COVID-19: Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit Human Heme Metabolism

Wenzhong Liu [1,2,*], Hualan Li [2]

[1] School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong, 643002, China;

[2] School of Life Science and Food Engineering, Yibin University, Yibin, 644000, China;

[*] Correspondence: liuwz@suse.edu.cn;

Abstract

The novel coronavirus pneumonia (COVID-19) is an infectious acute respiratory infection caused by the novel coronavirus. The virus is a positive-strand RNA virus with high homology to bat coronavirus. In this study, conserved domain analysis, homology modeling, and molecular docking were used to compare the biological roles of certain proteins of the novel coronavirus. The results showed the ORF8 and surface glycoprotein could bind to the porphyrin, respectively. At the same time, orf1ab, ORF10, and ORF3a proteins could coordinate attack the heme on the 1-beta chain of hemoglobin to dissociate the iron to form the porphyrin. The attack will cause less and less hemoglobin that can carry oxygen and carbon dioxide. The lung cells have extremely intense poisoning and inflammatory due to the inability to exchange carbon dioxide and oxygen frequently, which eventually results in ground-glass-like lung images. The mechanism also interfered with the normal heme anabolic pathway of the human body, is expected to result in human disease. According to the validation analysis of these finds, chloroquine could prevent orf1ab, ORF3a, and ORF10 to attack the heme to form the porphyrin, and inhibit the binding of ORF8 and surface glycoproteins to porphyrins to a certain extent, effectively relieve the symptoms of respiratory distress. Since the ability of chloroquine to inhibit structural proteins is not particularly obvious, the therapeutic effect on different people may be different. Favipiravir could inhibit the envelope protein and ORF7a protein bind to porphyrin, prevent the virus from entering host cells, and catching free porphyrins. This paper is only for academic discussion, the correctness needs to be confirmed by other laboratories. Due to the side effects and allergic reactions of drugs such as chloroquine, please consult a qualified doctor for treatment details, and do not take the medicine yourself.

Keywords: Novel Coronavirus; Respiratory distress; Ground-glass-Like Lung; E2 glycoprotein; ORF8; orf1ab; chloroquine; Blood; Diabetic; Fluorescence resonance energy transfer; Ancient virus; Cytokine Storm

1 Introduction

The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease. Patients with the coronavirus pneumonia have a fever, and the temperature above 38 degrees with symptoms such as dry cough, fatigue, dyspnea, difficulty breathing, and frost-glass-like symptoms in the lungs[1-3]. A large amount of mucus can be discovered in the dissected tissue without obvious virus inclusions. This pneumonia was first discovered in December 2019 in the South China Seafood Market Hubei Province, China[4]. The disease is highly transmitted[5,6]. Now the number of infected people has

reached tens of thousands around the world, and the infected people are not restricted by race and borders.

Researchers performed virus isolation tests and nucleic acid sequencing to confirm the disease was caused by a novel coronavirus[7,8]. It is noted that the nucleic acid of the novel coronavirus is a positive-stranded RNA[8]. Its structural proteins include: Spike Protein (S), envelope protein (E), membrane protein (M), and nucleocapsid phosphoprotein. Transcribed non-structural proteins include: orf1ab, ORF3a, ORF6, ORF7a, ORF10 and ORF8. The novel coronavirus is highly homologous to the coronavirus in bats[9,10], and has significant homology with SARS virus[11,12]. Researchers have studied the function of novel coronavirus structural proteins and some non-structural proteins[13,14]. But, the novel coronavirus has potential genomic characteristics, some of which are mainly the cause of human outbreaks[15,16]. For example, CoV EIC (Coronavirus envelope protein ion channel) been implicated in modulating virion release and CoV – host interaction[17]. Spike proteins, ORF8 and ORF3a proteins are significantly different from other known SARS-like coronaviruses, and they may cause more serious pathogenicity and transmission differences than SARS-CoV[18]. Earlier studies find that the novel coronavirus enters epithelial cells through the spike protein interacting with the human ACE2 receptor protein on the surface, causing human infection. However, structural analysis of the spike protein (S) protein of the novel coronavirus reveals that the S protein only weakly binds to the ACE2 receptor compared to SARS coronavirus[19]. Due to the limitations of existing experimental methods, the specific functions of virual proteins such as ORF8 and surface glycoprotein are still unclear. The pathogenicity mechanism of the novel coronavirus remains mysterious[20].

Literature[21] disclosed biochemical examination indexes of 99 patients with novel coronavirus pneumonia, and the report also reflected the abnormal phenomenon of hemoglobin-related biochemical indexes of patients. This report demonstrates that the hemoglobin and neutrophil counts of most patients have decreased, and the index values of serum ferritin, erythrocyte sedimentation rate, C-reactive protein, albumin, and lactate dehydrogenase of many patients increase significantly. This trace implies that the patient's hemoglobin is decreasing, and the heme is increasing, and the body will accumulate too many harmful iron ions, which will cause inflammation in the body and increase C-reactive protein and albumin. Cells react to stress due to inflammation, producing large amounts of serum ferritin to bind free iron ions to reduce damage. Hemoglobin consists of four subunits, 2-α and 2-β, and each subunit has an iron-bound heme[22,23]. Heme is an important component of hemoglobin. It is a porphyrin containing iron. The structure without iron is called porphyrin. When iron is divalent, hemoglobin can release carbon dioxide and capture oxygen atoms in alveolar cells, and iron is oxidized to trivalent. When hemoglobin is made available to other cells in the body through the blood, it can release oxygen atoms and capture carbon dioxide, and iron is reduced to divalent.

