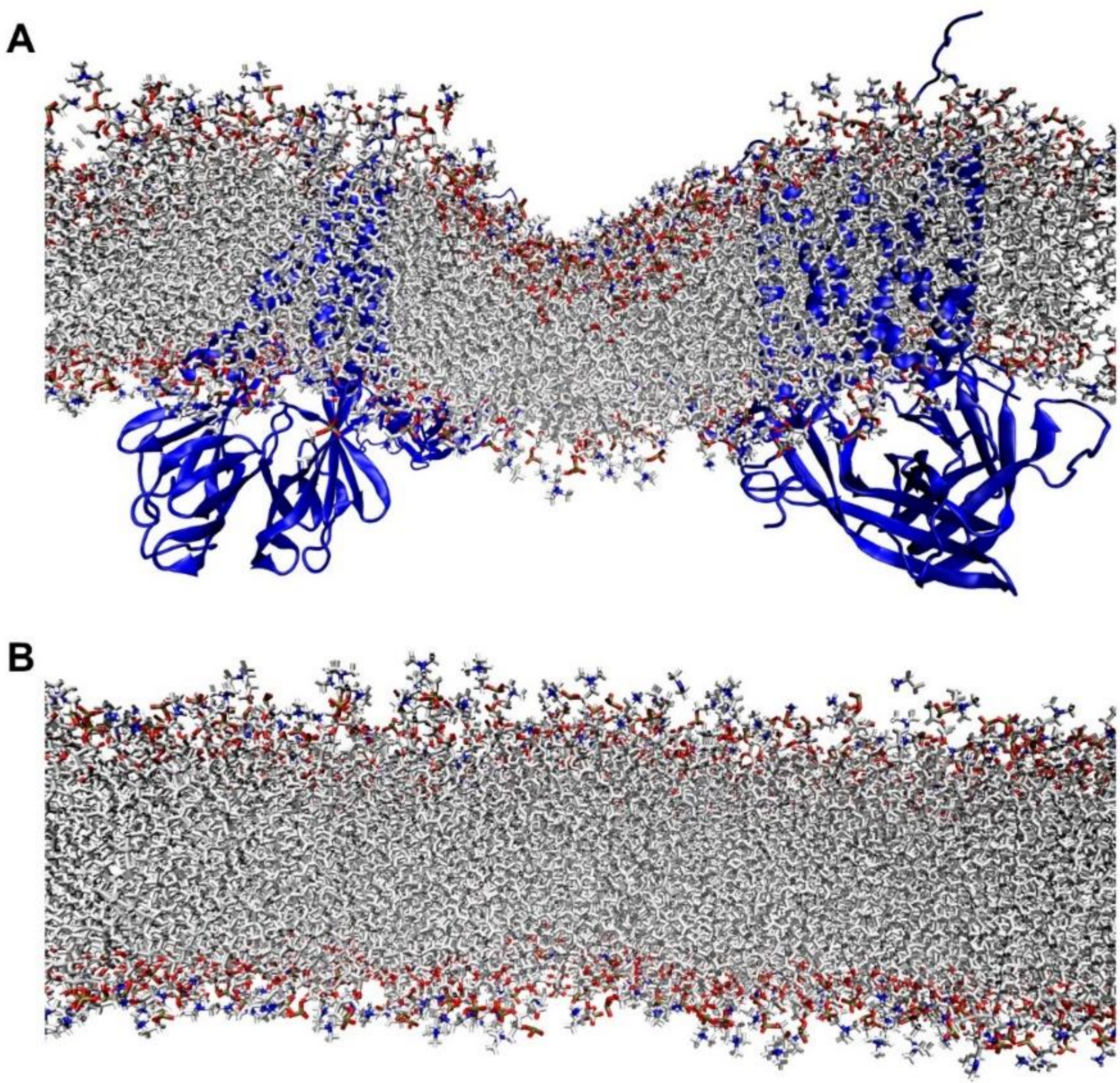

Figure 1 Representative side-view snapshots of **(A)** the high level of curvature in the 4M system at 800 ns and **(B)** the lack of curvature in the mem system at 800 ns.

