



# Probing the formation, structure and free energy relationships of M protein dimers of SARS-CoV-2



Yipeng Cao<sup>a,b</sup>, Rui Yang<sup>c</sup>, Wei Wang<sup>a</sup>, Shengpeng Jiang<sup>a</sup>, Chengwen Yang<sup>a</sup>, Ningbo Liu<sup>a</sup>, Hongji Dai<sup>a,e</sup>, Imshik Lee<sup>d</sup>, Xiangfei Meng<sup>b,\*</sup>, Zhiyong Yuan<sup>a,\*</sup>

<sup>a</sup> Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin 300060 PR China

<sup>b</sup> National Supercomputer Center in Tianjin, 300457 PR China

<sup>c</sup> Department of Infection and Immunity, Tianjin Union Medical Center, Nankai University Affiliated Hospital, 300031, PR China

<sup>d</sup> Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, PR China

<sup>e</sup> College of Physics, Nankai University, Tianjin 300071, PR China

## ARTICLE INFO

### Article history:

Received 27 August 2021

Received in revised form 4 January 2022

Accepted 8 January 2022

Available online 11 January 2022

### Keywords:

COVID-19

SARS-CoV-2

Coronavirus

Membrane (M) protein

Free energy calculations

MMPBSA

Molecular dynamic simulations

Dimer function

Binding residues

PMF

## ABSTRACT

The M protein of the novel coronavirus 2019 (SARS-CoV-2) is the major structural component of the viral envelope and is also the minimum requirement for virus particle budding. M proteins generally exist as dimers. In virus assembly, they are the main driving force for envelope formation through lateral interactions and interactions with other viral structural proteins that play a central role. We built 100 candidate models and finally analyzed the six most convincing structural features of the SARS-CoV-2 M protein dimer based on long-timescale molecular dynamics (MD) simulations, multiple free energy analyses (potential mean force (PMF) and molecular mechanics Poisson-Boltzmann surface area (MMPBSA)) and principal component analysis (PCA) to obtain the most reasonable structure. The dimer stability was found to depend on the Leu-Ile zipper motif and aromatic amino acids in the transmembrane domain (TMD). Furthermore, the C-terminal domain (CTD) effects were relatively small. These results highlight a model in which there is sufficient binding affinity between the TMDs of M proteins to form dimers through the residues at the interface of the three transmembrane helices (TMHs). This study aims to help find more effective inhibitors of SARS-CoV-2 M dimers and to develop vaccines based on structural information.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Severe acute respiratory syndrome is caused by novel coronavirus 2019 (SARS-CoV-2). SARS-CoV-2 and its mutants have caused a large number of infections and deaths worldwide. At present, it is the most serious crisis in the global community [1,2], and there is a risk of reduced effectiveness or even ineffectiveness of vaccines in the face of novel mutations, such as the delta mutation, suggesting that we need to better understand the structural properties of SARS-CoV-2 to improve treatment [3 4 5].

In the last two years, great progress has been made in research on the SARS-CoV-2 virus particle structure: for example, Li et al. determined the high-resolution structure of whole virion particles with cryo-EM [6]. However, due to resolution limitations, the details of the properties of the structural proteins are still vague.

Improvement in treatment and prevention strategies is clearly limited by the lack of structural information about viral proteins, and a deeper understanding of these structures can lead to the development of more effective anti-coronavirus drugs and vaccines [7 8].

SARS-CoV-2 virus particles include four structural proteins—the spike (S), nucleocapsid (N), envelope (E), and membrane (M) proteins—that are important for the infectivity and pathogenicity of coronaviruses [9 10 11]. The S protein, which is prominent on the virus surface, contains immune recognition sites and is the most important protein for inducing an immune response [12 13]. The N protein immobilizes the viral RNA in the capsid [14]. The E protein appears in smaller numbers in the virus, and its role is less understood. It may form a pentamer channel to regulate the ion concentration in the host cell. [15] The M protein is a viral membrane protein that can interact with the S, N and E proteins in the viral envelope to stabilize other structural proteins [16 17 18].

The shape and structural protein homeostasis of the viral envelope are mainly determined by the M protein, which is the most

\* Corresponding authors.

E-mail addresses: [vgsplyer1@163.com](mailto:vgsplyer1@163.com) (Y. Cao), [mengxf@nsc-tj.cn](mailto:mengxf@nsc-tj.cn) (X. Meng), [zyuan@tmu.edu.cn](mailto:zyuan@tmu.edu.cn) (Z. Yuan).

abundant structural protein in the coronavirus family and is thought to be important for the formation of viral membranes [19–20]. The M protein consists mainly of the transmembrane domain (TMD), which is embedded in the viral envelope, and the C-terminal domain (CTD). The CTD of the M protein is localized inside the virion and plays an important role in the structural stability of the N protein [21]. The structure of the M protein is still unknown, mostly because the virus TMD complicates the identification of the M protein by cryo-EM [6].

M proteins interact with other structural proteins to form new virus particles in the endoplasmic reticulum-Golgi apparatus (ERGIC) [22–23]. In previous studies, the M protein was found to form functional dimers in virus particles, and its abundance was found to affect the curvature of the viral envelope. The binding between the M protein dimer and the S protein is closely related to the virulence of the virus [24]. This suggests that the M protein dimer forms a matrix in the viral envelope and could be defined as a network of molecular interactions able to maintain the balance of the whole virion [17]. It has been shown that the M proteins of SARS-CoV-2 are more immunogenic in terms of T cell response than nonstructural viral proteins [25], and directly targeting viral replication can target the M protein [26]. The M protein also antagonizes type I and III interferon production by targeting retinoic acid-inducible gene I/melanoma differentiation-associated protein 5 (RIG-I/MDA-5) signaling, enhancing viral replication and reducing the body's immunity to the virus [27].

No crystal structure has been obtained for the SARS-CoV-2 M protein that can demonstrate the dimer conformation. According to the previous cryo-EM structural protein models of SARS-COV produced by Neuman et al., M–M interactions may take various forms [19]. Yu et al used coarse-grained molecular dynamics simulations (CGMD) to simulate M–M interactions and thereby explain the role of M–M dimers in the membrane remodeling of the virus [28], and Ouzounis et al. predicted low-resolution three-dimensional models of M proteins based on Orf-3a as a structural template [29]. Although there are few reports of M protein dimers, these studies still elucidate the dimer formation process through M–M interactions to some extent.

In this study, we established a series of M protein dimer models using long timescale molecular dynamics simulations, free energy calculations and other methods to further study both qualitative and quantitative aspects of the dimerization of M proteins in a complex membrane environment. This study provides new insight into the molecular mechanisms of the structure of the M protein, its influence on other structural proteins and the morphology of the viral envelope.

## 2. Methods

### 2.1. Construction of M protein dimer models

Because the initial model has a large effect on the accuracy of the simulation results of the M protein, we chose an M protein monomer model based on AlphaFold 2 and finely optimized it with the Feig laboratory. The model was highly rated in CASP14 in 2020 [30–32]. Subsequently, the HDock server was used to construct initial models of the M–M dimer [33]. HDock is a fast protein–protein docking server that integrates a variety of bioinformatics and structural biology methods to achieve robust and accurate models [34]. It has been successfully applied to the construction of many proteins complex models [35–36].

