

# **Computational Study of Ions and Water Permeation and Transportation Mechanisms of the SARS-CoV-2 Pentameric E Protein Channel**

Yipeng Cao,<sup>1,2</sup> Rui Yang,<sup>3</sup> Wei Wang,<sup>1</sup> Imshik Lee,<sup>4</sup> Ruiping Zhang,<sup>1</sup> Wenwen Zhang,<sup>1</sup> Jiana Sun,<sup>1</sup> Bo Xu,<sup>1,5</sup> # and Xiangfei Meng<sup>2</sup> #

1. Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin, 300060 P.R. China
2. National Supercomputer Center in Tianjin, 300457 P.R. China
3. Department of Infection and Immunity, Tianjin Union Medical Center, Nankai University Affiliated Hospital. 300031, P.R. China
4. College of Physics, Nankai University, Tianjin, 300071 P.R. China
5. Center for Intelligent Oncology, Chongqing University School of Medicine and Chongqing University Cancer Hospital, Chongqing 400030, P.R.China

# Corresponding Author

## **Abstract**

Coronavirus disease 2019 (COVID-19) is caused by a novel coronavirus (SARS-CoV-2) and represents the causative agent of a potentially fatal disease that is of public health emergency of international concern. Coronaviruses, including SARS-CoV-2, encode an envelope (E) protein, which is a small, hydrophobic membrane protein; the E protein of SARS-CoV-2 has high homology with that of severe acute respiratory syndrome coronavirus. (SARS-CoV) In this study, we provide insights into the function of the SARS-CoV-2 E protein channel and the ion and water permeation mechanisms on the basis of combined in silico methods. Our results suggest that the pentameric E protein promotes the penetration of monovalent ions through the channel. Analysis of the potential mean force (PMF), pore radius and diffusion coefficient reveals that Leu10 and Phe19 are the hydrophobic gates of the channel. In addition, the pore demonstrated a clear wetting/dewetting transition with monovalent cation selectivity under transmembrane voltage, which indicates that it is a hydrophobic voltage-dependent channel. Overall, these results provide structural-basis insights and molecular-dynamic information that are needed to understand the regulatory mechanisms of ion permeability in the pentameric SARS-CoV-2 E protein channel.

## **Introduction**

COVID-19 is a severe and highly contagious respiratory illness that was first reported in China in early December 2019. Subsequently, the virus has spread worldwide. As of May 6, 2020, millions of cases have been confirmed, and hundreds of thousands have died. The World Health Organization (WHO) announced a global pandemic for COVID-19 in March 2020. In addition to the hazards of the disease itself, it has also led to a severe turbulence in international financial markets and may cause serious consequences such as a financial crisis. COVID-19 is a disease caused by a new coronavirus named SARS-CoV-2. It is speculated that it originated from bats and was transmitted to humans through an intermediate host (some kind of wildlife). Its

symptoms include fever, general malaise, dry cough, shortness of breath, and respiratory distress. Comparing similar diseases, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), which are also caused by coronaviruses, the mortality rates for SARS and MERS were 10% and 36%, respectively <sup>1</sup>. Currently, COVID-19 has a mortality rate of 1 to 10% in different countries but appears to be more contagious than SARS and MERS <sup>2 3 4 5</sup>.

Similar to other coronaviruses, SARS-CoV-2 is a long positive-sense, long single-stranded (30 kb) RNA virus. The structure of different human coronaviruses (HCoVs) is similar. The viral genome is packed by nucleocapsid (N) proteins, forming a helicoidal nucleocapsid protected by a lipid envelope <sup>6</sup>. Several viral proteins, including the spike (S), envelope (E), and membrane (M) proteins, are embedded within a lipid envelope <sup>7 8</sup>. Studies have shown that S, M, and N proteins play important roles in receptor binding and virion budding. For example, the M protein participates in virus germination and interacts with the N and S proteins. The S protein has immune recognition sites and can be used to design vaccines <sup>9</sup>.

Currently, the importance of the E protein has not been fully revealed. Evidence suggests that the E protein maintains its morphology after virus assembly by interacting with the M protein <sup>10 11 12</sup>. When an E protein gene mutation occurs, it promotes apoptosis. Recent studies have shown that coronaviruses have a viroporin that can self-assemble into a pentameric structure and have ion selectivity. When there is a transmembrane voltage, the ion channel characteristics of viroporins are more significant. In addition, the Asn25Ala and Val18Phe mutations could destroy ion channel activity <sup>13</sup>. This indicates that the E protein may play an important role in regulating the ion equilibrium inside and outside the viral envelope.

The ion channel activity of the E protein can lead to increased levels of the inflammatory cytokines IL-1 $\beta$ , TNF and IL-6 in the lungs, leading to the occurrence of an "inflammatory storm" <sup>14</sup>. This plays a key role in the progression of the disease and may cause the patient's condition to suddenly deteriorate and lead to death. In addition, its evolutionary conservatism may be an important cause of viral cross-host infection <sup>15 16 17</sup>.

Although previous studies have suggested that the E protein of coronaviruses such as SARS and MERS oligomerizes and has ion permeability, the specific mechanism of ion permeability and channel properties remain to be explored due to the lack of a crystal structure for the E protein. In 2014, Li et al.<sup>17</sup> identified the SARS-CoV E protein monomer structure. In 2018, Surya et al. extracted the pentamer structure of the SARS E protein by nuclear magnetic resonance (NMR)<sup>18</sup>, which provided strong support for the study of the SARS-CoV-2 E protein pentamer ion permeability mechanism.

In this study, we obtained the amino acid sequence of the SARS-CoV-2 E protein from the National Center for Biotechnology Information (NCBI) database<sup>19</sup>. The E protein pentamer model of SARS-CoV-2 was built by using the homology modeling method, and the reasonableness of the model was evaluated. Subsequently,  $\mu$ s-level molecular dynamics (MD) simulations were performed to evaluate the pentamer's stability in the membrane environment. We tried to use potential mean force (PMF) to reveal the permeability of different physiological ions and water molecules in the pores of the E protein pentamer. The characteristics of the pentameric channel were analyzed by combining the channel diffusion coefficient and geometric properties. In addition, computational electrophysiology was applied for different transmembrane voltages of the system to reveal the effect of the voltage on the ion permeability of the pentameric E protein.

Overall, exploring the mechanisms of the pentameric SARS-CoV-2 E protein not only provides valuable insights into the conduction of the channel but also has important implications for our understanding of the difference between SARS-CoV-2 and other coronaviruses.