Results:

We employed g_lomepro[23] to generate 2D time-averaged mean curvature heatmaps over selected 100 ns intervals (Fig. 2A-E, Fig. S1A) as well as 3D plots of the same data (Fig. 3A-E, Fig. S1B). The 1M system showed a small bulge which grew more pronounced over time, indicating that even lone M protein dimers might induce kinks in the membrane. The 4M system showed by far the highest levels of curvature. Remarkably, the 4M system's curvature grew both in magnitude and in orderliness over time. In 4M's 700-800 ns interval, a cylindrical hill passed through the membrane, demonstrating the ability of the M proteins to work together in an organized fashion. Only small amounts of curvature were visible in the 1E, 4E, and mem simulations, indicating that E protein pentamers may play a role during budding which does not directly involve the induction of curvature. The 3M1E system showed moderate curvature, which was less pronounced than in the 4M system. In summary, these data indicate that E proteins likely do not induce substantial curvature, that isolated M proteins create bulges in the membrane, and that many M proteins together can act cooperatively to induce larger amounts of curvature.

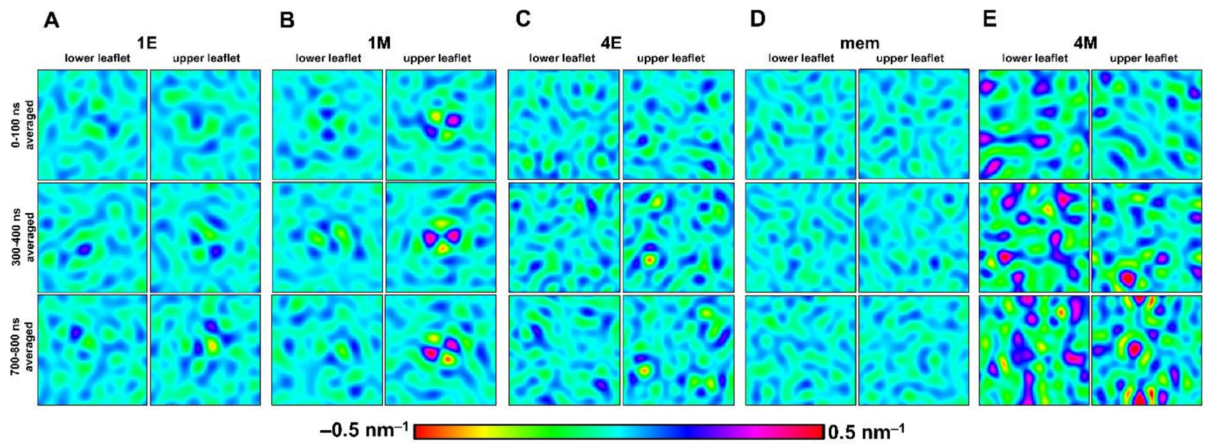

Figure 2 Time-averaged mean curvature heatmaps over selected time intervals for the **(A)** 1E system, **(B)** 1M system, **(C)** 4E system, **(D)** mem system, and **(E)** 4M system.