The HDock server has established a total of 100 M protein dimer models. Previous research has shown that the formation of M protein dimers depends mainly on the surface interaction of the TMD; therefore, the initial models were selected based on

the following criteria: 1. the M protein TMD has different interacting surfaces, requiring the dimer models to cover all possible TMD binding surfaces; 2. The TMDs are on the same plane, ensuring that the transmembrane region is always embedded in the membrane; 3. the CTD planes are positively parallel, oriented in the same direction; and 4. the favored region greater than 95% when MOLPROBITY software is used to evaluate the models [37]. We identified six models by filtering on the criteria above as well as using the HDock server's docking score provided by the server as the initial model. Subsequently, we used additional protein–protein docking server tools (CLUSPRO, ZDOCK and HADDOCK) [38–39, 40] to build M protein dimer models for comparison with the HDock results.

### 2.2. System construction and simulation

The construction of the simulation systems was based on our previous study [15]. The membrane composition was obtained from the reference Mandala et al. [41] and some related studies on lipid bilayer models in ERGIC-like environments and defined as a mixture of neutral phosphatidylcholine/phosphatidylethanolamine lipids (PC/PE), charged phosphatidylinositol and phosphatidylserine lipids (PI/PS), cholesterol (CHOL) and a small portion of sphingomyelin (PSM). The proportion of each part is shown in Table 1. The software TMHMM [42] was used to determine the location of the M protein TMD part in the membrane. The Chemistry at Harvard Macromolecular Mechanics (CHARMM)-GUI server was used to build six simulation systems containing M protein dimers, pre-equilibrium membranes, water and ions, and the system size was  $\sim 17 \times 17 \times 10 \text{ nm}^3$ . The total atom counts were between 370,000 and 410,000 due to the different dimer topologies [43].

### 2.3. MD simulation

All the simulation systems had 0.15 M NaCl added to simulate the physiological environment, and based on CHARMM36, all atomic force fields were applied [44]. First, the fastest descent algorithm was used to perform the energy minimization for 20–50 K steps. Then, six 5 ns equilibrium cycles were performed by gradually closing the position limits of lipid molecules, and the pre-equilibrium systems were optimized for a 20–50 ns NPT simulation. Finally, molecular dynamics simulations were performed for each of the six M protein dimer-membrane systems at 3000 ns. The molecular dynamics (MD) time step was set to 2 fs. Electrostatic interactions were described using the particle mesh Ewald (PME) algorithm with a cutoff of 1.0–1.2 nm. The LINear Constraint Solver (LINCS) algorithm was used to constrain the bonds. The pressure was maintained semi-isotropically at 1 bar in both the x- and y-directions using the Parrinello-Rahman barostat algorithm [45], and the system temperature was maintained at 310 K using the Nose-Hoover thermostat.

**Table 1**  
The membrane components.

name		chemical formula	ratio (%)
phosphatidylcholine lipids	POPC	C <sub>42</sub> H <sub>82</sub> NO <sub>8</sub> P	15
	DPPC	C <sub>40</sub> H <sub>80</sub> NO <sub>8</sub> P	15
phosphatidylethanolamine lipids	POPE	C <sub>39</sub> H <sub>76</sub> NO <sub>8</sub> P	15
	DPPE	C <sub>40</sub> H <sub>80</sub> NO <sub>8</sub> P	15
phosphatidylinositol lipids	POPI	C <sub>43</sub> H <sub>80</sub> O <sub>13</sub>	10
	POPS	C <sub>40</sub> H <sub>76</sub> NO <sub>10</sub> P	10
cholesterol	CHOL	C <sub>27</sub> H <sub>46</sub> O	15
sphingolipids	PSM	C <sub>39</sub> H <sub>79</sub> O <sub>6</sub> 2P	5

## 2.4. Umbrella sampling and PMF

The initial systems used for umbrella sampling simulations were derived from the 3000 ns simulation of the M protein dimer mentioned above. The gmX trjconv protocol was used to keep the M protein dimer in the center of the system. One of the dimer chains was restricted, and the other chain was slowly pulled with 1000 kJ/mol energy using the method described by Lu et al [46]. After 5–10 ns, one chain was pulled to the edge of the system, and the distance between the two chains' TMD center of mass was approximately 5 nm. For the umbrella sampling simulation, from the center of mass to 3 nm, each successive window was 0.1 nm, and from 3 nm to 5 nm, the windows were 0.2 nm. Each system had approximately 30–40 windows in total. To maximize the convergence of the membrane protein systems, Domański et al.'s method [47] was applied. Each window was subjected to a long-term simulation for equilibration. A 50–70 ns simulation with a timestep of 3 fs and a force constant of 1,000 kJ/mol<sup>1</sup>nm<sup>-2</sup> harmonic potential was used. The initial 20–30 ns were used for system equilibration, and the subsequent 30–50 ns were applied for analysis.

For the CTD of the M protein, we chose the closest CTD conformations (dimer 1, see Fig. 1 for the umbrella sampling simulation). The CTD did not have a TMD, and we removed the TMD and N-terminal domain (NTD) and the GROMACS 'gmX solvate' to insert the CTD part into a water box. We performed 60 ns enhanced sampling for each window. The initial 20 ns was used to equilibrate the system. All simulation parameters were consistent with those mentioned above. The reaction coordinate of all systems was the center of mass of the protein (or CTD).

Finally, the potential mean forces (PMFs) were computed using the weighted histogram analysis method (WHAM), and the profile was generated using the 'gmX wham' protocol [48]. A bootstrap analysis (N = 50) was performed to evaluate the statistical error. The PMF balance value was normalized to the vicinity of the y-axis 0, and the energy results were all set to kcal/mol.

## 2.5. mMPBSA

The molecular mechanics Poisson-Boltzmann surface area (MMPBSA) method was applied to calculate the binding energy between M proteins in all dimers. For all systems, 200 snapshots of the equilibrated last 400 ns were used to calculate the binding free energies with an interval of 2 ns [49]. In all calculations, water, ions and lipids were removed, and the implicit calculation model

was used. The binding free energy of the ligand-receptor (M–M proteins) complex could be calculated as:

$$\Delta G_{bind, aq} = \Delta H - T\Delta S - [\Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand})] \quad (1)$$

$$\Delta G_{bind, aq} = \Delta E_{MM} + \Delta G_{bind, solv} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta E_{Covalent} + \Delta E_{Elect} + \Delta E_{VDW} \quad (3)$$

$$\Delta G_{bind, solv} = \Delta E_{polar} + \Delta G_{non-polar} \quad (4)$$

$\Delta G_{protein}$  and  $\Delta G_{ligand}$  are defined as two chains of the M protein dimer. The binding energy was decomposed per residue to calculate each residue's contribution to the M protein dimer association.

## 3. PCA and aFEL

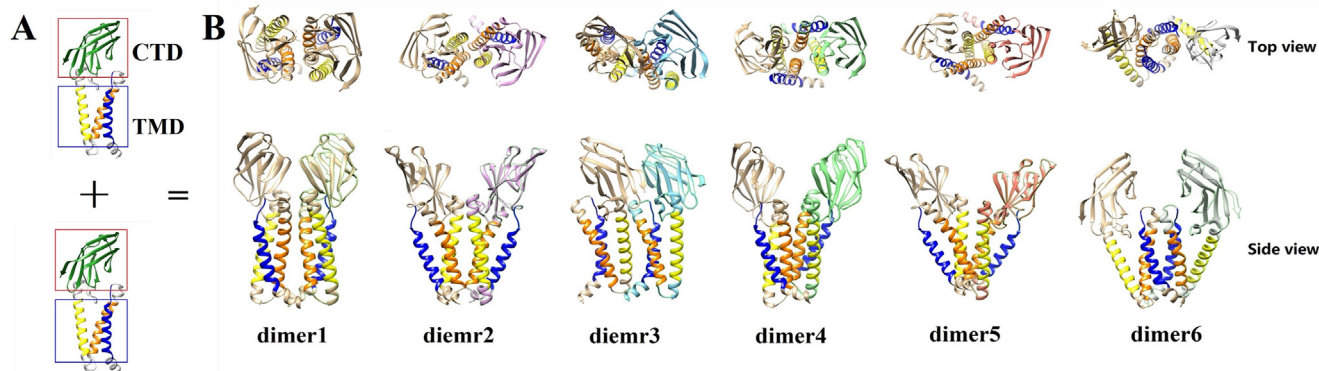
Principal component analysis (PCA) [50] was used as a part of the quasiharmonic analysis method to explore the structural and binding energy of six SARS-CoV-2 M protein dimers. The rotational and translational motions of the systems were eliminated by fitting the initial reference structure. The 5000 snapshots from the equilibrated 400 ns of all six M protein dimers were taken to generate the covariance matrix between the C $\alpha$  atoms of the M protein. The GROMACS tools gmX Sham, Covar and Ana eig were used to perform PCA and obtain the approximate free energy landscape (aFEL). Two feature vectors (PC1 and PC2) generated from PCA were used for the 2d aFEL.