## **Results**

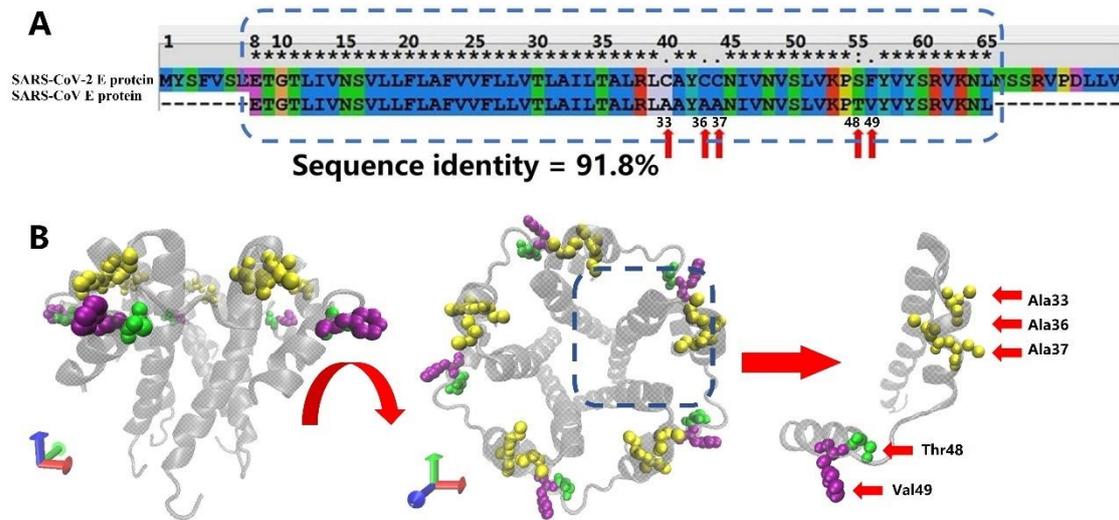


Figure 1. A, The sequence alignment for the E protein of SARS-CoV-2 and SARS-CoV. The SARS-CoV E protein sequence is from the protein data bank (PDB) database (PDB ID: 5X29). The SARS-CoV-2 E protein sequence was obtained from NCBI. The sequences in the rectangular box are used for homology modeling. The red arrows mark the different amino acids between the SARS-CoV-2 and SARS-CoV E proteins. B, Two views of the E protein model of SARS-CoV-2. The different amino acids between SARS-CoV and SARS-CoV-2 are marked with red arrows.

#### Sequence alignment and homology modeling

Figure 1A shows the sequence alignment between the E proteins of two human coronaviruses (SARS-CoV and SARS-CoV-2 sequence) made by Clustal X software. The blue dotted rectangle in Figure 1B is the area of SARS-CoV-2 that corresponds to the crystal structure of the SARS-CoV E protein. The E protein amino acid sequences of these two coronaviruses are highly homologous, and the overall and transmembrane (TM) region similarity reached 91.8% and 100%, respectively.

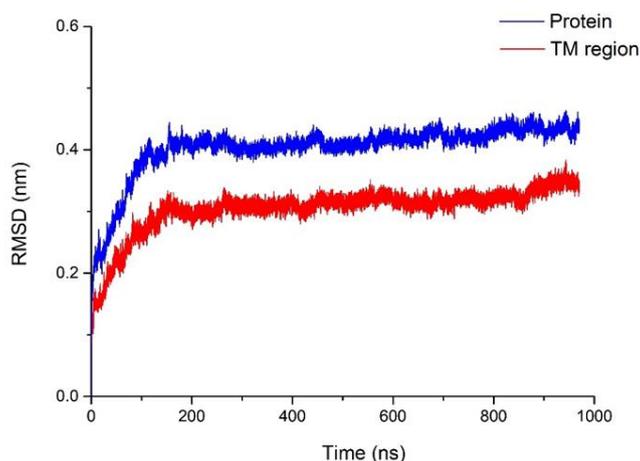


Figure 2. Structural stability of the SARS-CoV-2 E protein. The blue and red curves represent the root-mean-square deviation (RMSD) of the whole pentameric E protein and TM region, respectively.

#### SARS-CoV-2 E protein pentamer homology model evaluation

The Rampage online program <sup>20</sup> was used to evaluate the accuracy of the SARS-CoV-2 pentameric E protein model. More than 98.9% of the amino acids are within the acceptable range, suggesting that the SARS-CoV-2 E protein is similar to the SARS E protein NMR model. Subsequently, a 1000 ns (1  $\mu$ s) MD simulation was performed to evaluate the stability of pentameric E protein embedded in the membrane environment. The root-mean-square deviation (RMSD) of the model is shown in Figure 2, and the red and blue curves represent the whole protein and TM region, respectively. In the first 200 ns, the RMSD continued to rise, indicating that the model needed longer optimization (compared with other ion channels) to reach the pentameric structure equilibrium. During the last 800 ns, the curves plateaued. The RMSD of the whole protein and the TM region converged at  $\sim$ 0.4 and  $\sim$ 0.3 nm, respectively. There is a 0.1 nm difference between the whole protein and TM region. These findings are consistent with other ion channel data that show that the TM region has a higher stability than the other parts of the membrane protein <sup>21</sup>.

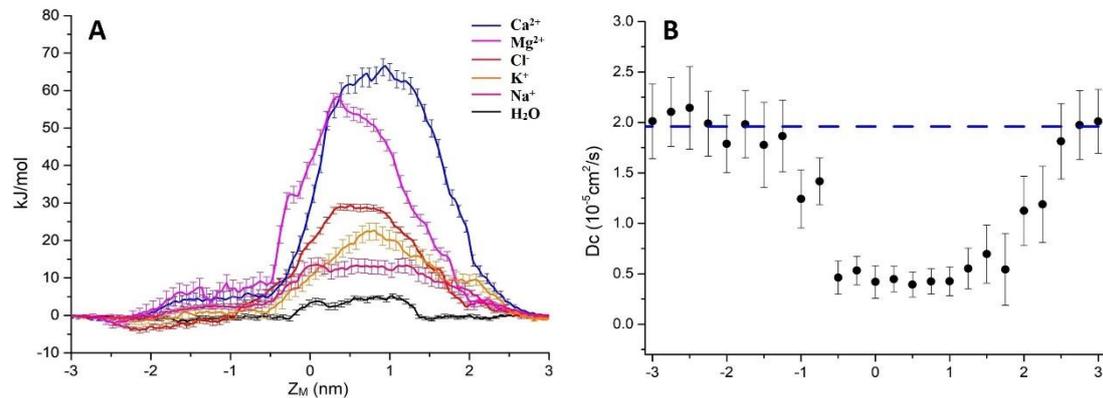


Figure 3. A, PMF profiles of single ions ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{H}_2\text{O}$  represented by blue, magenta, red, yellow, pink and black, respectively) as a function of position along the pentameric SARS-CoV-2 E protein pore axis. B, The diffusion coefficient profile. The dashed line is the experimental value of a single  $\text{K}^+$  diffusion coefficient in water,  $D_c=10^{-5} \text{ cm}^2/\text{s}$ . Error bars were estimated by bootstrapping and are the same color as the corresponding line.