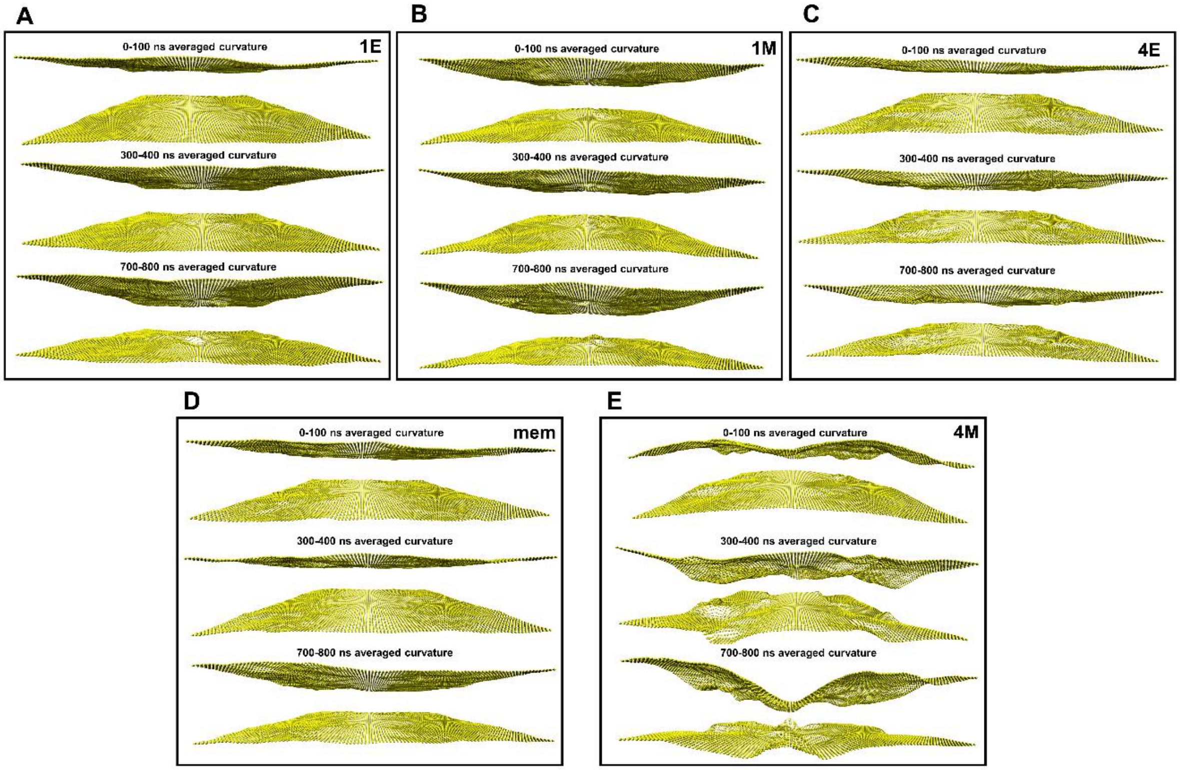

Figure 3 Time-averaged mean curvature 3D plots over selected time intervals for the **(A)** 1E system, **(B)** 1M system, **(C)** 4E system, **(D)** mem system, and **(E)** 4M system.

We characterized protein dynamics using MDanalysis[22] to perform root-mean-square deviation (RMSD) (Fig. 4A-D, Fig. S2A) and radius of gyration (R_g) (Fig. 4E-H, Fig. S2B) calculations. By comparison to the M proteins, the E proteins consistently reached higher RMSD values. This is likely due to the flexible hinge regions connecting the E protein cytosolic α-helices to their transmembrane α-helices. The comparative lack of flexibility in the M proteins may facilitate retention of their wedgelike shape, which could help induce membrane curvature. We tentatively suggest that the less restricted motion of the E protein cytosolic α-helices could play a role in sequestering M proteins to coordinate budding. Similarly, the R_g values of the M proteins remained relatively constant over time while the R_g values of the E proteins exhibited greater variability over time. These R_g data corroborate the notion that the M protein dimers have relatively rigid conformations while the E protein pentamers may have looser structures. The differences in flexibility between the M protein and E protein cytosolic domains provide hints as to their functional mechanisms.