The software Chimera [51] was used to visualize the protein structure. All of the simulations were performed using the GROMACS 2018 package [52] on the new-generation supercomputer in the National Supercomputer Center in Tianjin. The total simulation times were ~70  $\mu$ s.

## 4. Results

### 4.1. RMSD and RMSF

To build all possible M protein dimer models, we used the HDock online docking server of the Huang laboratory [33–53]. The HDock server based on a hybrid algorithm of template-based modeling and the ab initio method can accurately predict the structure of protein–protein or protein–DNA docking according to extensive experiments [54–55]. With the input of the M protein monomer predicted by the AlphaFold 2 AI algorithm and the opti-



**Fig. 1.** The M protein monomer (A) and six dimer model (B). A: The red rectangle is the CTD and the blue rectangle is the TMD of the M protein. B: The blue, orange and yellow helices represent TMH1, TMH2, and TMH3, respectively. A helix constitutes the transmembrane region of M and are, respectively. The top view and side view of six M protein dimers (dimer 1 to dimer 6) are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mization of M with the Feig laboratory (the highest prediction score on CASP14), the HDOCK server established a total of 100 possible M protein dimers. Through a filtering process (see the 'Methods' section for the filter criteria), six SARS-CoV-2 M protein dimer models were selected for subsequent simulation, as shown in Fig. 1A. To further confirm the accuracy of these models, we used different servers for M protein docking. The results of ClusPro, ZDOCK and HADDOCK included 10, 30 and 10 different dimer models respectively. Using the filtering process mentioned in the method section, 3, 3, and 2 models were obtained for each server, respectively. By comparing the dimer models, we found that they were consistent with the six HDOCK models. There were only slight differences in the CTD (the results are presented in Supplementary Figs. 1–3, and the results indicate that all six dimers are representative.

The M protein contains a TMD, and TMHMM was used to define the TMDs of the M protein as three helices, transmembrane helix (TMH)1 (W20–A40), TMH2 (K50–V70) and TMH3 (I80–F100). The CTD (L120–G202) was defined as well and is located within the envelope of the virus particle. Some evidence suggests that the CTD binds viral RNA, and its stability is important for the interaction of RNA with the M protein and M with the N protein [56]. Of the six M protein dimers, the associated formation of TMH1–TMH3 (Fig. 1B) is different. Dimer 1 and dimer 3–dimer 5 show a symmetrical dimerization structure, but dimer 2 shows a parallel dimerization structure. The three transmembrane helices of the M protein monomer interact with another monomer, and the binding positions of THM1–THM3 are different in each dimer. For the CTD, the two states – open (dimer 2, dimer 4–dimer 6) and closed (dimer 1 and dimer 3) – can be clearly observed in Fig. 1B.

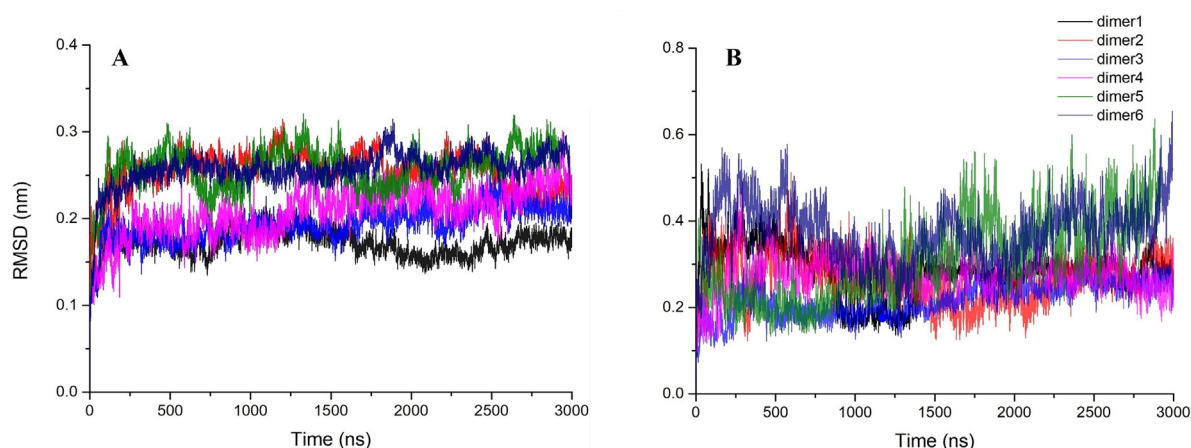
To evaluate the stability of the M protein dimer model in a complex membrane environment, six 3000 ns (3  $\mu$ s) atom molecular dynamics simulations for each system were performed. RMSD of which is shown in Fig. 2. We evaluated the stability of the TMD and CTD models by analyzing the root mean square deviation (RMSD), as shown in Fig. 2A (TMD) and 2B (CTD). For the TMD, the RMSD of all dimers converges in the 3000 ns simulation. The RMSD of dimer 1 is less than 0.2 nm when the curves plateau, and the fluctuation range is only 0.01 nm, showing high stability. For dimer 3 and dimer 4, the RMSD values remain stable at approximately 0.22 nm, and dimer 2, dimer 5 and dimer 6 have RMSD values of approximately 0.26 nm. Although the RMSD values indicate that the TMDs in the six systems are all stable, the RMSD of the other dimers fluctuates greatly, approximately 0.03–0.07 nm, com-

pared with that of dimer 1. In contrast, the CTDs in the different systems show completely different structural characteristics. The RMSD values and amplitudes of all the dimers are significantly increased, by approximately 0.1–0.3 nm, and some of the dimers (dimer 2, dimer 5 and dimer 6) do not converge during the simulation. These results suggest that the M protein dimer TMD has higher stability than the CTD, similar to most membrane protein system simulations.