#### Water and ion permeability

The permeability of the SARS-CoV-2 pentameric E protein channel is very important for understanding the replication ability of viruses in cells and how they are secreted into the extracellular medium<sup>22</sup>. It is possible to examine a profile of the free energy of a single ion or water as a function of its position along the pore axis by calculation of PMF. In this simulation, the ions or water molecules are restrained to a continuous position along the z-axis and move freely in the pore's xy plane. Moreover, other parts of the system (proteins, ions, water molecules, and lipids) can move freely and reach equilibrium.

The PMF of important physiological ions ( $\text{Mg}^{2+}$ ;  $\text{Ca}^{2+}$ ;  $\text{Cl}^-$ ;  $\text{K}^+$  and  $\text{Na}^+$ ) and the water molecules as a function of their position along the pore (z)-axis were calculated separately. Figure 3A shows the PMF of ions and water molecules permeating through the SARS-CoV-2 E protein pentamer pore.  $Z_M$  is defined as the axial direction along the pore, which is from -3 ~ 3 nm (the length of umbrella sampling is 6 nm in total).

From the PMF curve, we found that the energy barriers of monovalent and divalent ions have a significant difference. The maximum energy barriers of the two divalent ions  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  are 60 kJ/mol and 70 kJ/mol, respectively. In contrast, the PMFs of the monovalent ions  $\text{Na}^+$  and  $\text{K}^+$  are  $\sim 12$  kJ/mol and  $\sim 20$  kJ/mol, respectively, while the  $\text{Cl}^-$  energy barrier is  $\sim 30$  kJ/mol, which is significantly smaller than that of  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ . From an energy perspective, the SARS-CoV-2 pentameric E protein is almost impermeable to divalent ions. This confirms the previous hypothesis that the SARS-CoV and MERS-CoV E protein channel is a monovalent cation channel<sup>23 24</sup>. The maximum energy barrier in descending order is as follows:  $\text{Na}^+ < \text{K}^+ < \text{Cl}^- < \text{Mg}^{2+} < \text{Ca}^{2+}$ . Additionally, the barrier of water molecules is only  $\sim 5$  k/mol, which is similar to the energy threshold of free diffusion in liquid. Interestingly, there is no obvious peak in the PMF curve for  $\text{Na}^+$  ions and water molecules. This result indicates that the E protein pentamer pores are not permeable to divalent ions. The difference in the energy barrier between the monovalent cations is large, suggesting that the channel has selective permeability for specific ions.

To explore the permeability of the channel, we calculated the diffusion coefficient ( $D_C$ ) using a single  $\text{K}^+$  ion as a probe. By fitting the  $\text{K}^+$  mean square displacement (MSD) with the Einstein equation ( $\text{MSD} = 2D_C t$ ), we obtained the diffusion coefficient. As shown in Figure 3B, the average diffusion coefficient of  $\text{K}^+$  outside the pore is  $\sim 2.0 \times 10^{-5} \pm 0.4$   $\text{cm}^2/\text{s}$ , which is consistent with the experimental data of  $\text{K}^+$  in the liquid environment<sup>25</sup>. However, the  $D_C$  decreases by approximately 70% when the  $\text{K}^+$  probe is in the channel pore. The  $D_C$  curve of the internal channel pore has a flat bottom and is similar to the PMF. The mid-value of  $D_C = 0.35 \times 10^{-5} \pm 0.2$   $\text{cm}^2/\text{s}$  indicates that there is a large obstacle for probe diffusion in the inner pore, showing a similar value to other narrow hydrophobic pores<sup>26 27</sup>.

Next, we used the solubility-diffusion equation (Eq 1)<sup>28 29</sup> to evaluate the permeability coefficient (P) between two different ions.

$$\frac{1}{P} = \int_{z_1}^{z_2} \frac{\exp\left(\frac{PMF(z)}{k_b T}\right)}{D(z)} dz \quad \text{Eq 1}$$

In Eq 1,  $PMF(z)$  and  $D(z)$  are functions of PMF and DC along the  $z$ -axis,  $z_1$  denotes a position on one side of the pore and  $z_2$  denotes a position on the other side.  $K_b$  is the Boltzmann constant, and  $T$  is the simulation temperature. The part of the PMF in the TM region is approximate to the graph of the quadratic function, so the  $PMF(Z)$  uses quadratic function fitting. The linear fitting of the corresponding  $D_C$  gives the function  $D(z)$ . The ratios of the  $P$  for  $Na^+$ ,  $K^+$  and  $Cl^-$  in the maximum energy barrier region were simply evaluated.  $P_{Na^+}/P_{K^+} = 22.42$ ,  $P_{K^+}/P_{Cl^-} = 5.82$ , and  $P_{H_2O}/P_{Na^+} = 402.57$ .

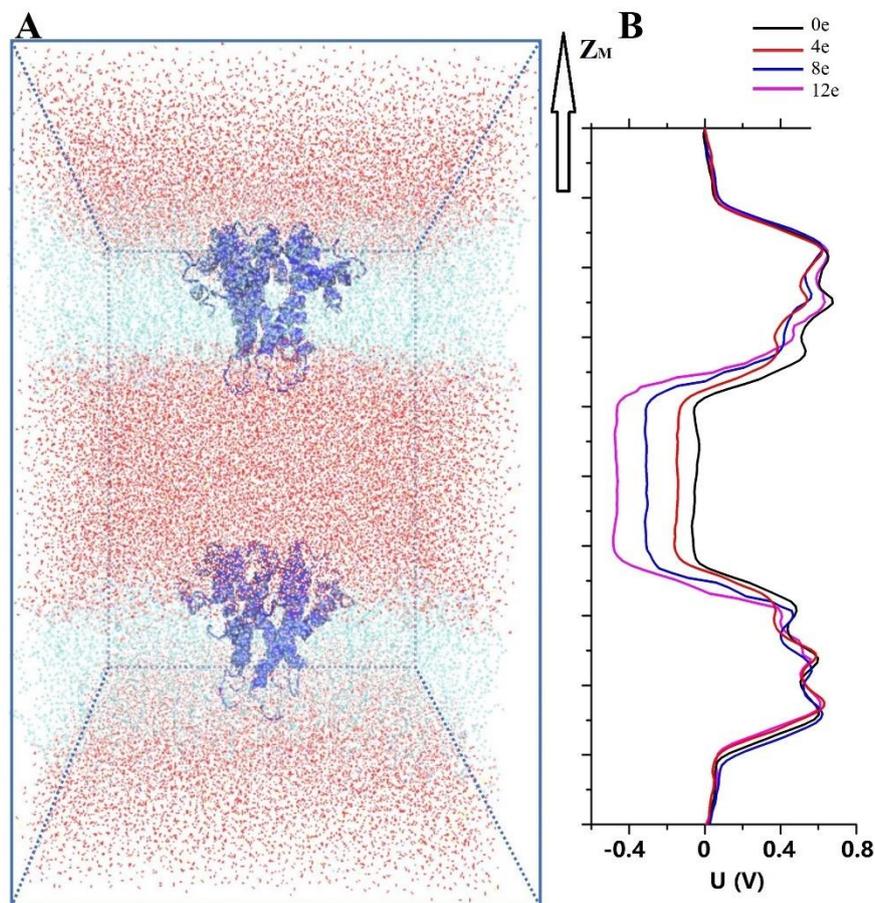


Figure 4. A, A sandwich simulation system of SARS-CoV-2 at 0.15 M NaCl concentration, and this includes the SARS-CoV-2 pentameric E proteins, cell membrane, ions and water molecules. B, Average time of the electrostatic potential  $\Delta U$  along the  $z$ -axis, arising from imbalances  $\Delta Q$  between the elementary charges of 0 and 12(color curve).