There are no particularly effective drugs and vaccines to control the disease against novel coronaviruses[24]. However, there are several old drugs has found in the latest clinical treatments, which can inhibit some functions of the virus, for example, cloroquine phosphate has a definite effect on the novel coronavirus pneumonia[25]. chloroquine phosphate is an antimalarial drug that has been utilized clinically for more than 70 years. Experiments show that erythrocytes infected by malaria can cause a large amount of chloroquine to accumulate in it. The drug leads to the loss of hemoglobin enzyme, and parasite death due to insufficient amino acids in the growth and development of the parasite. The therapeutic effect of chloroquine phosphate on novel coronavirus pneumonia suggests that novel coronavirus pneumonia might be closely related to abnormal hemoglobin metabolism in humans. Meanwhile, one detail we can notice is that chloroquine is also a commonly used drug for treating

porphyria[26,27].

Therefore, it is believed that combining viral proteins and porphyrins will cause a series of human pathological reactions, such as a decrease in hemoglobin. Because of the severe epidemic, and the existing conditions with limited experimental testing methods for the proteins' functions, it is of great scientific significance to analyze the proteins' function of the novel coronavirus by bioinformatics methods.

In this study, conserved domains prediction, homology modeling and molecular docking techniques were used to analyze the functions of virus-related proteins. This study found that ORF8 and surface glycoprotein had a function to combine with porphyrin to form a complex, while orf1ab, ORF10, ORF3a coordinately attack the heme on the 1-beta chain of hemoglobin to dissociate the iron to form the porphyrin. This mechanism of the virus inhibited the normal metabolic pathway of heme, and made people show symptoms of the disease. Built on the above research results, we also verified the role of chloroquine phosphate and Favipiravir by molecular docking technology to assist clinical treatment.

2 Materials and Methods

2.1 Data set

The protein sequences were downloaded from NCBI: All proteins of Wuhan novel coronavirus; Heme-binding protein; Heme oxidase; Protein sequences were utilized to analyze conserved domain.
All proteins of Wuhan novel coronavirus were also used to construct three-dimensional structures by homology modeling.

At the same time, the PDB files were downloaded from the PDB database: Crystal structure of MERS-CoV nsp10_nsp16 complex--5yn5; HEM; Human Oxy-Hemoglobin 6bb5; DEOXY HUMAN HEMOGLOBIN 1a3n; 0TX; RP. MERS-CoV nsp10_nsp16 complex--5yn5 was used to homology modeling. HEM, 0TX and 1RP were used to molecular docking. Two Oxy-Hemoglobin were used to protein docking

2.2 Flow view of bioinformatics analysis

A series of bioinformatics analysis was performed based on published biological protein sequences in this study. The steps are illustrated in Figure 1: 1. Conserved Domains of Viral Proteins are analyzed by MEME[28-30] Online Server. Conserved domains were used to predict function differences of viral proteins and human proteins. 2.The three-dimensional structure of viral proteins was constructed by homology modeling of Swiss-model[31,32]. When the sequence length exceeded 5000nt, the homology modeling tool of Discovery-Studio 2016 was adopted. 3. Using molecular docking technology (LibDock tool) of Discovery-Studio 2016[33], the receptor-ligand docking of viral proteins with human heme (or porphyrins) was simulated. Depending on the results of bioinformatics analysis, a life cycle modelof the virus was constructed, and the related molecular of the disease was proposed.
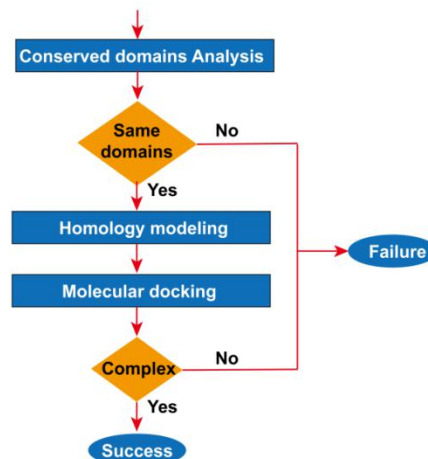
Figure 1. Flow view of Bioinformatics Analysis

The workflow is based on evolutionary principles. Although the biological sequence characteristic of advanced life forms and viral is different, molecules with similar structures can always achieve similar biological roles. The homology modeling method uses the principle the similar primary structure of protein sequences has a similar spatial structure. Molecular docking technology is built on homology modeling or really three-dimensional molecular.

2.3 Analysis of conserved domain

MEME Suite is an online website that integrates many tools of predicting and annotation motif. The maximum expectation (EM) algorithm is the basis for MEM's identification of the motif. The motif is a conserved domain of a small sequence in a protein. Motif-based models could assess the reliability of phylogenetic analysis. After opening the online tool MEME, the protein sequences of interest are merged into a text file, and the file format remains fasta. Then select the number of motifs you want to find, and click the "Go" button. At the end of the analysis, the conserved domains are displayed after clicking the link.