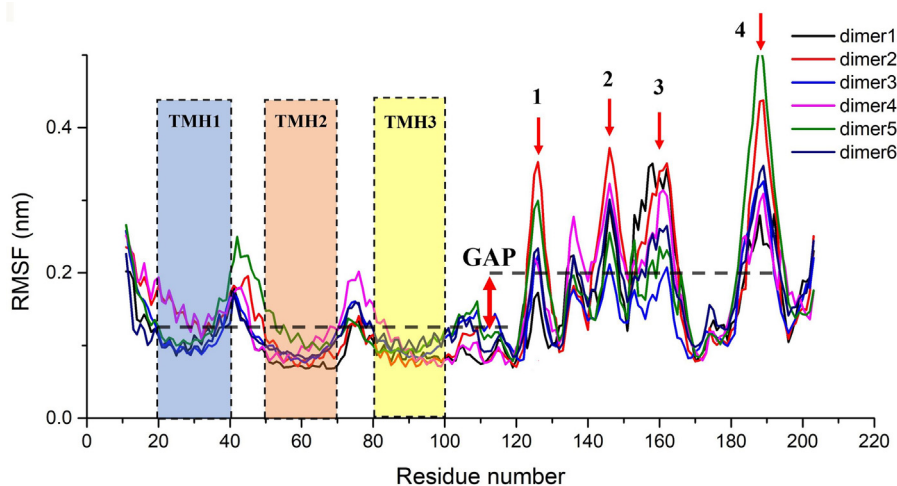
#### 4.2. PMF and MMPBSA

To characterize the dynamic behavior of each amino acid of the M protein, we analyzed the root mean square fluctuation (RMSF) of all the systems, as shown in Fig. 3. Consistent with the RMSD, the RMSF of the TMD has less fluctuation than that of the CTD, especially the three transmembrane helices, which have the smallest RMSF fluctuations (TMH1–TMH3). Of the six systems, dimer 1 has the smallest RMSF, while dimer 2 and dimer 5 have the largest values. The RMSF of THM1–3 of dimer 1 is approximately 0.1–0.15 nm, and that of the other five dimers reaches 0.1–0.2 nm. The CTD RMSF values of the six dimers are more obviously different, and the maximum region reaches  $\sim$  0.3 nm. A comparison of the mean values of TMD and CTD shows that the gap between the two domains reaches  $\sim$  0.08 nm. The CTD part in the profile (Fig. 3 shows four peak positions, corresponding to the amino acid positions of H125–G126 (peak-1), R146–G147 (peak-2), R158–K162 (peak-3) and V187–S191 (peak-4), and the RMSFs are all more than 0.3 nm. The four peak regions of the M protein dimer are shown in Supplementary Fig. 4. The results further confirmed the stability of the TMD as well as the high flexibility of the CTD.

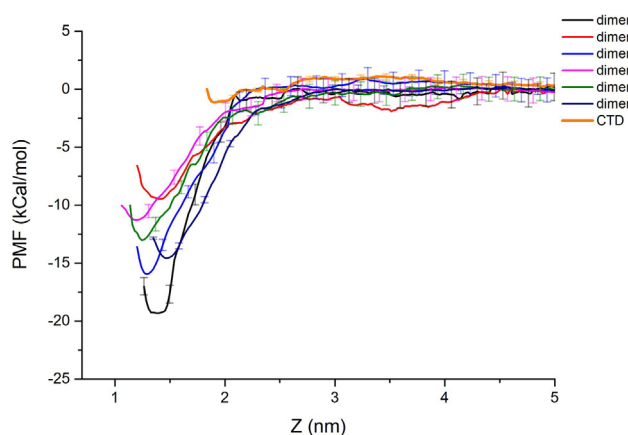
The association energy of the SARS-CoV-2 M protein dimer with different structural characteristics can be characterized by the potential of mean force (PMF), which describes the change in free energy between two M protein monomers along a particular reaction coordinate and is derived from the probability distribution along this coordinate. We analyzed the free energy distribution of M protein dimerization in a complex membrane system by using the umbrella sampling method and calculated the PMF for the lateral interaction of M protein monomers in six conformations. A steered MD simulation was performed in which a force was applied to pull the M protein monomer away from its binding site on the dimer. This generated a lateral 1D reaction coordinate perpendicular to the M protein surface, ranging from the bound to unbound state. The reaction coordinate Z was defined as the dis-



**Fig. 2.** Structural stability of the TMD (A) and CTD (B) Ca of six M protein dimers in 3000 ns MD simulations. Different colored lines (black, red, blue, pink, green, and purple) represent the RMSD values of the six M protein dimers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** RMSF profiles of six M protein dimers. The three  $\alpha$ -helices of TMD, TMH1–TMH3, are shown in the same color rectangles as those in Fig. 1. The red arrows indicate four spikes of the CTD. The black dotted line is the average RMSF between the TMD and CTD. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** PMF profiles of dimer 1–dimer 6 associations. The PMFs of dimer 1–dimer 6 are represented by black, red, blue, pink, green and purple lines, and the PMFs of the CTD are represented by orange bold lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tance between the center of mass of the M protein TMD binding site. The PMF difference of the six dimers determines the dimerization capacity. According to the PMF profile, dimer 1 has the largest binding energy, at  $\sim 20$  kcal/mol, when it is dimerized, while the PMFs of dimers 4, 3, 6, 5 and 2 gradually decrease, and the PMFs range from 18 to 10 kcal/mol. The maximum energy difference  $\Delta G$  is  $\sim 10$  kcal/mol. To evaluate whether the TMD or CTD of the M protein has different influences on the formation of the dimer, the PMF of dimer 1 CTD was also calculated. When the system has only CTD, the maximum PMF is only  $\sim 3$  kcal/mol. These observations indicate that the TMDs of the M protein are more affected by dimerization than CTDs. The PMF results are consistent with previous relevant SARS-CoV M protein dimer experimental data. [57]

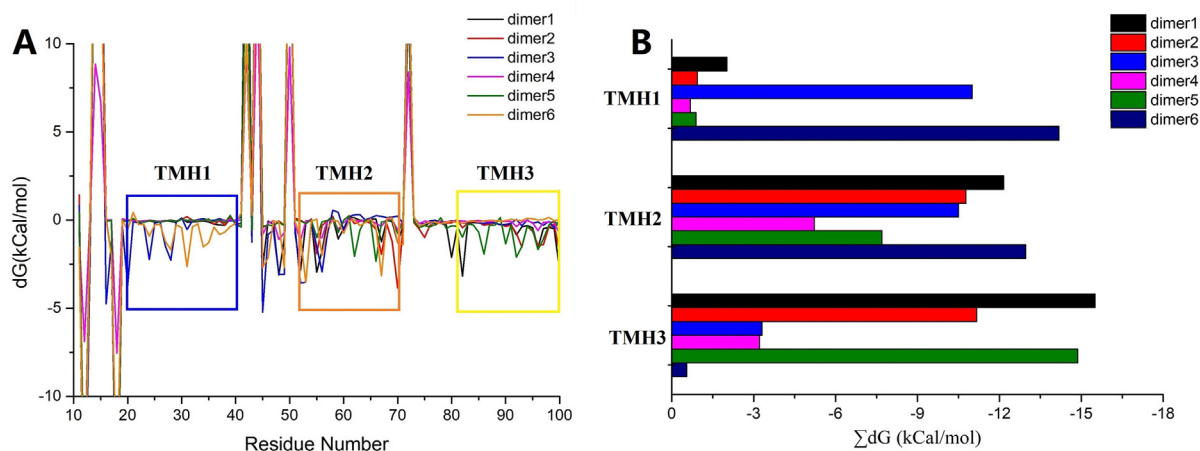
The binding energetics of six dimers of SARS-CoV-2 M protein were investigated by the MMPBSA method. The binding energy was partitioned into its individual components including the van der Waals, electrostatic, polar solvation, and solvent accessible surface area (SASA) energies. The PMF demonstrated that the TMD has the greatest effect on the association of M proteins. By analyzing the 3000 ns simulation trajectories (see methods section), we cal-

culated the binding energies of all the residues in the TMD (E11–F100). The binding energy results indicate that the amino acid positions with the largest difference are located in the three helices (TMH1–TMH3) of the THD. The binding free energies for M protein dimers were decomposed into a per-residue-based binding energy (the values for each residue are shown in Supplementary Fig. 5 to find the residue contribution. Interestingly, the binding affinities of the six M protein dimers are all different. We compared the contributions of free energies for each TMH to investigate the influence of TMH1–TMH3 on the association of the M protein (shown in Fig. 5B). For dimer 1, dimer 2, dimer 4 and dimer 5, the higher binding affinity is concentrated in TMH2 and TMH3. For dimer 6, the major contributors to M protein association are TMH1 and TMH2. For dimer 4, the binding energy is equally distributed between TMH1–TMH3. The MMPBSA results clarify that the TMD maintains the dimers through the TMH interaction in the simulation process, but the binding energy of decomposition to each dimer is different, suggesting that the TMHs have an important influence on the stability of M protein dimers with different conformations.