## Computational electrophysiology

Previous electrochemical experiments have shown that the transmembrane voltage has an important effect on the pore permeability for the E protein in both SARS-CoV and MERS-CoV<sup>23 24</sup>. In addition, viroporins and similarly nanopores are sensitive to transmembrane voltage, causing electrowetting of the pores and changing the tension on the hydrophobic surface inside the pores, which is enough to functionally open the channels<sup>30 31 32 33</sup>. We used the computational electrophysiology first proposed by Kutzner et al. This approach expands the membrane protein structure into a "sandwich" and simulates the transmembrane voltage by adjusting the number of ions at different layers. It can directly simulate and analyze the ion penetration in the presence of transmembrane voltage at an atomic resolution. The sandwich structure of the SARS-CoV-2 E protein pentamer is presented in Figure 4A. A total of 4 models were established to simulate the charge differences between the two sides of the channel ( $\Delta Q = 12e, 8e, 4e,$  and  $0e$ ). Using the method described above, the transmembrane voltage was calculated as  $\Delta U = \sim 0.45 \text{ V}, 0.3 \text{ V}, 0.15 \text{ V},$  and  $0 \text{ V}$ , respectively (Fig 4B).

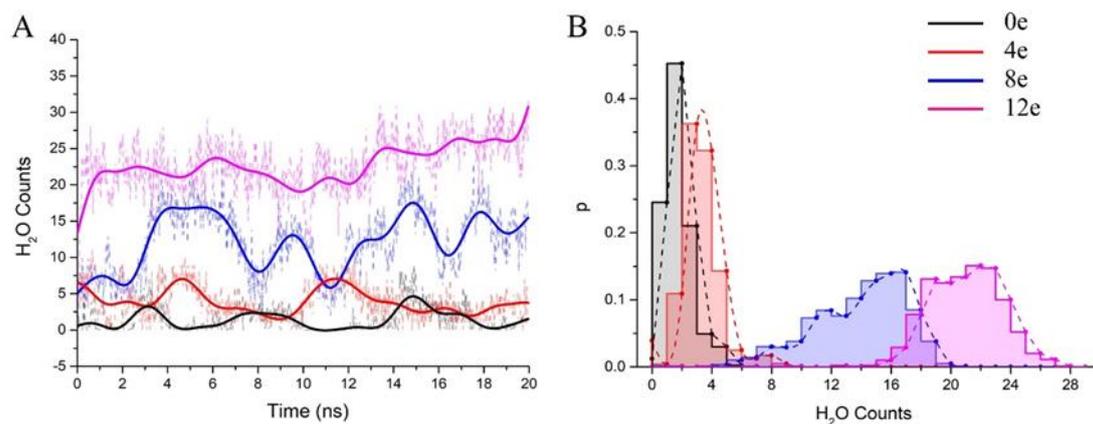


Figure 5. A, The change in water molecule counts in the inner pore of the channel under different transmembrane voltages. The scatter represents the absolute counts of water molecules, and the color curve represents after smoothing ( $n = 25$ ) using the Fast Fourier Transformation (FFT) method. B, Normalized probability histogram of

the water counts in the TM region for 20 ns simulations. The different colors represent different models: 0e (blue), 4e (red), 8e (blue) and 12e (magenta).

#### Effect of transmembrane voltage

Trick et al. indicated that the behavior of water molecules in pores can be used as a reasonable method to study the ion permeability of channels <sup>27</sup>. Many works have shown that water influx into hydrophobic pores may help ions overcome the energy barrier on hydrophobic surfaces <sup>34 35</sup>. Overall, we could speculate the ion penetrability of the E protein pentamer by the statistical number of water molecular counts in the inner pore.

To accurately assess the behavior of water molecules in the SARS-CoV-2 E protein pentamer pores, we performed six 20 ns simulations for each transmembrane voltage, and the total simulation time reached  $\sim 0.5 \mu\text{s}$ . Subsequently, the water molecules at different transmembrane voltages were counted (the statistical data of the changes in the molecule counts are shown in the Supplementary Word Film). We extracted representative water molecule diagrams for each voltage, as shown in Figure 5A. When  $\Delta Q = 0\text{e}$ , the average time for the channel to keep wetting is  $\sim 3 \text{ ns}$ .  $\Delta Q = 4\text{e}$  is slightly longer ( $\sim 5 \text{ ns}$ ). The absolute number of water molecules in the pore is small, with the average ranging from 0 to 5 (Fig 5B). Although the pores are functionally open for water permeation, due to the small stream of water, they may not be able to maintain sufficient electrical wetting, resulting in ions that cannot permeate unimpeded. We speculate that at the maximum physiological voltage ( $\sim 0.2 \text{ V}$ ), the SARS-CoV-2 pentameric E protein pore can only permeate a few water molecules.

The number of water molecules increased significantly when  $\Delta Q = 8\text{e}$  or  $12\text{e}$  ( $\Delta U = \sim 0.3 \text{ V}$ ,  $0.45 \text{ V}$ ), and the water distributions were 4-20 (8e) and 15-27 (12e) (The process of water permeation see supplementary Movie 1). The pores remained in a wetting state. The experiment shows that the maximum  $\Delta U = 0.45 \text{ V}$  is not enough to cause electrical breakdown of the lipid bilayer <sup>36</sup>, indicating that the SARS-CoV-2 E protein channel is activated by transmembrane voltage. Additionally, we observed that a few water molecules were able to be cotransported with  $\text{Na}^+$  and  $\text{K}^+$  during the

simulations (See Supplementary Movie 2 and 3). Interestingly, as the voltage increased, the permeability of water molecules also increased synchronously, showing a periodic change regardless of the magnitude of the transmembrane voltage (even if  $\Delta U = 0$ ). The calculation suggests that the SARS-CoV-2 E protein pentamer channel has obvious wetting/dewetting transitions and that it has the typical characteristics of voltage-dependent hydrophobic channels.