2.4 Homology modeling

SWISS-MODEL is a fully automatic homology modeling server for protein structure, which can be accessed through a web server. The first step is to enter the Swiss-model, enter the sequence, and click "Search Template" to perform a simple template search. After the search is completed, you can choose a template for modeling. A template search will be performed clicking "Build Model" and a template model is chosen automatically. As can be seen, several templates were searched, and then numerous models were built. Just a model is chosen here. The model in PDB format is downloaded and visualized in VMD. SWISS-MODEL only models protein models with sequence lengths less than 5000nt. You can use Discovery-Studio's homology modeling tool to model where the protein sequence exceeds 5000nt.

Before using Discovery-Studio to model homology of an unknown protein (such as orf1ab), the pdb structure file of the template protein, such as MERS-CoVnsp10_nsp16 complex 5yn5, should be downloaded from the PDB database. Next, the sequence alignment tool of Discovery-Studio is utilized to align homologous sequences between 5yn5 and orf1ab. Then the spatial structure file of orf1ab was constructed based on the template protein 5yn5.

2.5 Molecular docking technology

Molecular docking is the process of finding the best matching pattern between two or more molecules through geometric matching and energy matching. The steps for using LibDock molecular docking with Discovery-Studio are as follows:

1. Preparation of a ligand model. Open a ligand file such as HEM, and click "Prepare Ligands" in the "Dock Ligands" submenu of the "Receptor-Ligand Interactions" menu to generate a heme ligand model for docking. First delete FE (iron atom) in HEM, and then click the "Prepare Ligands" button, then the porphyrin ligand model will be generated. With 0 XT is opened, click "Prepare Ligands" again to get the chloroquine ligand model.

2. Prepare a protein receptor model. Open the protein's pdb file (generated by homology modeling), and click "Prepare protein" in the "Dock Ligands" submenu of the "Receptor-Ligand Interactions" menu to generate a protein receptor model for docking.

3. Set docking parameters to achieve docking. Select the generated protein receptor model. From the "Define and Edit Binding Site" submenus in the "Receptor-Ligand Interactions" menu, click "From receptor Cavities". A red sphere appears on the protein receptor model diagram. After you right-click the red ball, you can modify the radius of the red ball. Then, in the "Receptor-Ligand Interactions" menu, select "Dock Ligands (LibDock)" in the "Dock Ligands" submenu. In the pop-up box, select the ligand as the newly established ligand model-ALL, and select the receptor as the newly established receptor model-ALL, and the sites sphere as the sphere coordinates just established. Finally click RUN to start docking.

4. Calculate the binding energy and choose the pose with the largest binding energy. After docking is complete, many locations of ligand will be displayed. Open the docked view, and click the "Caculate Binding Energies" button in the "Dock Ligands" submenu of the "Receptor-Ligand Interactions" menu. In the pop-up box, select the receptor as the default value, select ligand as the docked model -ALL, and then start to calculate the binding energy. Finally, compare the binding energy and choose the pose with the largest binding energy. The better the stability of the complex, the greater the binding energy.

5. Export the joint section view. For the docked view, after setting the display style of the binding area, click the "Show 2D Map" button in the "View Interaction" submenu of the "Receptor-Ligand Interaction" menu to pop up the view of the binding section. This view can be saved as a picture file.

2.6 Protein docking technology

Discovery-Studio's ZDOCK is another molecular docking tool for studying protein interactions. We used it to study the attack of hemoglobin by viral non-structural proteins. The following is the docking of orf1ab and hemoglobin, and other docking methods with virus non-structural proteins are the same. After opening the PBD files of Human Oxy-Hemoglobin 6bb5 and orf1ab protein, click the "Dock proteins (ZDOCK)" button of "Dock and Analyze Protein Comlexes" under the "macromolecules" menu. In the pop-up interface, select Human Oxy-Hemoglobin 6bb5 as the receptor, orf1a as the ligand, and then click the "Run" button. After the computer finishes computing, click on the "proteinpose" interface and select the pose and cluster with the highest ZDOCK score. It could obtain the position of orf1ab on Human Oxy-Hemoglobin 6bb5. Deloxy HUMAN HEMOGLOBIN 1a3n has a similar docking pattern with orf1ab protein.

3 RESULTS

3.1 Virus structural proteins binding porphyrin

In humans, hemoglobin can be degraded into globin and heme. Heme is composed of a porphyrin and an iron ion, and the iron ion is in the middle of the porphyrin. Heme is insoluble in water and can be combined with heme-binding proteins to form a complex and be transported to the liver. The porphyrin is degraded into bilirubin and excreted from the bile duct, and Iron in the molecule can be reused by the body. If virus proteins can bind to the porphyrin of the heme, they should have the similar binding ability to the human heme-binding protein, that is, the virus proteins and heme-binding proteins should have the similar conserved domains. To examine the binding of virus structure proteins and porphyrin, the following bioinformatics methods were applied in this paper.