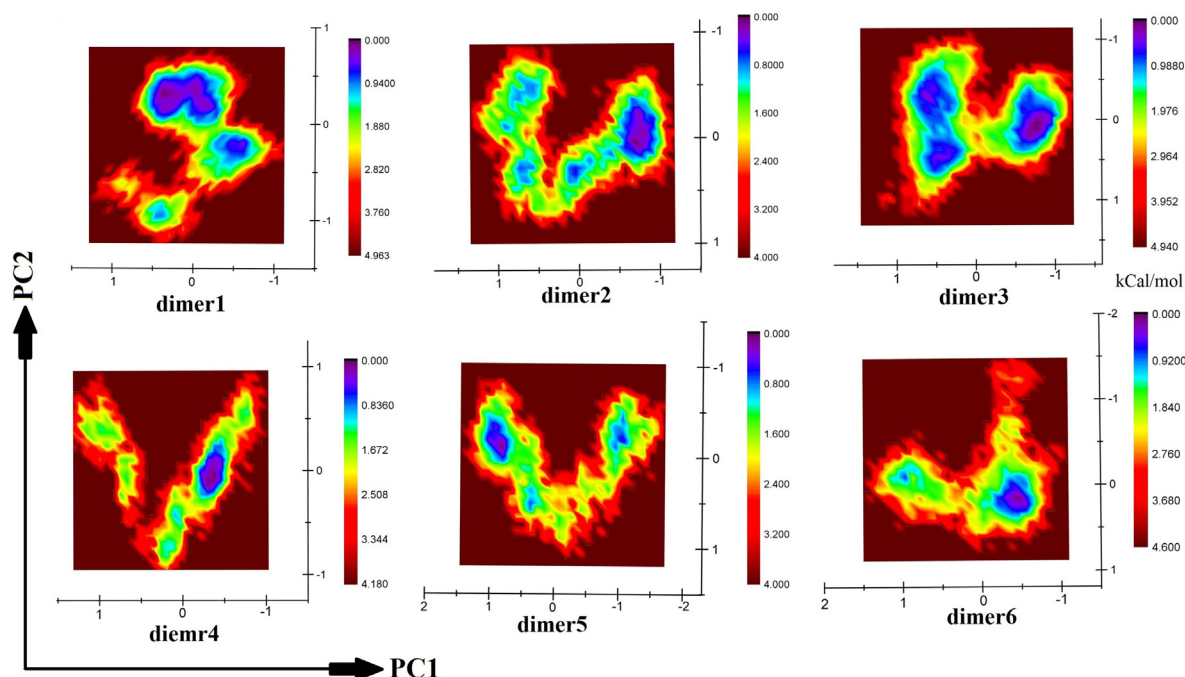
Neuman et al. indicated that the tightness of the M protein dimer could affect the curvature of the viral membrane [19]. Our results show that the TMD of the M protein dimer being in close contact would be better for maintaining the stability of the viral membrane. However, when the TMD is loosely bound, the dimer is more likely to disturb the membrane. The transition of the dimer between these conformations changes the shape of the virus. In summary, the TMD of the M protein is critical for maintaining the stability of the virus structure and morphology, as well as its virulence and infectious ability. In summary, TMDs are critical to viral morphology, virulence, and infectious ability.

#### 4.3. PCA and aFEL

To identify the dominant motions in the TMD of the M protein dimer, PCA was performed on the 400 ns MD simulation trajectory of each system after equilibrium to obtain the decomposition diagram defined by two feature vectors PC1 and PC2. Then, the feature vector is converted to an aFEL, which displays the variance in the TMD conformational motion, as shown in Fig. 6. The blue to red bars represent the free energy from the lowest to the highest. The results show that all the aFEL dimers separate into four free energy gradients, 0, 0.8, 1.6 and 3.5 kcal/mol. Obviously, dimer 1



**Fig. 5.** Binding energies (MMPBSAs) of six M protein dimers. A. Binding energy decomposition per residue (E11–F100) of the M protein dimer. B. Sum of the binding energies of the three TMHs (TMH1–TMH3) of the TMD. M protein dimer 1–dimer 6 are shown in different colors.



**Fig. 6.** Mapping of the principal components of the six dimers from the 400 ns simulations for the SARS-CoV-2 M protein. The color bar is relative to the free energy state.

and dimer 3’s aFELs show a lower metastable state, and the lowest free energies are  $-4.963$  and  $-4.940$  kcal/mol. Dimer 2, dimer 4–dimer 6 have smaller regions of blue and purple than dimer 1 and dimer 3, demonstrating that there are large motions in the TMD. Overall, compared with the PMF and MMPBSA, these plots show similar characteristics demonstrating the involvement of these TMDs in the dynamic stability of the M protein dimers.

### 5. Discussion

COVID-19 is an unprecedented threat to humans, and currently, the vaccines available are effective against nonmutant SARS-CoV-2 [58]. However, the vaccines are less effective against mutant variants, especially the delta mutation. Researchers studying the

pathogenesis and infection routes of COVID-19 have mainly focused on the S protein, but information on other structural proteins is still lacking. A more detailed understanding of the role of structural proteins in viruses is very important for revealing how the virus enters a cell, replicates and is released. As the most abundant structural protein in a coronavirus, the M protein is the main coordinator of virus assembly. The M protein is a transmembrane protein that also has a major function as a dimer in interactions with S, E and N proteins. The M protein includes two main domains, the TMD and CTD. The TMD binds stably to the virus membrane as well as the transmembrane region of the S and E proteins. There is evidence that the CTD of the M protein, which is located in the interior of virus particles, immobilizes viral RNA by N and M protein interactions [59 60]. Therefore, the SARS-

CoV-2 M protein is important for maintaining virus morphology, stabilizing the position of other structural proteins and the size of the coronavirus.

In this study, we thoroughly investigated the SARS-CoV-2 M protein dimer structure binding free energy and structural characteristics. By using the optimized AlphaFold-2-built M protein model, the HDOCK server modeled six candidate dimers, and the CLUSPRO, ZDOCK and HADDOCK servers were used to build a total of 50 models to verify the binding coverage of the dimers. The conformations of the M protein dimer can be divided into CTD open states (dimer 2 and dimers 4–6) and CTD closed states (dimer 1 and dimer 3). The difference in the TMD in the dimer is mainly due to the difference in the transmembrane interaction surfaces (Fig. 2 and Supplementary Fig. 6. After long timescale MD simulations, we find that all six systems are stable and do not dissociate in water or the membrane region, but the dimer conformation shows different steady-state dynamic behaviors. The RMSD values of the TMD and CTD suggest the critical influence of the initial structure on the overall structural stability, especially for the open state of the CTD of the M protein. The RMSD shows higher fluctuation, but the closed state does not appear. This is due to the binding affinity of the residues in CTD being very low. Notably, dimer 4 shows a “semi-open” state, and the CTD shows structural stability between dimer 1 and dimer 2. Compared with dimer 1, the TMDs of dimer 4 interaction surfaces are different. We analyzed the MD trajectory of dimer 4, which does not show a CTD closed state similar to that of dimer 1.