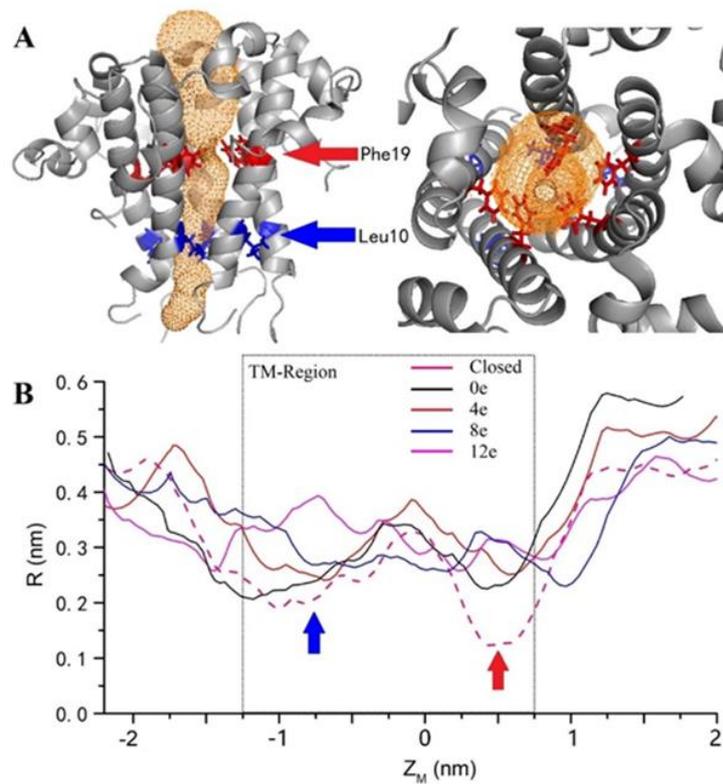


Figure 6. A, The channel structure in the closed state. The red and blue arrows indicate the two residues with the largest radius changes at different voltages. B, The variation in the inner pore radius. The z axis shows the distance from the center of mass. The black, red, blue, and magenta solid curves represent  $\Delta Q=0e$ ,  $4e$ ,  $8e$ , and  $12e$ , respectively, and the pink dashed curve represents the closed state. The calculated channel as measured by CAVER 3.01<sup>37</sup>.

#### Pore conformation analysis

By calculating the channel radius at different voltages ( $\Delta U = \sim 0.45$  V,  $0.3$  V,  $0.15$  V,  $0$

V and 0 V (closed state)) (Fig 6B), we found that in the pentameric E protein channel in the closed state ( $\Delta U = 0$  V), there are two positions with the minimum radius in the TM region,  $r = 0.2$  nm ( $Z_M = -0.8$  nm) and  $0.15$  nm ( $Z_M = -0.5$  nm), corresponding to Leu10 and Phe19 (blue and red arrows, respectively). Phe19 and Leu10 are located at the top and bottom of the pore, respectively. The two positions show a candidate gated area of the channel. Figure 6A shows that these two locations have obvious "bottlenecks", which are not large enough for water permeation. This was also confirmed by the calculation of the channel radius on different transmembrane voltages. The radius of the pore obviously increased with transmembrane voltage at the two locations. In the open state ( $\Delta U=0$  V), the pore radius increased to  $2$  nm and  $2.2$  nm. When the transmembrane voltage increased to  $\Delta U = \sim 0.15$  V and  $0.3$  V, the radius of the pore at the Leu10 position was  $\sim 0.23$  nm and  $0.28$  nm, and the radius of the pore at Phe19 was  $\sim 0.25$  nm and  $0.33$  nm, respectively. The voltage had a greater influence on Phe19 than on Leu10. However, for  $\Delta U = 0.45$  V, the radius of the pore was still  $\sim 0.33$  nm at Phe19 but had expanded to  $0.4$  nm at Leu10, indicating that the large transmembrane voltage may make Leu10 excessively available, which may cause the unidirectional permeation to be destroyed. The narrowest locations of the inner pore are consistent with the energy barrier. Apparently, the radius of the inner pore showed that Leu10 and Phe19 of the SARS-CoV-2 E protein are candidate hydrophobic gates.

## Discussion

COVID-19 is an unprecedented threat to humanity. To understand the infection, the replication mechanism of the virus in the human body is particularly important. Currently, research on the pathogenic mechanism and infection route of COVID-19 mainly focuses on the S protein<sup>38 39</sup>, but research on the E protein is lacking. Previous reports note that the E protein pentamers from different coronavirus subtypes (SARS-CoV, MERS-CoV) may be viroporins, which play an important role in the virus life cycle<sup>40</sup>. Deleting the encoding gene of the E protein will greatly reduce the

replication and pathogenicity of coronaviruses. Moreover, the pentameric E protein channel activity is necessary for inflammasome activation and is the determinant of coronavirus virulence<sup>41</sup>. This study explored the permeability mechanism of the SARS-CoV-2 pentameric E protein. We used the NMR conformation of the SARS-CoV E protein pentamer to build the SARS-CoV-2 E protein pentamer model by homology modeling. A combination of MD simulations, PMF, diffusion coefficient calculations, computational electrophysiology analyses, and channel pore geometry analyses was used. The results were mutually corroborated and characterized the TM dynamics of the SARS-CoV-2 pentameric E protein channel. It was found that the pentamer is a monovalent cation-selective voltage-dependent hydrophobic channel, and the two amino acids Leu10 and Phe19 are candidate hydrophobic gates.

The SARS-CoV-2 E protein pentamer has been considered an unknown nature ion channel. There is significant similarity in the TM region sequences (reaching 91.8%) and high gene homology between the SARS-CoV and SARS-CoV-2 E proteins; five different residues are all located at the region of interaction between the protein and membrane, while the sequences of the TM region are completely identical. The highly conserved TM region also suggests similar E protein functions in different coronavirus subtypes. As the RMSD of the E protein pentamer during the 1  $\mu$ s MD simulation, the structure has a certain flexibility, showing that its geometric deformation can make the channel permeable to ions and water molecules. Figure 2 shows that the TM region of the SARS-CoV-2 E protein remains relatively stable during the simulation, indicating the stability of the pentamer in the membrane environment, which is consistent with previous studies of SARS-CoV and MERS-CoV.