First, MEME's online server was employed to search for conserved domains in each viral structure protein and human heme-binding protein (ID:NP_057071.2 heme-binding protein 1, ID: EAW47917.1 heme-binding protein 2). Figure 2 presents that three viral proteins (surface glycoprotein, envelope protein and nucleocapsid phosphoprotein) and heme-binding proteins have conserved domains, but membrane glycoprotein do not have any conserved domains. p-value values are small, there were statistically significant. The domains in three viral proteins are different, suggesting the structural protein's ability to bind porphyrin may be slightly different. Membrane glycoprotein could not bind to theporphyrin.
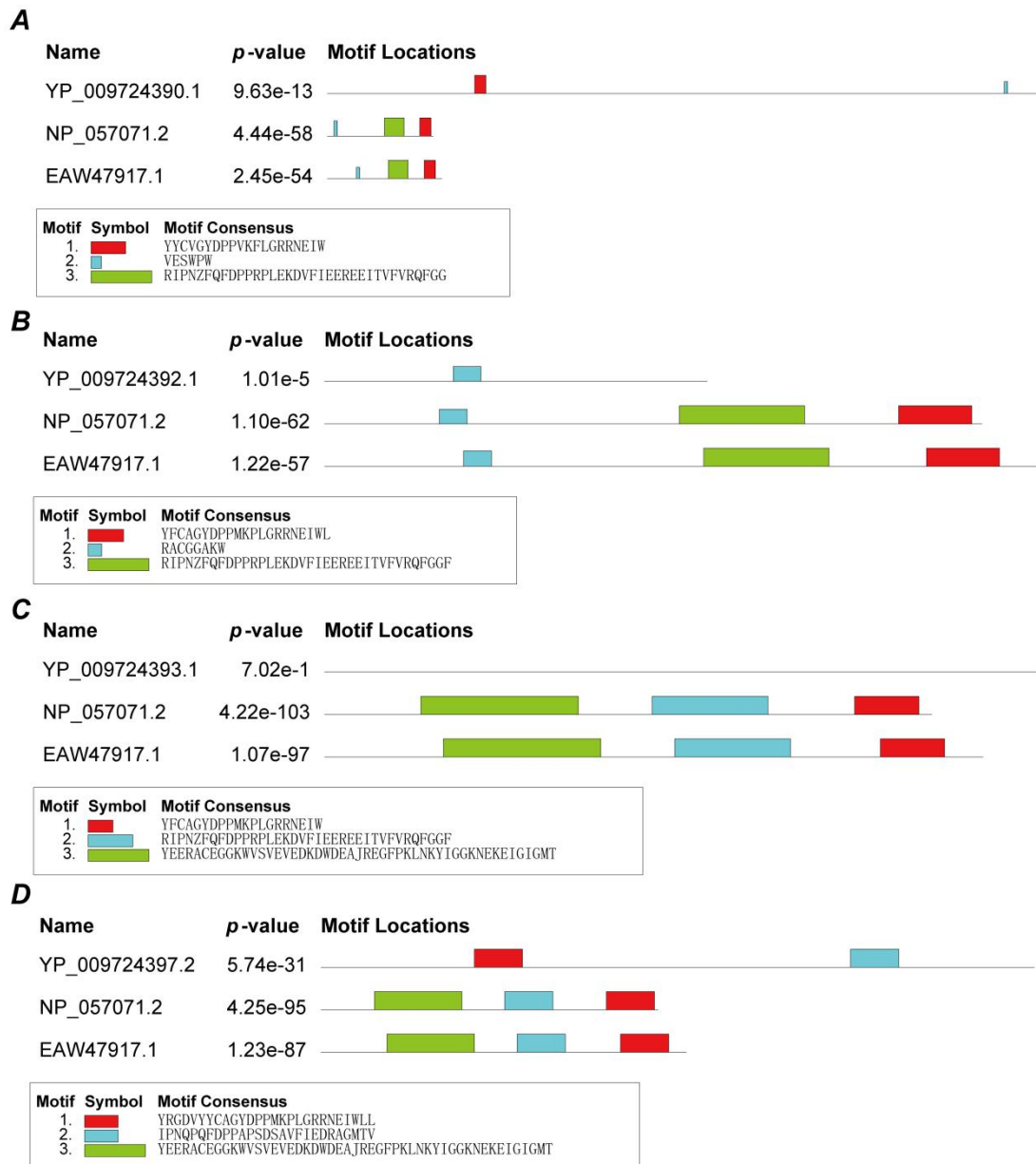
Figure 2. Conserved domains between structure proteins and human heme-binding proteins. *A*. Conserved domains of surface glycoprotein. *B*. Conserved domains of envelope protein. *C*. Conserved domains of membrane glycoprotein. *D*. Conserved domains of nucleocapsid phosphoprotein.

Next, the Swiss-model online server modeled the surface glycoproteins to produce a three-dimensional structure, and two kinds of files based on the Spike and E2 templates were selected. The 3D-structural file of heme was downloaded from the PDB database.

In the end, Discovery-Studio realized molecular docking of surface glycoproteins and the porphyrin. The docking of the Spike protein with heme (and porphyrin) failed first. E2 glycoprotein (Figure 3.A) is derived from templates 1zva.1.A. The docking of E2 glycoprotein and heme was also fruitless. When the iron ion was removed, the heme became a porphyrin, many kinds of docking were finalized between the E2 glycoprotein and the porphyrin. Calculating the binding energy, the docking pose with the highest binding energy (7,530,186,265.80kcal/mol) was accepted. The docking result is exhibited in Figure 4.A-1, which is the molecular model of E2 glycoprotein binds to the porphyrin.

Figure 4.A-2 provides a two-dimensional view of the binding section, where 18 amino acids of the E2 glycoprotein interact with the porphyrin.