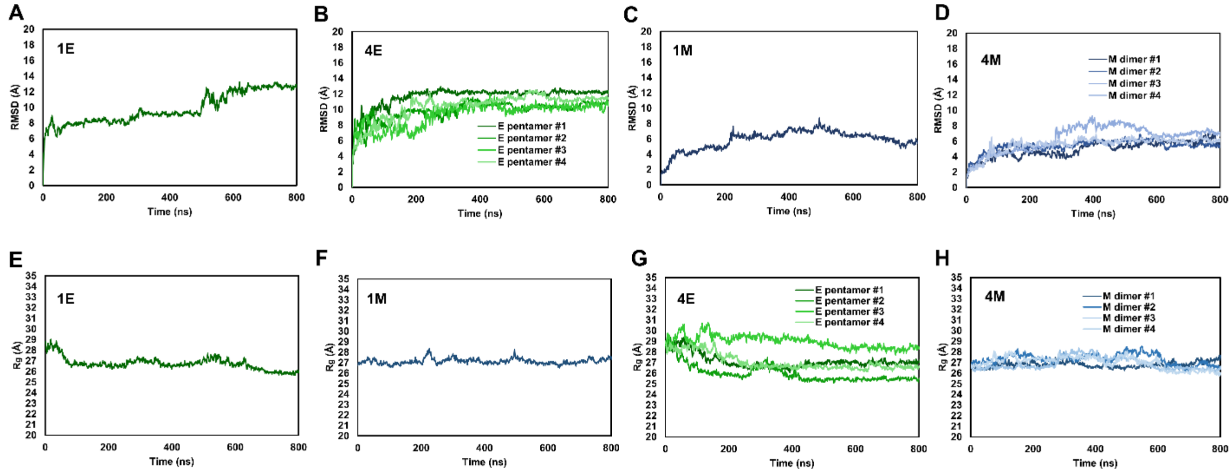

Figure 4 RMSD plots for the **(A)** 1E simulation, **(B)** 4E simulation, **(C)** 1M simulation, and **(D)** 4M simulation as well as R_g plots for the **(E)** 1E simulation, **(F)** 1M simulation, **(G)** 4E simulation, and **(H)** 4M simulation.

Discussion:

Our atomistic MD simulations uncovered insights around the roles of M and E proteins in the budding of SARS-CoV-2. Multiple M protein dimers induced curvature in a cooperative fashion. Because coronaviruses are known to produce large numbers of M proteins in the ERGIC membranes of infected cells,[10] we hypothesize that this cooperative effect might increase further in the biological reality, leading to enough curvature to encapsulate the RNA genome of the virus. The lack of curvature in the 1E and 4E simulations indicates that the E protein likely does not directly facilitate membrane curvature during SARS-CoV-2 budding. But since experimental results indicate that E proteins are essential for budding coronaviruses,[2–4] the E protein likely still plays another role in budding. One possibility is that the E protein introduces a planar region into the membrane's overall curvature profile, eventually creating a viral envelope with a larger radius of curvature than would be possible with only the M proteins.

Based on the results of our simulations, we propose that the M protein dimer may represent a valuable target for drugs intended to treat COVID-19 and other coronavirus diseases. Due to the high level of conservation of the M protein across different types of coronaviruses,[12] we postulate that a drug affecting the M protein might have a broad degree of efficacy. Pharmaceuticals which target the M protein could provide a powerful approach by which to mitigate the effects of coronavirus infections.

Methods:

Six MD simulations of M and E proteins in lipid membrane were used in this study. All of the simulations were carried out at atomic resolution using GROMACS 2019.4.[24] Structures of the E protein pentamer and M protein dimer were obtained from the Feig laboratory's predicted models.[21] Six initial configurations were constructed: a membrane-only system (mem), a system with a single E protein pentamer in membrane (1E), a system with four E protein pentamers in membrane (4E), a system with a single M protein dimer in membrane (1M), a system with four M protein dimers in membrane (4M), and a system with three M protein dimers and one E protein pentamer in membrane (3M1E). To mimic the biological ERGIC, the membrane composition used for all six systems was as follows: 57% POPC, 25% POPE, 10% POPI, 2% POPS, 6% Cholesterol.[21] All the systems were solvated using explicit water molecules and ions. The CHARMM36 force field[25] was used for all lipids, ions, and proteins, while the TIP3P[26] model was implemented for the water molecules. All hydrogen atoms were constrained with the LINCS algorithm,[27] and long-range electrostatics were evaluated with particle-mesh Ewald summation.[28] All simulations used 2 fs time step with Leap-Frog integrator[29] and a 1.4 nm cutoff for all of the interactions. A standard energy minimization procedure was performed using the steepest descent method.[30] For each simulation, a small NPT equilibration run was performed followed by a production run using a Nose-Hoover thermostat[31] at 300K and a Parrinello-Rahman barostat[32] at 1 atm. The lengths of the production runs were as follows: 800 ns for mem, 1E, 4E, 1M, and 4M and 400 ns for 3M1E.