The RMSF values (Fig. 3 show that the three transmembrane helices, TMH1–TMH3, of the six dimers have high stability, with a fluctuation range of 0.1–0.2 nm, and the average RMSF is 0.12 nm. The CTD shows a large RMSF, especially for four spikes (red arrows in Fig. 3, and the value exceeds 0.4 nm. The average RMSF of the CTD is 0.18 nm, 0.06 nm higher than that of the TMD. Supplementary Fig. 4 shows the positions of the amino acids for the four spikes in the M protein CTD, which are all located at the hairpin of the B-sheet, indicating that the B-sheet affects the overall stability of the CTD. Combined with the RMSD and RMSF, although the overall stability of the M protein depends on both the CTD and TMH, the TMD plays a key role in the stability of the M protein dimer.

Regarding the TMD interactions of different M protein dimers, the PMF was calculated. The PMF is a monotonically decreasing function of the degree of separation; it describes the change in free energy between two M protein monomers along a particular reaction coordinate and is derived from the probability distribution along this coordinate. Starting from the critical separation distance of ~ 2.5 nm A, the M protein binding energy landscape forms a steep downslope funnel depending on the M protein center-of-mass distance. In particular, each M protein dimer has a different PMF profile. Dimer 1 needs more free energy for disassociation, approximately 20 kcal/mol, while dimer 2 needs less, approximately 10 kcal/mol, as shown in Fig. 4. The PMF result of dimer 4 is the closest to that of dimer 1, which corresponds to the structural characteristics between the two dimers. We also built a system to calculate the association energy of the CTD of dimer 1 (the center of mass of dimer 1's CTD is the closest). Under the same MD simulation conditions, the CTD disassociation free energy is only ~ 3 kcal/mol. Although the separate CTD simulation system cannot fully reflect the role of the M protein CTD in dimerization, some simulations suggest that the CTD contributes very little (less than 10%) to M protein binding [61 62 63]. The PMF results further demonstrate that the difference in the binding affinity of the six dimers is greater for the three THM binding surfaces than the CTD of the M protein.

Notably, the structural characteristics of dimer 3 are significantly different from those of other dimers. The M protein mono-

mers contact in parallel, and the dynamic behavior is very stable. The disassociation free energy of the dimer is 16 kcal/mol, indicating that it is also a high probability M protein dimer conformation. We suspect that in some cases, the M protein monomer can form a head–tail contact oligomer structure through a centered dimer 3-like model; if the virus particles need to be recombinant or changed in structure, the M protein oligomer can dissociate into dimers and interact with N, E, S, proteins or the membrane. Our hypothesis can also explain the results of some previous experiments [23 20].

Since the current umbrella sampling method cannot achieve complete convergence for membrane proteins, it has some influence on the accuracy of the PMF results. The MMPBSA method was used to quantitatively calculate the binding energy of each residue of the M protein dimer during the stability period of MD simulation, especially the contact residues in the TMH1–TMH3 surfaces. Fig. 5B shows that the TMHs of the dimer do not contribute the same binding energy. The energy contributions of dimer 1, dimer 2, dimer 4 and dimer 5 are mainly from TMH2 and TMH3. Those of Dimer 6 are from TMH1 and TMH2, and those of dimer 3 are from all three TMHs (each residue binding energy is shown in Supplementary Fig. 5 and Table 2. The differences in the binding free energies of the six dimer TMDs suggest the stability of the M protein dimers in the complex membrane environment, and the MMPBSA results are in agreement with the PMF results, indicating that the PMF value is convergent.

According to the MMPBSA results for each residue, the large numbers of leucine, isoleucine, phenylalanine and tryptophan residues in areas including TMH1–TMH3 suggest that aromatic amino acids and Leu-Ile zippers may play an important role in maintaining dimer stability, which further confirms previous studies based on structural characteristics [64 65].

The aFEL from the PCA and MMPBSA results shows that TMH2 interacts in all six dimers, demonstrating that it plays a key role in the steady state of the dimers, followed by TMH3 and TMH1. This is mainly determined by the topological structure of the M protein. TMH2 is located at the pocket of the M protein and connected to the other two TMHs, which increases the helix stability during the simulation. The MMPBSA results show that of all the TMHs, three residues—Trp55 Leu56 and Phe96—have the lowest binding energy in four dimers and Ile52 and Leu67 have the lowest binding energy in three dimers, suggesting that these residues play an important role in maintaining the homeostasis of the M protein dimer, even though their conformations are different. Dimer 5 has special structural characteristics compared with the other dimers. The six dimers have different lowest binding energy positions of

**Table 2**  
High-affinity residues in the six dimers.

	Amino acid names	Binding energy (kCal/mol)	TMH locations
Dimer1	Ile52, Trp55, Leu56,	–2.862, –2.954, –1.745,	TMH2, TMH3
	Leu67, Val70, Ile82,	–1.291, –1.277, –3.174,	
	Phe96, Phe100	–1.567, –2.826	
Dimer2	Trp55, Leu67, Val70,	–1.906, –1.933, –3.851,	TMH2, TMH3
	Leu93, Phe96, Ile97,	–1.521, –1.416, –1.392,	
	Phe100	–1.845	
Dimer3	Ile24, Leu27, Phe28,	–2.242, –1.498, –2.27,	TMH1, TMH2
	Ile52, Phe53, Trp55,	–3.562, –3.525, –1.492,	
	Leu56	–2.93	
Dimer4	Ile52, Trp55, Leu56,	–1.174, –0.944, –1.006,	TMH2, TMH3
	Phe96	–0.599	
Dimer5	Leu62, Val66, Cys86,	–2.046, –2.337, –1.931,	TMH2, TMH3
	Trp92, Leu93, Phe96,	–2.104, –1.493, –1.67,	
	Ile97	–1.08	
Dimer6	Leu27, Phe28, Trp31,	–1.223, –1.686, –2.639,	TMH1, TMH2
	Leu34, Leu35, Phe53,	–1.516, –1.281, –3.54,	
	Leu56, Leu67	–1.998, –3.15	

TMH2. We speculate that the CTD of the initial structure is widely open and that the axial slope of the TMD's interaction surface changes, leading to the deviation of the two M protein binding sites. Interestingly, from the RMSD, PMF, MMPBSA and aFEL results, the change in the binding site does not affect the M protein dimer 5 (shown in Table 2). It is suggested that the state of the CTD has little effect on the overall structural stability of the M protein dimer. This also confirms the hypothesis of Haan et al. that when all three TMDs are heterosubstituted, the interactions of M proteins are severely reduced [20]. They also suggested that the interaction of the M protein dimer TMD with the S protein fixed the position of S with the viral envelope, independent of the CTD [24].

Lu et al. proposed an important M + N + RNA three-compartment structure model that suggested that M and RNA could reject each other without eliminating their interactions [66]. It is well known that the N protein binds to the M protein through the CTD [67]. Our calculations indicated that the open (over-open) state of the flexible CTD does not promote dimer stability. We speculate that this might be why the viral RNA can be fixed on the novel coronavirus envelope through the N and M proteins inside the virus despite its long length and flexibility.