Analysis of envelope protein adopted the same methods. The template 5x29.1.A was selected as the 3D structure template of envelope protein (Figure 3.B). Discovery-Studio found several kinds of docking of the envelope protein and the porphyrin, where the docking pose with the highest binding energy (219,317.76kcal/mol) was chosen. Figure 4.B-1 shows the docking result, which is the molecular model of envelope protein binding to the porphyrin. Figure 4.B-2 is the two-dimensional view of the binding section, where 18 amino acids of the envelope protein interact with the porphyrin.

Same methods were utilized to analyze the nucleocapsid phosphoprotein. The template of the nucleocapsid phosphoprotein was 1ssk.1.A (Figure 3.C). Discovery-Studio provides the docking between the nucleocapsid phosphoprotein and the porphyrin with the highest binding energy (15,532,506.53kcal/mol). Figure 4.C-1 shows the docking result, which is the molecular model of the nucleocapsid phosphoprotein bind to the porphyrin. Figure 4.C-2 is the two-dimensional view of the binding section, where 22amino acids of thenucleocapsid phosphoprotein are bound to the porphyrin. Membrane protein is derived from templates 1zva.1.A.　The docking of membrane protein with heme (and porphyrin) failed. These results signal the surface glycoprotein, envelope protein and nucleocapsid phosphoprotein could bind to the porphyrin to form a complex.

It was found the binding energy of envelope protein was the lowest, the binding energy of E2 glycoprotein was the highest, and the binding energy of nucleocapsid phosphoprotein was medium. It means that binding E2 glycoprotein to the porphyrin is the most stable, the binding of nucleocapsid phosphoprotein to the porphyrin is unstable, and binding envelope protein to the porphyrin is the most unstable.

After that, the following analysis was carried out to discover whether structural proteins attacked the heme and dissociate the iron atom to form porphyrins. Heme has an oxidase called heme oxidase, which oxidizes heme and dissociates the iron ion. If structural proteins could attack heme and dissociate iron ions, it should have a similar conserved domain as a heme oxidase. MEME's online server was manipulated to search for conserved domains of structural proteins and heme oxidase proteins (NP_002124.1: heme oxygenase 1;BAA04789.1: heme oxygenase-2;AAB22110.2: heme oxygenase-2). As a result, conserved domains of structural proteins were not found (Figure 5). Combining the results of the previous analysis, that is, structural proteins could only combine with the porphyrin. It can be possible to inferthe structural proteins did not attack heme and dissociate the iron atom to form the porphyrin.
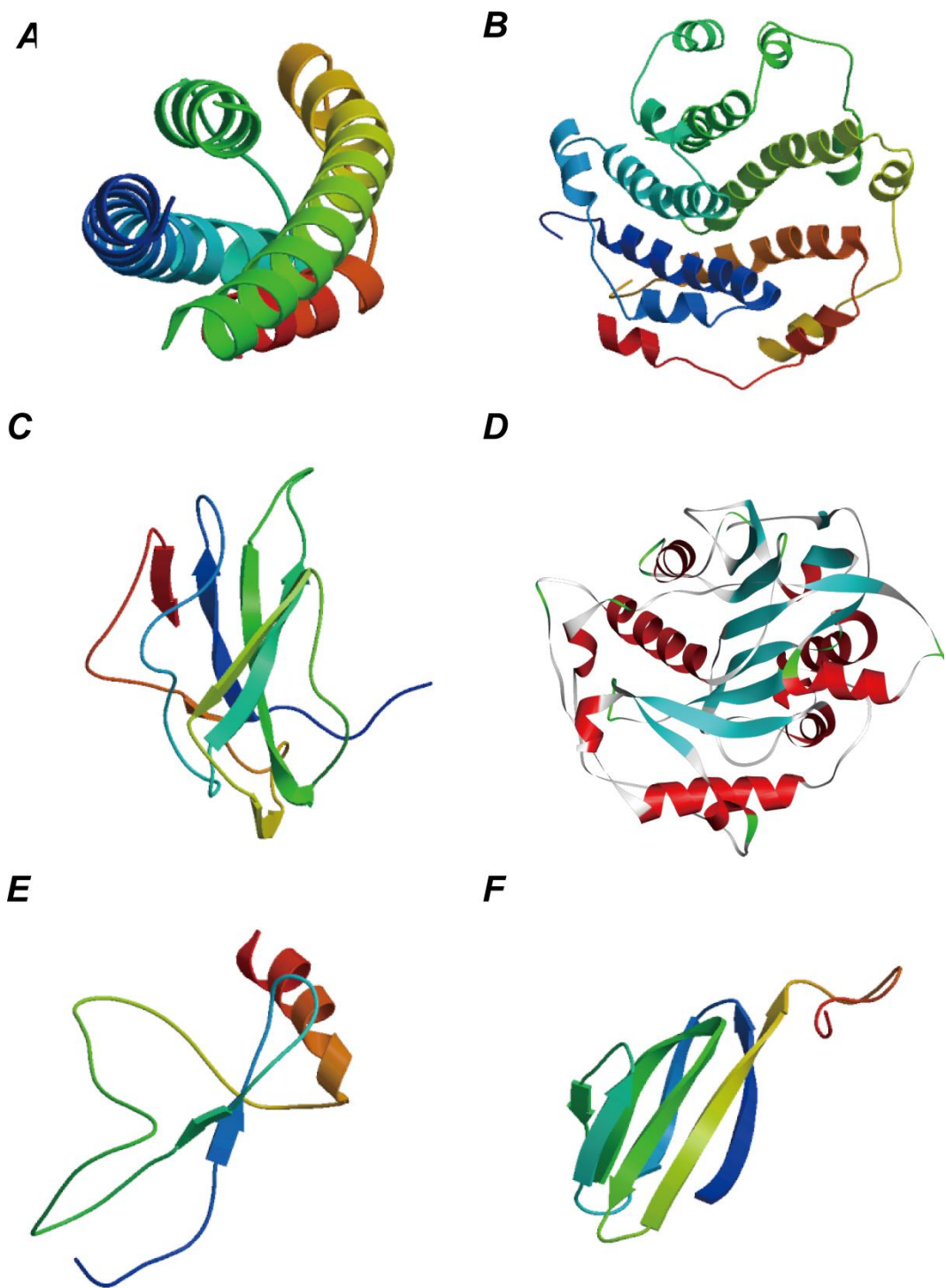
Figure 3. 3D structure schematics of the novel coronavirus proteins by the homology modeling. *A*. E2 glycoprotein of the surface glycoprotein. *B*. Envelope protein. *C*. nucleocapsid phosphoprotein. *D*. orf1ab protein. *E*. ORF8 protein. *F*. ORF7a protein