The free energy calculation of the ions permeating through the SARS-CoV-2 pentameric E protein channel strongly suggests that the pore has selection permeability for monovalent ions. The maximum energy barriers of Na<sup>+</sup>, K<sup>+</sup>, and Cl<sup>-</sup> are 15 kJ/mol, 20 kJ/mol, and 30 kJ/mol, respectively. The comparison of the P values for Na<sup>+</sup>, K<sup>+</sup> and Cl<sup>-</sup> shows that the P of Na<sup>+</sup> is 22 times greater than that of K<sup>+</sup>, while the P of K<sup>+</sup> is 6 times greater than that of Cl<sup>-</sup>. Although the computational

permeability coefficient is two times higher or lower than that observed experimentally for SARS-CoV, it still shows a very large gap in the permeation between ions and water molecules. Additionally, there is no obvious peak in the PMF when  $\text{Na}^+$  and water molecules permeate through the channel, further supporting that  $\text{Na}^+$  and water molecules have less penetration resistance at the inner pore. This further confirmed the selectivity for  $\text{Na}^+$ . The ion selectivity of SARS-CoV-2 is different from that of SARS-CoV<sup>23</sup>. Because the maximum energy barrier of  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$  is 2 ~ 3 times higher than that of the monovalent ions, divalent ions cannot overcome such a very large energy barrier under the simulation environment. The conformational analysis indicates that it is mainly caused by the geometric size of the pore (Fig 6).

The importance of transmembrane voltage on viroporin has been found in many viruses, including HIV1 Vpu oligomers and the hepatitis C virus (HCV) p7 protein, etc.<sup>42 43 44</sup>. Although it has been observed in many experiments, this mechanism has not been revealed. Because water molecules play an important role in hydrophobic ion channels, their behavior at the inner pores under different voltages ( $\Delta U = \sim 0.45$  V, 0.3 V, 0.15 V, and 0 V) is the focus of this simulation. Computational electrophysiology enabled the determination of a conducive conformation for ions. The channel wetting/dewetting transitions appeared under different transmembrane voltages, and the radius of the pore increased. Meanwhile, the capacity to contain water molecules also increased. Further analysis of the trajectories revealed that the monovalent cations ( $\text{Na}^+$  and  $\text{K}^+$ ) will permeate through the pore with water molecules when the transmembrane voltage is  $> 0.3$  V.

Computational electrophysiology also explained that the transmembrane voltage led to an easier wetting transition for the channel. The range for keeping the channel open should be between 0.15 V and 0.45 V.  $\text{Na}^+$  and  $\text{K}^+$  could be cotransported during water permeation, which is very similar to that in many hydrophobic channels<sup>45 46 47</sup>.<sup>48</sup> Intriguingly, we did not observe  $\text{Cl}^-$  conduction, which may be because chloride ions could not overcome the energy barrier under the maximum transmembrane voltage. In addition, the change in the pore radius at different voltages highlights two

important sites, where the geometric radius changes the most, corresponding to Leu10 and Phe19. Phe19 contains a benzene ring group at the hydrophobic surface of the inner pore as a hydrophobic gate. The isopropyl group of Leu10 at the bottom of the pore prevents the reverse penetration (Fig 6A). Previous studies indicated that Asn15Val and Val25Phe mutations will make the channel dysfunctional. This may be due to the additional side chain groups in the pores caused by these mutations (especially the benzene ring in Val25Phe), making 1) the radius of the pores decrease and 2) the hydrophobicity of the pores increase. In short, these mutations may increase the energy barrier, negatively affecting the wetting transition and causing the channel to be functionally closed.

Therefore, the SARS-CoV-2 E protein pentamer is a voltage-dependent hydrophobic channel. We propose that the E protein may play an essential role in the virus infection and replication processes through the following mechanisms: 1) The monovalent selective permeation of the E protein pentamer ion channel may change the intracellular pH, providing a suitable microenvironment for virus replication. 2) Selective permeation can form a transmembrane voltage, creating feedback regulation and maintaining an intracellular microenvironment that is suitable for viral growth. 3) The disintegration of the ion equilibrium of the intracellular area affect the charges coming from the cell through ion channels in the cell membrane, changing the pH and making it easier for the virus to fuse with the cell membrane. This is the infection mechanism of the avian coronavirus as well as some types of influenza viruses<sup>49 50</sup>. 4) It has a certain signal transduction function.

## **Conclusion**

Although the pentameric SARS-CoV-2 E protein crystal structure is not yet available, the channel conformation was built by homology modeling from the SARS-CoV E protein pentamer. We characterized and estimated the channel's stability, energy barrier, voltage dependence and geometric properties by using MD simulations. Our simulation results demonstrate that the SARS-CoV-2 E protein pentamer is a

voltage-dependent hydrophobic channel with monovalent cation selectivity. Water molecules and monovalent cations can penetrate through the channel spontaneously under a transmembrane voltage. Differences in the penetrability of E protein channels may be responsible for the infectivity, inflammatory response and incubation period of SARS-CoV-2, which is different from that of SARS-CoV and MERS-CoV. The results further suggested that Leu10 and Phe19 are hydrophobic gates for ion regulation and permeability. Our research reveals the nature of the SARS-CoV-2 E protein pentamer, providing theoretical support for future research on vaccines against E proteins from SARS-CoV-2 and other coronaviruses.

## **Experimental methods**

### Preparation of the SARS-CoV-2 pentameric E protein-membrane simulation system

The amino acids sequence of the human coronaviruses E proteins were downloaded from the National Center for Biotechnology Information (NCBI). The Cluster-X software was used for Multiple Sequence Alignment (MSA) for the new human coronavirus and the SARS-CoV E protein. The model was built for the SARS-CoV-2 E protein using the SWISSMODEL web server<sup>51</sup>. As the SARS E protein (PDB ID: 5X29)<sup>18</sup> pentameric structure is the closest E protein to SARS-CoV-2, it was used as a template for building the SARS-CoV-2 E protein. The Rampage web tool was used to evaluate the rationality of the SARS-CoV-2 E protein model. Then, the model was inserted in a DOPC/POPC,1:1 lipid bilayer via pre-equilibrium. The simulation system was set as a  $11 \times 11 \times 12 \text{ nm}^3$  cubic box by CHARMM-GUI<sup>52</sup>, comprising the whole pentameric E-protein, 330 lipid molecules, ~30,000 TIP3P water molecules, and concentration is 0.15 M  $\text{Na}^+$  and  $\text{Cl}^-$  ions. Resulting in a simulation system size of ~120,000 atoms.

### MD Simulaiton

To obtain the equilibrated pentameric SARS-CoV-2 E protein model, Charmm36 all-atom force field<sup>53</sup> was chosen for MD simulation. The MD time step was set at 2 fs. Electrostatic interactions were described using the Particle Mesh Ewald (PME)

algorithm<sup>54</sup> with a cut-off of 1.2 nm. The LINCS algorithm<sup>55</sup> was used to constrain the bond lengths. The pressure was maintained semi-isotropically at 1 bar at both x and y directions using the Perinello-Rahman barostat algorithm<sup>56</sup> and the system temperature was maintained at 310 K by the Nose-Hoover thermostat<sup>57</sup>. Then, a 1000 ns (1  $\mu$ s) MD simulations were performed for the E protein pentameric.