The morphology and mechanism of SARS-CoV-2 M protein dimers may be varied. Although the topological stability and binding free energies of the six dimers differ, they do not directly lead to dissociation of the dimers, suggesting that the dimers may coexist in various forms (see Fig. 7). We propose the following reasons: (1) By changing the binding site, the different dimer conformations can change the curvature of the viral envelope and control its shape to adapt to different environments. (2) The TMD of the dimer can bind with the S and E proteins, which requires changing the M protein dimer conformation to expose its binding sites to interact with other structural proteins. (3) When the CTD of the M protein interacts with the N protein and viral RNA, the stability of N and RNA in the virus is not compromised because of the specific morphology of the M protein dimer. (4) When the M dimer is not necessary, M exists as an oligomer with lower binding energy that can easily and quickly dissociate when the environment changes to form the dimer. So far, the virus is still mutating. The amino acid sequences of the M protein of the two common mutations (DELTA andOMICRON) showed that only 1–2 amino acids were mutated compared with the original strain, indicating that the M protein is highly conserved (the amino acid sequence multiple alignments

results are shown in Supplementary Fig. 7). This conservation may be closely related to the function of the M protein in the virus, such as maintaining the viral envelope morphology, other structural proteins and genetic material stability. These factors provide SARS-CoV-2 with high adaptability to the human environment, which will provide valuable clues for future research on SARS-CoV-2 M protein and its mutations.

## 6. Conclusion

Although the crystal structure of the novel coronavirus 2019 M protein has not yet been determined, the M protein monomer could still be built based on AlphaFold 2, and the dimers were obtained by the HDock server. The dimers were characterized, and their structural properties were estimated by all-atom MD simulations in the complex membrane composition. Long time-scale MD simulation and free energy results reveal the most stable state of the M protein dimer. The structural differences in M protein dimers are mainly from their binding surface, and the free energy results suggest that the TMD plays an important role in the stability of dimers. The important amino acids for dimers were concentrated in the Leu-Ile zipper and benzene ring amino acids in THM1 to THM3. In addition, this study suggests that the flexible part of the CTD of the M protein is located within the virus envelope but that it has little effect on the overall structure of the dimer. We suggest that the TMD of the M protein dimer may maintain the stability of other structural proteins (S, N, and E) on the viral envelope through similar mechanisms, but the interaction mechanisms need more study. In conclusion, our study revealed the mechanisms of M protein dimerization at the molecular level. In the future, relevant strategies should be used to destroy the assembly of the M protein to reduce the replication ability of the novel coronavirus and to even kill the virus. This study provides novel insights into important structural features of interface residues for the advancement of effective therapeutic strategies to target the role of M proteins in the viral life cycle.

## CRedit authorship contribution statement

**Yipeng Cao:** Conceptualization, Methodology, Data curation, Writing & review & editing. **Rui Yang:** Conceptualization, Metho-

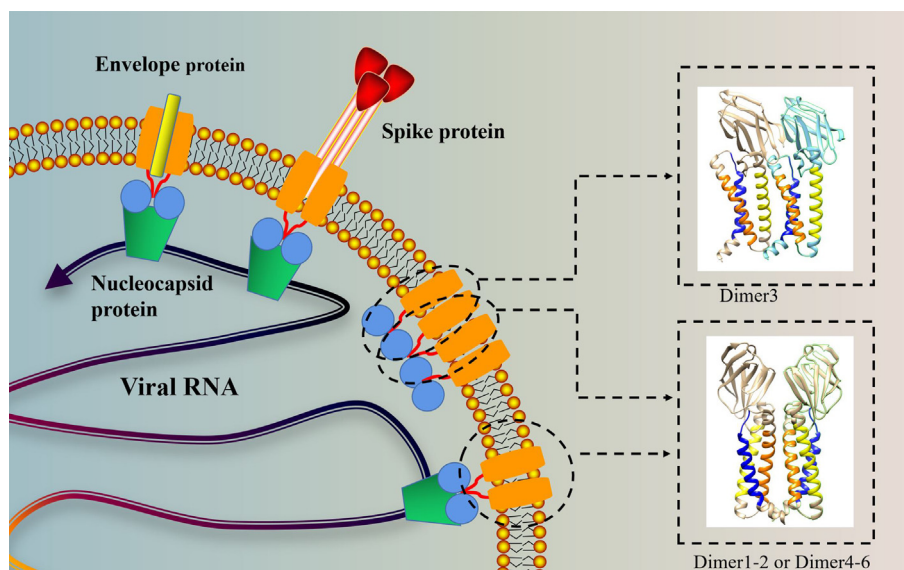


Fig. 7. Model of SARS-CoV-2 M protein dimer functions on the viral membrane.



dology, Data curation. **Wei Wang:** . **Shengpeng Jiang:** Data curation. **Chengwen Yang:** Data curation. **Ningbo Liu:** Data curation. **Hongji Dai:** Data curation. **Imshik Lee:** Data curation. **Xiangfei Meng:** Methodology, Data curation, Writing & review & editing, Supervision. **Zhiyong Yuan:** Methodology, Data curation, Writing & review & editing, Supervision.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This study was supported by the National Natural Science Foundation of China (NSFC-31900894) and the China Postdoctoral Science Foundation (2021M700975). We thank the National Supercomputer Center in Tianjin (NSCC-TJ) for the long-timescale MD simulations performed on the New Generation Exascale Tianhe Supercomputer.