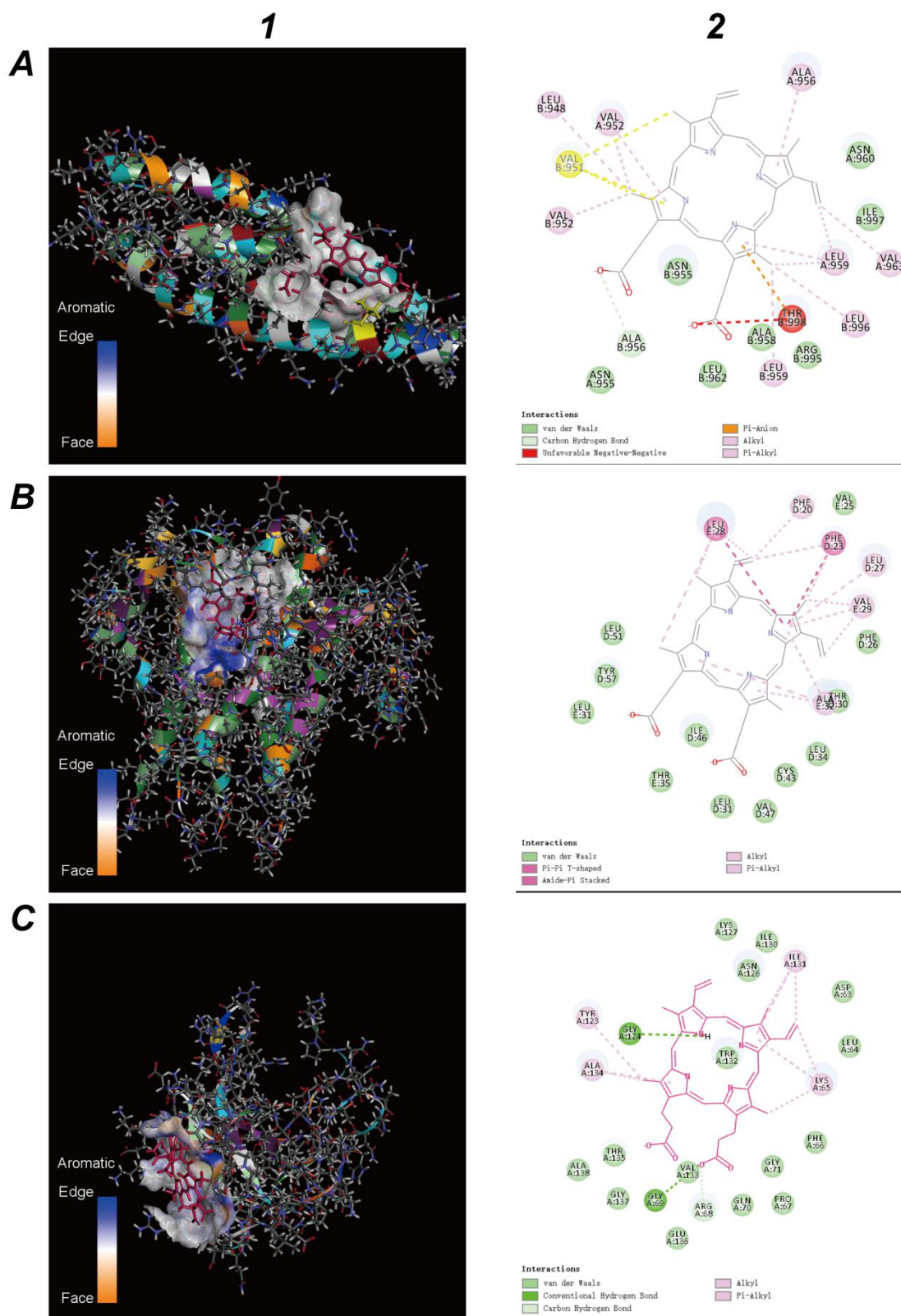**Figure 4.** Molecular docking results of viral structure proteins and the porphyrin (red structure). *A*.Molecular docking results of E2 glycoprotein and the porphyrin. *B*.Molecular docking results of the envelope protein and the porphyrin. *C*.Molecular docking results of the nucleocapsid phosphoprotein and the porphyrin. *1*.Viral structure proteins. *2.