Analyses of the results of the simulations included RMSD, R_g, and time-averaged mean curvature of the membranes. MDanalysis[22] was used to calculate RMSD and R_g while g_lomepro[23] was used for the membrane curvature calculations. Each protein's RMSD was calculated at 0.1 ns intervals by comparing its conformation at a given time step to a reference conformation consisting of the initial equilibrated structure. To correct for the effects of proteins undergoing translations and rotations during the simulation runs, RMSD was adjusted by translating with a vector $\mathbf{t}$ and rotating with a matrix $\mathbf{R}$. In this way, only the changes in the proteins relative to their initial reference structures were included in the final RMSD outputs. The RMSD was calculated using the coordinates of all of the α-carbon atoms in the given protein where $\mathbf{x}$ describes the coordinates in the current conformation, $\mathbf{x}_{\text{ref}}$ are the coordinates of the reference conformation, and n is the number of α-carbon atoms in the protein.

$$\text{RMSD} = \frac{1}{n}\sum_{i=1}^{n}|(\mathbf{R}\cdot\mathbf{x}_i + \mathbf{t}) - \mathbf{x}_{\text{ref}}|$$

Similarly, R_g was calculated for the α-carbon atoms of each protein at 0.1 ns intervals to analyze changes in the compactness of the proteins. R_g was computed using the displacement vector $\mathbf{r}$ between a given protein's center of mass and each α-carbon of that protein. These calculations were weighted by the mass m of the atom in question.

$$\text{R}_\text{g} = \sqrt{\frac{\sum_{i=1}^{n} m_i|\mathbf{r}_i|^2}{\sum m_i}}$$

For membrane curvature calculations, the g_lomepro[23] software package was used to calculate mean curvature as averaged over the frames of the 0-100 ns, 300-400 ns, and 700-800 ns time periods of the mem, 1E, 1M, 4E 4M, and mem simulations. The same was done for the 3M1E simulation, but only with the 0-100 ns and 300-400 ns time intervals (since that simulation was 400 ns in length). Performing quantitative analyses helped us to decipher insights from our simulations.

Acknowledgements:
Time and resources on the Frontera supercomputer were awarded to Conduit through the COVID-19 High-Performance Computing Consortium project MCB200139.

Conflict of interest:
The authors are affiliated with Conduit Computing, a company which is developing a home diagnostic test for COVID-19 (as well as other infectious diseases) called nanoSPLASH.

References:

(1) Masters, P. S. The Molecular Biology of Coronaviruses; Academic Press, 2006; Vol. 66, pp 193–292. https://doi.org/10.1016/S0065-3527(06)66005-3.

(2) Mortola, E.; Roy, P. Efficient Assembly and Release of SARS Coronavirus-like Particles by a Heterologous Expression System. *FEBS Lett.* **2004**, *576* (1), 174–178. https://doi.org/https://doi.org/10.1016/j.febslet.2004.09.009.

(3) Vennema, H.; Godeke, G. J.; Rossen, J. W.; Voorhout, W. F.; Horzinek, M. C.; Opstelten, D. J.; Rottier, P. J. Nucleocapsid-Independent Assembly of Coronavirus-like Particles by Co-Expression

of Viral Envelope Protein Genes. *EMBO J.* **1996**, *15* (8), 2020–2028. https://doi.org/10.1002/j.1460-2075.1996.tb00553.x.