### References

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Yi, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497–506.
- Pedersen SF, Ho Y-C. SARS-CoV-2: a storm is raging. *J Clin Invest* 2020;130:2202–5.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19(7):409–24.
- Shen X, Tang H, Pajon R, Smith G, Glenn GM, Shi W, et al. Neutralization of SARS-CoV-2 variants B. 1.429 and B. 1.351. *N Engl J Med* 2021;384(24):2352–4.
- Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 2021;596(7871):276–80.
- Yao H, Song Y, Chen Y, Wu N, Xu J, Sun C, et al. Molecular architecture of the SARS-CoV-2 virus. *Cell* 2020;183(3):730–738.e13.
- Brouwer PJM, Brinkkemper M, Maisonnasse P, Dereuddre-Bosquet N, Grobben M, Claireaux M, et al. Two-component spike nanoparticle vaccine protects macaques from SARS-CoV-2 infection. *Cell* 2021;184(5):1188–1200.e19.
- He L, Lin X, Wang Y, Abraham C, Sou C, Ngo T, et al. Single-component, self-assembling, protein nanoparticles presenting the receptor binding domain and stabilized spike as SARS-CoV-2 vaccine candidates. *Sci Adv* 2021;7:eabf1591.
- Bartlam M, Yang H, Rao Z. Structural insights into SARS coronavirus proteins. *Curr Opin Struct Biol* 2005;15(6):664–72.
- Shang J, Ye G, Shi Ke, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020;581(7807):221–4.
- Mariano G, Farthing RJ, Lale-Farjat SLM, Bergeron JRC. Structural characterization of SARS-CoV-2: Where we are, and where we need to be. *Front Mol Biosci* 2020;7:344.
- Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol* 2009;7(3):226–36.
- Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 2020;181(4):894–904.e9.
- Clarson CR, Asfaha JB, Ghent CM, Howard CJ, Hartooni N, Safari M, et al. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol Cell* 2020;80:1092–103.
- Cao Y, Yang R, Wang W, Lee I, Zhang R, Zhang W, et al. Computational Study of the Ion and Water Permeation and Transport Mechanisms of the SARS-CoV-2 Pentameric E Protein Channel. *Front Mol Biosci* 2020;7:565797. <https://doi.org/10.3389/fmolb.2020.565797>.
- Hu Y, Wen J, Tang L, Zhang H, Zhang X, Li Y, et al. The M protein of SARS-CoV: basic structural and immunological properties. *Genom Proteom Bioinform* 2003;1(2):118–30.
- He R, Leeson A, Ballantine M, Andonov A, Baker L, Dobie F, et al. Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. *Virus Res* 2004;105(2):121–5.
- Dey D, Borkotoky S, Banerjee M. In silico identification of Tretinoin as a SARS-CoV-2 envelope (E) protein ion channel inhibitor. *Comput Biol Med* 2020;127:104063. <https://doi.org/10.1016/j.compbiomed.2020.104063>.
- Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* 2011;174(1):11–22.
- de Haan CAM, Vennema H, Rottier PJM. Assembly of the coronavirus envelope: homotypic interactions between the M proteins. *J Virol* 2000;74(11):4967–78.
- Khan A, Tahir Khan M, Saleem S, Junaid M, Ali A, Shujait Ali S, et al. Structural insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Comput Struct Biotechnol J* 2020;18:2174–84.
- Collins LT, Elkholy T, Mubin S, Williams R, Ezike K, Singhal A. Understanding SARS-CoV-2 budding through molecular dynamics simulations of M and E protein complexes. *BioRxiv* 2021.
- Perrier A, Bonnin A, Desmarests L, Danneels A, Goffard A, Rouillé Y, et al. The C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal. *J Biol Chem* 2019;294(39):14406–21.
- de Haan CAM, Smeets M, Vernooij F, Vennema H, Rottier PJM. Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. *J Virol* 1999;73(9):7441–52.
- Keller MD, Harris KM, Jensen-Wachspress MA, Kankate V V, Lang H, Lazarski CA, et al. SARS-CoV-2-specific T cells are rapidly expanded for therapeutic use and target conserved regions of the membrane protein. *Blood* 2020;136:2905–17.
- Chan C-M, Ma C-W, Chan W-Y, Chan HYE. The SARS-Coronavirus Membrane protein induces apoptosis through modulating the Akt survival pathway. *Arch Biochem Biophys* 2007;459(2):197–207.
- Zheng Yi, Zhuang M-W, Han L, Zhang J, Nan M-L, Zhan P, 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* 2020;5(1). <https://doi.org/10.1038/s41392-020-00438-7>.
- Yu A, Pak AJ, He P, Monje-Galvan V, Casalino L, Gaieb Z, et al. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys J* 2021;120(6):1097–104.
- Ouzounis CA. A recent origin of Orf3a from M protein across the coronavirus lineage arising by sharp divergence. *Comput Struct Biotechnol J* 2020;18:4093–102.
- Callaway E. “It will change everything”: DeepMind’s AI makes gigantic leap in solving protein structures. *Nature* 2020;588(7837):203–4.
- Heo L, Janson G, Feig M. Physics-Based Protein Structure Refinement in the Era of Artificial Intelligence. *Proteins Struct Funct Bioinforma* 2021;89(12):1870–87.
- Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins Struct Funct Bioinforma* 2021;89:1687–99.
- Yan Y, Tao H, He J, Huang S-Y. The HDock server for integrated protein–protein docking. *Nat Protoc* 2020;15(5):1829–52.
- Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* 2017;45:W365–73.
- Yan Y, He J, Feng Y, Lin P, Tao H, Huang S. Challenges and opportunities of automated protein–protein docking: HDock server vs human predictions in CAPRI Rounds 38–46. *Proteins Struct Funct Bioinforma* 2020;88:1055–69.
- Aderinwale T, Christoffer CW, Sarkar D, Alnabati E, Kihara D. Computational structure modeling for diverse categories of macromolecular interactions. *Curr Opin Struct Biol* 2020;64:1–8.
- Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 2018;27(1):293–315.
- de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 2010;5(5):883–97.
- Pierce BG, Wiehe K, Hwang H, Kim B-H, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* 2014;30(12):1771–3.
- Kozakov D, Hall DR, Xia B, Porter KA, Padhorna D, Yueh C, et al. The ClusPro web server for protein–protein docking. *Nat Protoc* 2017;12(2):255–78.
- Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat Struct Mol Biol* 2020;27(12):1202–8.
- Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001;305:567–80.
- Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 2008;29(11):1859–65.
- Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, 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* 2010;31:671–90.
- Melchionna S, Ciccotti G, Lee Holian B. Hoover NPT dynamics for systems varying in shape and size. *Mol Phys* 1993;78(3):533–44.
- Lu H, Schulten K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins Struct Funct Bioinforma* 1999;35(4):453–63.
- Domański J, Hedger G, Best RB, Stansfeld PJ, Sansom MSP. Convergence and sampling in determining free energy landscapes for membrane protein association. *J Phys Chem B* 2017;121(15):3364–75.
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 1992;13(8):1011–21.

- [49] Li J, Zhou N, Cai P, Bao J. In silico screening identifies a novel potential PARP1 inhibitor targeting synthetic lethality in cancer treatment. *Int J Mol Sci* 2016;17(2):258. <https://doi.org/10.3390/ijms17020258>.
- [50] Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. *J Phys Chem* 1996;100(7):2567–72.
- [51] Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 2021;30:70–82.
- [52] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1-2:19–25.
- [53] Gordon DE, Hiatt J, Bouhaddou M, Rezelj V V, Ulferts S, Braberg H, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* (80-) 2020;370.
- [54] Pitsillou E, Liang J, Ververis K, Lim KW, Hung A, Karagiannis TC. Identification of small molecule inhibitors of the deubiquitinating activity of the SARS-CoV-2 papain-like protease: in silico molecular docking studies and in vitro enzymatic activity assay. *Front Chem* 2020;8:1171.
- [55] Rostøl JT, Xie W, Kuryavyi V, Maguin P, Kao K, Froom R, et al. The Card1 nuclease provides defence during type III CRISPR immunity. *Nature* 2021;590:624–9.
- [56] Chen C-Y, Chang C-k, Chang Y-W, Sue S-C, Bai H-I, Riang L, et al. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol* 2007;368(4):1075–86.
- [57] Cao Y, Wu X, Yang R, Wang X, Sun H, Lee I. Self-assembling study of sarcolipin and its mutants in multiple molecular dynamic simulations. *Proteins-Structure Funct Bioinforma* 2017;85(6):1065–77.
- [58] Farinholt T, Doddapaneni H, Qin X, Menon V, Meng Q, Metcalf G, et al. Transmission event of SARS-CoV-2 Delta variant reveals multiple vaccine breakthrough infections. *MedRxiv* 2021.
- [59] Mortola E, Roy P. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett* 2004;576:174–8.
- [60] Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* 2011;85(9):4122–34.
- [61] He S, Moutaoufik MT, Islam S, Persad A, Wu A, Aly KA, et al. HERG channel and cancer: a mechanistic review of carcinogenic processes and therapeutic potential. *Biochim Biophys Acta (BBA)-Reviews Cancer* 2020;1873:188355.
- [62] Naughton FB, Kalli AC, Sansom MSP. Association of peripheral membrane proteins with membranes: Free energy of binding of GRP1 PH domain with phosphatidylinositol phosphate-containing model bilayers. *J Phys Chem Lett* 2016;7(7):1219–24.
- [63] Lapelosa M. Conformational dynamics and free energy of BHRF1 binding to Bim BH3. *Biophys Chem* 2018;232:22–8.
- [64] Tseng Y-T, Chang C-H, Wang S-M, Huang K-J, Wang C-T, Pöhlmann S. Identifying SARS-CoV membrane protein amino acid residues linked to virus-like particle assembly. *PLoS ONE* 2013;8(5):e64013.
- [65] Wong NA, Saier MH. The SARS-coronavirus infection cycle: a survey of viral membrane proteins, their functional interactions and pathogenesis. *Int J Mol Sci* 2021;22(3):1308. <https://doi.org/10.3390/ijms22031308>.
- [66] Lu S, Ye Q, Singh D, Cao Y, Diedrich JK, Yates JR, et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat Commun* 2021;12(1). <https://doi.org/10.1038/s41467-020-20768-y>.
- [67] Rahman MS, Islam MR, Alam ASMRU, Islam I, Hoque MN, Akter S, et al. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J Med Virol* 2021;93(4):2177–95.