* View of the binding sections
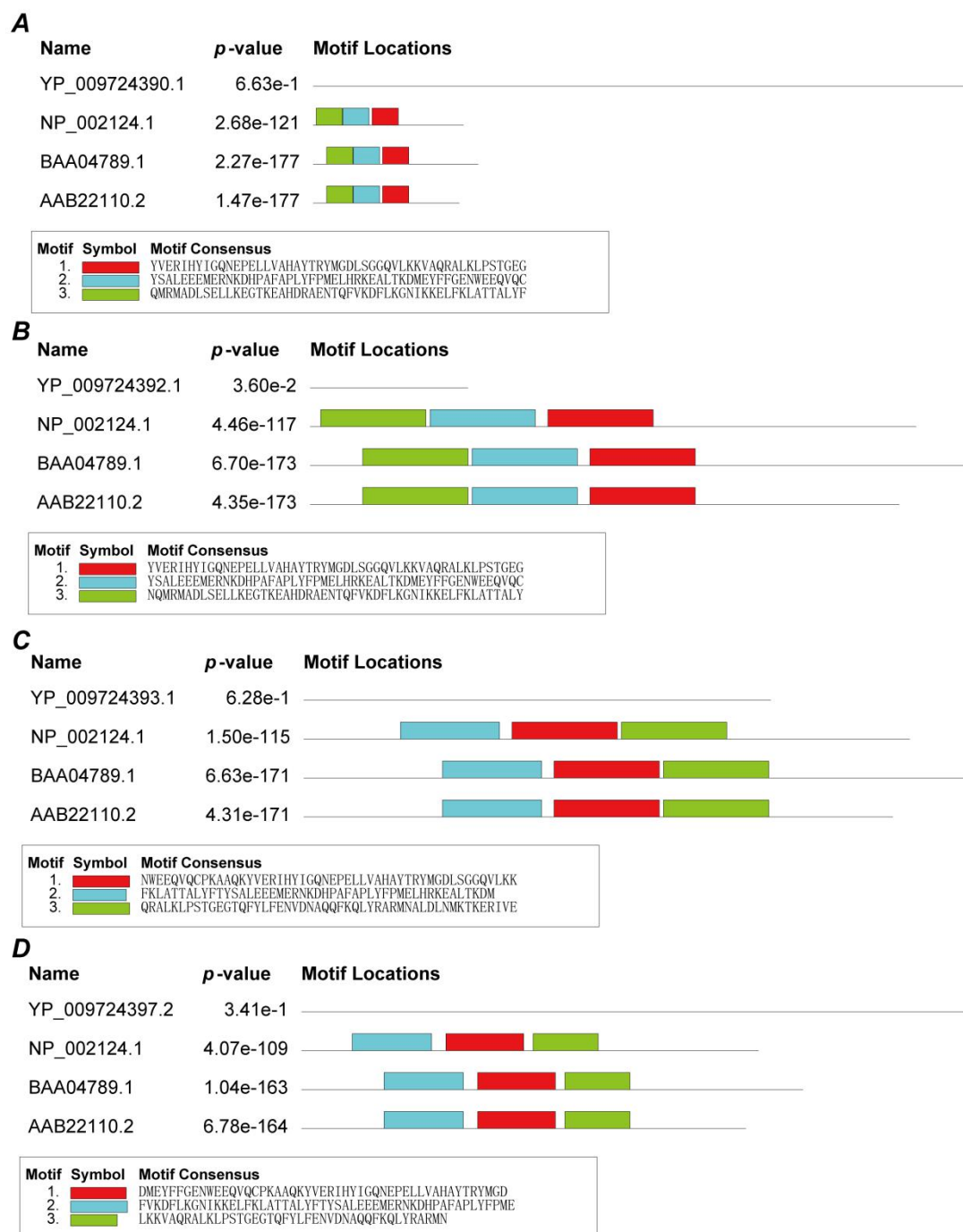
Figure 5. Conserved domains between the structure proteins and human heme oxygenase proteins. *A*. Conserved domains of surface glycoprotein. *B*. Conserved domains of envelope protein. *C*. Conserved domains of membrane. *D*. Conserved domains of nucleocapsid phosphoprotein.