(4) Baudoux, P.; Carrat, C.; Besnardeau, L.; Charley, B.; Laude, H. Coronavirus Pseudoparticles Formed with Recombinant M and E Proteins Induce Alpha Interferon Synthesis by Leukocytes. *J. Virol.* **1998**, *72* (11), 8636 LP – 8643. https://doi.org/10.1128/JVI.72.11.8636-8643.1998.

(5) Neuman, B. W.; Kiss, G.; Kunding, A. H.; Bhella, D.; Baksh, M. F.; Connelly, S.; Droese, B.; Klaus, J. P.; Makino, S.; Sawicki, S. G.; Siddell, S. G.; Stamou, D. G.; Wilson, I. A.; Kuhn, P.; Buchmeier, M. J. A Structural Analysis of M Protein in Coronavirus Assembly and Morphology. *J. Struct. Biol.* **2011**, *174* (1), 11–22. https://doi.org/https://doi.org/10.1016/j.jsb.2010.11.021.

(6) Siu, Y. L.; Teoh, K. T.; Lo, J.; Chan, C. M.; Kien, F.; Escriou, N.; Tsao, S. W.; Nicholls, J. M.; Altmeyer, R.; Peiris, J. S. M.; Bruzzone, R.; Nal, B. 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.* **2008**, *82* (22), 11318 LP – 11330. https://doi.org/10.1128/JVI.01052-08.

(7) Klein, S.; Cortese, M.; Winter, S. L.; Wachsmuth-Melm, M.; Neufeldt, C. J.; Cerikan, B.; Stanifer, M. L.; Boulant, S.; Bartenschlager, R.; Chlanda, P. SARS-CoV-2 Structure and Replication Characterized by in Situ Cryo-Electron Tomography. *Nat. Commun.* **2020**, *11* (1), 5885. https://doi.org/10.1038/s41467-020-19619-7.

(8) Duart, G.; García-Murria, M. J.; Grau, B.; Acosta-Cáceres, J. M.; Martínez-Gil, L.; Mingarro, I. SARS-CoV-2 Envelope Protein Topology in Eukaryotic Membranes. *Open Biol.* **2021**, *10* (9), 200209. https://doi.org/10.1098/rsob.200209.

(9) Schoeman, D.; Fielding, B. C. Coronavirus Envelope Protein: Current Knowledge. *Virol. J.* **2019**, *16* (1), 69. https://doi.org/10.1186/s12985-019-1182-0.

(10) Cavanagh, D. Coronaviridae: A Review of Coronaviruses and Toroviruses BT - Coronaviruses with Special Emphasis on First Insights Concerning SARS; Schmidt, A., Weber, O., Wolff, M. H., Eds.; Birkhäuser Basel: Basel, 2005; pp 1–54. https://doi.org/10.1007/3-7643-7339-3_1.

(11) McBride, R.; Van Zyl, M.; Fielding, C. B. The Coronavirus Nucleocapsid Is a Multifunctional Protein. *Viruses* . 2014. https://doi.org/10.3390/v6082991.

(12) Tiwari, V.; Beer, J. C.; Sankaranarayanan, N. V.; Swanson-Mungerson, M.; Desai, U. R. Discovering Small-Molecule Therapeutics against SARS-CoV-2. *Drug Discov. Today* **2020**, *25* (8), 1535–1544. https://doi.org/https://doi.org/10.1016/j.drudis.2020.06.017.

(13) Mohamed, K.; Yazdanpanah, N.; Saghazadeh, A.; Rezaei, N. Computational Drug Discovery and Repurposing for the Treatment of COVID-19: A Systematic Review. *Bioorg. Chem.* **2021**, *106*, 104490. https://doi.org/https://doi.org/10.1016/j.bioorg.2020.104490.