### Umbrella sampling

The initial system for umbrella sampling simulations was derived from the equilibrated pentameric SARS-CoV-2 E protein mentioned above. A single ion that maintain physiological activity ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ) or water molecule was placed at successive positions along the central pore axis by using GROMACS pull code. Energy minimization was performed before simulation for optimizing the water and ions position. The reaction coordinate defined from z +3 to -3 nm with the mass center at z = 0 nm, with a spacing of 0.1 nm between successive windows, resulting in 60 umbrella sampling simulation systems. The probe ion or water molecules were harmonically restrained by a force constant of 2000 kJ mol/nm<sup>2</sup> same with the pore direction. Each window was performed a 5 ns total umbrella sampling simulation. The initial 2 ns for system equilibration, then a subsequent 3 ns was applied for analysis. The PMFs were computed by the weighted histogram analysis method (WHAM)<sup>58</sup>, and the profile was generated by Gromacs protocol 'g\_wham'<sup>59</sup>. Bootstrap analysis (N = 50) was used to estimate statistical error and the '-cycl' parameter was used to make the value equal.

### Diffusion-coefficient calculation

The Diffusion-coefficient was calculated using the method described by Shirvanyants et al.<sup>60</sup>. We restrained the E-protein pentamer helix backbones and made the side chains, membrane, water and ions free to move. Due to the ion selectivity is related to the side chain of inner pore<sup>61</sup>, the backbone restraint maintained the E-protein pentamer tertiary structure but the side-chain flexibility in the inner pore was not influenced. A  $\text{K}^+$  as the probe ion, the calculation system was used in the same

manner as for umbrella sampling. The ion mean-square displacement (MSD) was calculated along the pore's Z-axis. A total of 25 simulation systems were obtained, each widows interval distance of the  $K^+$  being set as 0.24 nm. The umbrella restraint was used to maintain the  $K^+$  ion's position on the x-y plane of the pore. The Einstein equation  $MSD=2D(z)t$  was used to calculate the diffusion coefficient by the protocol `g_msd`. The umbrella restraint can be disregarded for these analyses because of the restraint force was negligible compared to thermally induced RMS fluctuations.

### CE simulations

Computational electrophysiology (CE) is a good tool for simulating the ion conduction ability of a channel under transmembrane voltage. We established a sandwich structure including three parts (membrane-protein-water) as described by Kutzner et al.<sup>62</sup> as shown in Fig 4A, each water layer contained a different number of ions. Due to the imbalance of the ion distribution of the water layer, the ions gradient will produce transmembrane voltage. Specific transmembrane voltage can be applied to the simulation system by adjusting the number of ions between the water layers. In this study, we built the SARS-CoV-2 E protein sandwich system. The initial single layer system was equilibrium for 50 ns to make the structures stable. Then the single-layer system was duplicated along the z direction. The resulting system had a size of around  $11 \times 11 \times 22 \text{ nm}^3$ . The transmembrane voltage set with different ion number of each parts.

$$\Delta e = num(1)_{cl-} - num(2)_{cl-} = 0e, 4e, 8e, 12e$$

The transmembrane potential was calculated using the implementation of the Poisson equation  $\nabla^2 U = -\rho/\epsilon$  in the GROMACS `g_potential` tool. After that, six 20 ns simulations for each transmembrane voltage were executed, the total simulation time of CE is  $6 \times 20 \times 4 = 480$  ns.

The VMD software<sup>63</sup> were used to visualize MD trajectories. All of the simulations were performed using the Gromacs 5.1 package<sup>64</sup> on Tianhe I Supercomputer in the National Supercomputer Tianjin Center of China. The total simulation times were ~8

μs.

## **Acknowledgments**

This study was supported by the National Natural Science Foundation of China (NSFC-31900894 and NSFC-81974464) and The Science & Technology Development Fund of Tianjin Education Commission for Higher Education (2018KJ067). We thank National Supercomputer center – Tianjin provide computing power support.

## **Declaration of interests**

All other authors declare no competing interests.

## References

1. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523 (2016).
2. Baud, D. *et al.* Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* (2020).
3. Onder, G., Rezza, G. & Brusaferro, S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *Jama* (2020).
4. Mahase, E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. (2020).
5. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* (2020).
6. Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D. & Huang, T. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res.* **103**, 39–50 (2014).
7. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* (80-. ). **300**, 1394–1399 (2003).
8. 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).
9. Ibrahim, I. M., Abdelmalek, D. H., Elshahat, M. E. & Elfiky, A. A. COVID-19 spike-host cell receptor GRP78 binding site prediction. *J. Infect.* (2020).
  10. Corse, E. & Machamer, C. E. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J. Virol.* **74**, 4319–4326 (2000).
  11. Vennema, H. *et al.* Nucleocapsid - independent assembly of coronavirus - like particles by co - expression of viral envelope protein genes. *EMBO J.* **15**, 2020–2028 (1996).
  12. Mortola, E. & Roy, P. Efficient assembly and release of SARS coronavirus - like particles by a heterologous expression system. *FEBS Lett.* **576**, 174–178 (2004).
  13. Torres, J. *et al.* Conductance and amantadine binding of a pore formed by a lysine - flanked transmembrane domain of SARS coronavirus envelope protein. *Protein Sci.* **16**, 2065–2071 (2007).
  14. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033–1034 (2020).
  15. Graham, R. L., Donaldson, E. F. & Baric, R. S. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat. Rev.*

- Microbiol.* **11**, 836–848 (2013).
16. Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog.* **10**, (2014).
  17. Li, Y., Surya, W., Claudine, S. & Torres, J. Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* **289**, 12535–12549 (2014).
  18. Surya, W., Li, Y. & Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta (BBA)-Biomembranes* **1860**, 1309–1317 (2018).
  19. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
  20. Lovell, S. C. *et al.* Structure validation by C $\alpha$  geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins Struct. Funct. Bioinforma.* **50**, 437–450 (2003).
  21. Khalili-Araghi, F. *et al.* Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **19**, 128–137 (2009).
  22. Ruch, T. R. & Machamer, C. E. A single polar residue and distinct membrane topologies impact the function of the infectious bronchitis coronavirus E protein. *PLoS Pathog.* **8**, (2012).
  23. Wilson, L., Mckinlay, C., Gage, P. & Ewart, G. SARS coronavirus

- E protein forms cation-selective ion channels. *Virology* **330**, 322–331 (2004).
24. Surya, W., Li, Y., Verdi àB àguena, C., Aguilera, V. M. & Torres, J. MERS coronavirus envelope protein has a single transmembrane domain that forms pentameric ion channels. *Virus Res.* **201**, 61–66 (2015).
25. Hille, B. Ion channels of excitable membranes (Sinauer, Sunderland, MA). (2001).
26. Roux, B., Allen, T., Berneche, S. & Im, W. Theoretical and computational models of biological ion channels. *Q. Rev. Biophys.* **37**, 15–103 (2004).
27. Trick, J. L. *et al.* Functional annotation of ion channel structures by molecular simulation. *Structure* **24**, 2207–2216 (2016).
28. Marrink, S. J. & Berendsen, H. J. C. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *J. Phys. Chem.* **100**, 16729–16738 (1996).
29. Sun, R. *et al.* Molecular transport through membranes: Accurate permeability coefficients from multidimensional potentials of mean force and local diffusion constants. *J. Chem. Phys.* **149**, 72310 (2018).
30. Bratko, D., Daub, C. D., Leung, K. & Luzar, A. Effect of field direction on electrowetting in a nanopore. *J. Am. Chem. Soc.* **129**,