(14) Chilamakuri, R.; Agarwal, S. COVID-19: Characteristics and Therapeutics. *Cells* . 2021. https://doi.org/10.3390/cells10020206.

(15) Chen, P.; Nirula, A.; Heller, B.; Gottlieb, R. L.; Boscia, J.; Morris, J.; Huhn, G.; Cardona, J.; Mocherla, B.; Stosor, V.; Shawa, I.; Adams, A. C.; Van Naarden, J.; Custer, K. L.; Shen, L.; Durante, M.; Oakley, G.; Schade, A. E.; Sabo, J.; Patel, D. R.; Klekotka, P.; Skovronsky, D. M. SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19. *N. Engl. J. Med.* **2020**, *384* (3), 229–237. https://doi.org/10.1056/NEJMoa2029849.

(16) Weinreich, D. M.; Sivapalasingam, S.; Norton, T.; Ali, S.; Gao, H.; Bhore, R.; Musser, B. J.; Soo, Y.; Rofail, D.; Im, J.; Perry, C.; Pan, C.; Hosain, R.; Mahmood, A.; Davis, J. D.; Turner, K. C.; Hooper, A. T.; Hamilton, J. D.; Baum, A.; Kyratsous, C. A.; Kim, Y.; Cook, A.; Kampman, W.; Kohli, A.; Sachdeva, Y.; Graber, X.; Kowal, B.; DiCioccio, T.; Stahl, N.; Lipsich, L.; Braunstein,

N.; Herman, G.; Yancopoulos, G. D. REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *N. Engl. J. Med.* **2020**, *384* (3), 238–251. https://doi.org/10.1056/NEJMoa2035002.

(17) Singh Tomar, P. P.; Arkin, I. T. SARS-CoV-2 E Protein Is a Potential Ion Channel That Can Be Inhibited by Gliclazide and Memantine. *Biochem. Biophys. Res. Commun.* **2020**, *530* (1), 10–14. https://doi.org/https://doi.org/10.1016/j.bbrc.2020.05.206.

(18) Das, G.; Das, T.; Chowdhury, N.; Chatterjee, D.; Bagchi, A.; Ghosh, Z. Repurposed Drugs and Nutraceuticals Targeting Envelope Protein: A Possible Therapeutic Strategy against COVID-19. *Genomics* **2021**, *113* (1, Part 2), 1129–1140. https://doi.org/https://doi.org/10.1016/j.ygeno.2020.11.009.

(19) Monje-Galvan, V.; Voth, G. A. Molecular Interactions of the M and E Integral Membrane Proteins of SARS-CoV-2. *bioRxiv* **2021**, 2021.04.29.442018. https://doi.org/10.1101/2021.04.29.442018.

(20) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *Biophys. J.* **2021**, *120* (6), 1097–1104. https://doi.org/10.1016/j.bpj.2020.10.048.

(21) 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**, 2020.03.25.008904. https://doi.org/10.1101/2020.03.25.008904.

(22) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319–2327. https://doi.org/https://doi.org/10.1002/jcc.21787.

(23) Gapsys, V.; de Groot, B. L.; Briones, R. Computational Analysis of Local Membrane Properties. *J. Comput. Aided. Mol. Des.* **2013**, *27* (10), 845–858. https://doi.org/10.1007/s10822-013-9684-0.

(24) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718. https://doi.org/https://doi.org/10.1002/jcc.20291.

(25) Huang, J.; MacKerell Jr, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/https://doi.org/10.1002/jcc.23354.

(26) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.

(27) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(29) Birdsall, C. K.; Langdon, A. B. *Plasma Physics via Computer Simulation*; Series in Plasma Physics and Fluid Dynamics; Taylor & Francis, 2004.

(30) Fliege, J.; Svaiter, B. F. Steepest Descent Methods for Multicriteria Optimization. *Math. Methods Oper. Res.* **2000**, *51* (3), 479–494. https://doi.org/10.1007/s001860000043.

(31) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069–4074. https://doi.org/10.1063/1.449071.

(32) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.