- 2504–2510 (2007).
31. Trick, J. L., Song, C., Wallace, E. J. & Sansom, M. S. P. Voltage gating of a biomimetic nanopore: Electrowetting of a hydrophobic barrier. *ACS Nano* **11**, 1840–1847 (2017).
  32. Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: structure and biological functions. *Nat. Rev. Microbiol.* **10**, 563–574 (2012).
  33. Catterall, W. A. Ion channel voltage sensors: structure, function, and pathophysiology. *Neuron* **67**, 915–928 (2010).
  34. Trick, J. L., Wallace, E. J., Bayley, H. & Sansom, M. S. P. Designing a hydrophobic barrier within biomimetic nanopores. *ACS Nano* **8**, 11268–11279 (2014).
  35. Yamashita, M. *et al.* STIM1 activates CRAC channels through rotation of the pore helix to open a hydrophobic gate. *Nat. Commun.* **8**, 1–13 (2017).
  36. Benz, R. & Zimmermann, U. 376-Relaxation studies on cell membranes and lipid bilayers in the high electric field range. *Bioelectrochemistry Bioenerg.* **7**, 723–739 (1980).
  37. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**, e1002708 (2012).
  38. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N. & Siddique, R. COVID-19 infection: origin, transmission, and characteristics of

- human coronaviruses. *J. Adv. Res.* (2020).
39. Smith, M. & Smith, J. C. Repurposing therapeutics for covid-19: Supercomputer-based docking to the sars-cov-2 viral spike protein and viral spike protein-human ace2 interface. (2020).
40. Gonzalez, M. E. & Carrasco, L. Viroporins. *FEBS Lett.* **552**, 28–34 (2003).
41. Jimenez-Guardeno, J. M. *et al.* Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. *PLoS Pathog.* **11**, (2015).
42. Mehnert, T. *et al.* Biophysical characterization of Vpu from HIV - 1 suggests a channel - pore dualism. *Proteins Struct. Funct. Bioinforma.* **70**, 1488–1497 (2008).
43. Schubert, U. *et al.* Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV - 1 - infected cells. *FEBS Lett.* **398**, 12–18 (1996).
44. Clarke, D. *et al.* Evidence for the formation of a heptameric ion channel complex by the hepatitis C virus p7 protein in vitro. *J. Biol. Chem.* **281**, 37057–37068 (2006).
45. Aryal, P., Sansom, M. S. P. & Tucker, S. J. Hydrophobic gating in ion channels. *J. Mol. Biol.* **427**, 121–130 (2015).
46. Powell, M. R., Cleary, L., Davenport, M., Shea, K. J. & Siwy, Z. S.

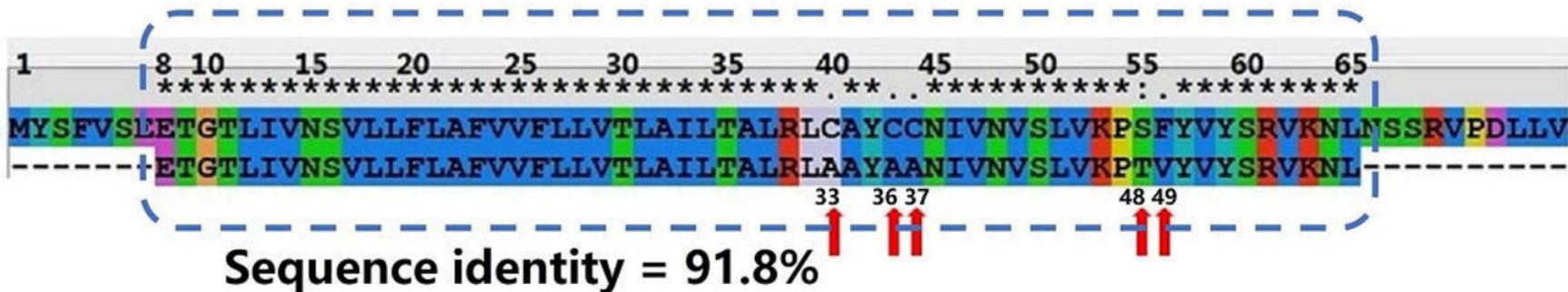
- Electric-field-induced wetting and dewetting in single hydrophobic nanopores. *Nat. Nanotechnol.* **6**, 798–802 (2011).
47. Smirnov, S. N., Vlassiouk, I. V & Lavrik, N. V. Voltage-gated hydrophobic nanopores. *ACS Nano* **5**, 7453–7461 (2011).
48. McCormack, K. *et al.* A role for hydrophobic residues in the voltage-dependent gating of Shaker K<sup>+</sup> channels. *Proc. Natl. Acad. Sci.* **88**, 2931–2935 (1991).
49. Chu, V. C., McElroy, L. J., Chu, V., Bauman, B. E. & Whittaker, G. R. The avian coronavirus infectious bronchitis virus undergoes direct low-pH-dependent fusion activation during entry into host cells. *J. Virol.* **80**, 3180–3188 (2006).
50. Costello, D. A., Whittaker, G. R. & Daniel, S. Variations in pH sensitivity, acid stability, and fusogenicity of three influenza virus H3 subtypes. *J. Virol.* **89**, 350–360 (2015).
51. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
52. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM - GUI: a web - based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
53. Vanommeslaeghe, K. *et al.* CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with

- the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
54. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
55. 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).
56. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
57. Evans, D. J. & Holian, B. L. The nose–hoover thermostat. *J. Chem. Phys.* **83**, 4069–4074 (1985).
58. Rosenbergl, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
59. Hub, J. S. g\_wham—A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Phys.* **6**, 3713–3720 (2015).
60. Shirvanyants, D. *et al.* Pore dynamics and conductance of RyR1 transmembrane domain. *Biophys. J.* **106**, 2375–2384 (2014).
61. Gouaux, E. & MacKinnon, R. Principles of selective ion transport

- in channels and pumps. *Science* (80-. ). **310**, 1461–1465 (2005).
62. Kutzner, C., Grubmüller, H., De Groot, B. L. & Zachariae, U. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys. J.* **101**, 809–817 (2011).
63. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
64. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).

**A**

SARS-CoV-2 E protein  
SARS-CoV E protein

**B**