3.2 Virus non-structural proteins bind to the porphyrin

If the non-structure proteins (ID:YP_009724396.1) can bind to the porphyrin of heme, it should have the similar binding ability to the human heme-binding protein. Then, HEME's online server was used to search for conserved domains between the non-structure proteins and human heme-binding proteins. Figure 6 shows that five viral proteins (orf1ab, ORF3a, ORF7a, ORF8 and ORF10) and heme-binding proteins have conserved functional domains, but ORF6 and heme-binding proteins do

not have any conserved functional domains. *p*-value values are small, there were also statistically significant. The domains in the five viral proteins are different, suggesting the non-structural protein's ability to bind porphyrin may be slightly different. ORF6 protein dose not bind to porphyrin.
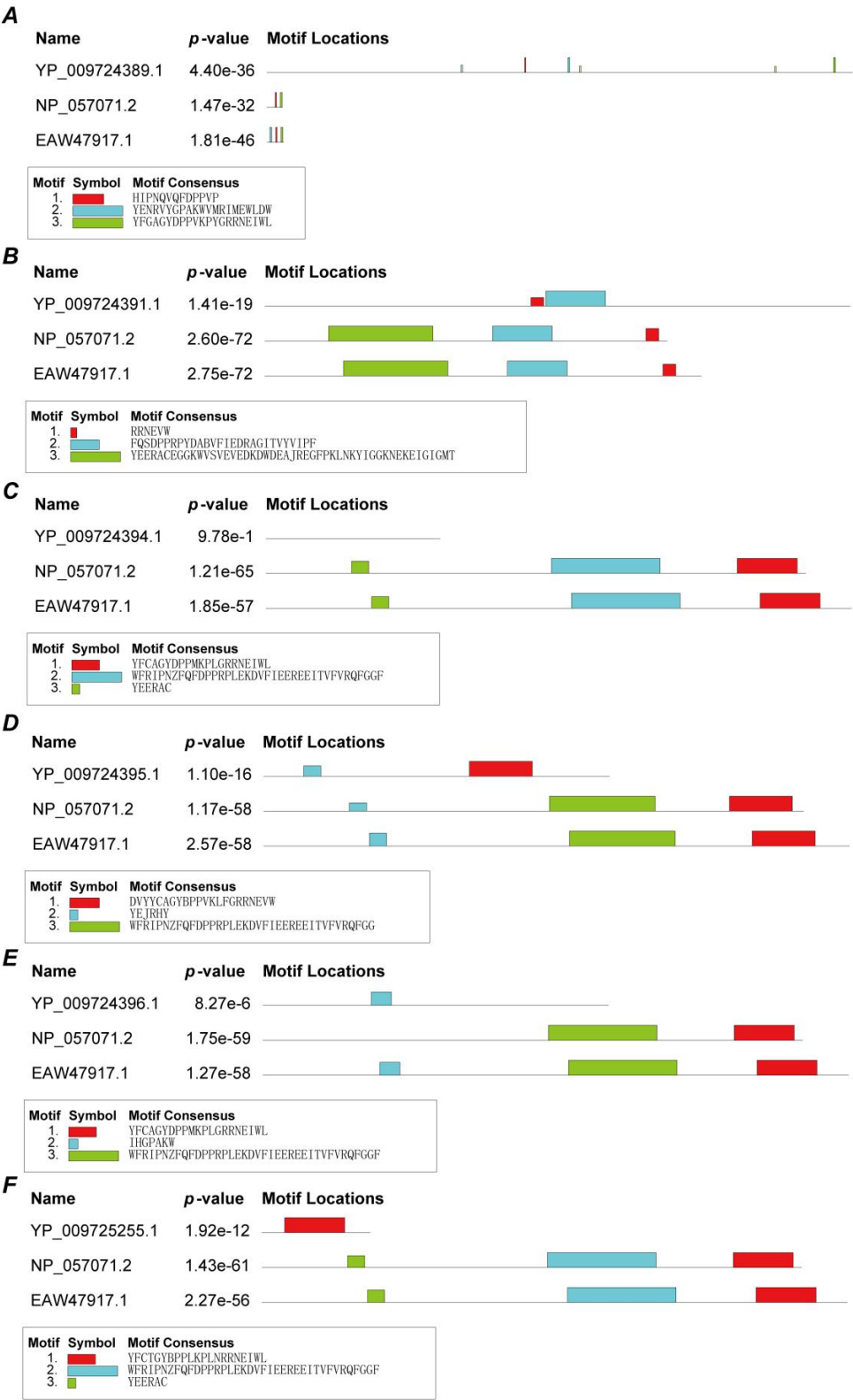

Figure 6. Conserved domains between non-structural proteins and human heme-binding proteins. *A*. Conserved domains of orf1ab. *B*. Conserved domains of ORF3a. *C*. Conserved domains of ORF6. *D*.

Conserved domains of ORF7a. *E*. Conserved domains of ORF8. *F*. Conserved domains of ORF10.

Homology modeling and molecular docking technology were applied to study the characteristics of orf1ab protein's ability to bind heme. Because swiss-model cannot model the 3D structure of orf1ab protein sequence with a sequence length exceeding 5000nt, Discovery-Studio was used to homology modeling. The crystal structure of MERS-CoV nsp10_nsp16 complex 5yn5 and heme were downloaded from the PDB database. In this study, the crystal structure of MERS-CoV nsp10_nsp16 complex 5yn5 was set as a template to create a homologous structure of orf1ab protein. The default homologous structure was selected as the orf1ab protein 3D-structure (Figure 3.D). Then molecular docking of orf1ab protein and porphyrin was finished by Discovery-Studio. orf1ab protein and heme could not complete the docking experiment, but by removing iron ions to make heme into a porphyrin, and the radius of action increased, then several types of docking were completed. By calculating the binding energy, a docking model with the highest binding energy (561,571.10kcal/mol) was selected. The docking result is shown in Figure 7.A-1, where is the molecular model of the orf1ab protein binding to the porphyrin. The binding part of the orf1ab protein acts like a clip. It was this clip that grasps the porphyrin without the iron ion. Figure 7.A-2 shows a two-dimensional view of the binding section. It can be seen that 18 amino acids of the orf1ab protein are bound to the porphyrin.

To study the binding properties of ORF8 protein to heme, the same analysis steps as the structural protein method were used. The structure file was created based on the ORF7 template (Figure 3.E). Several kinds of docking of the ORF8 protein and the porphyrin, where the docking pose with the highest binding energy (12,804,859.25kcal/mol) was selected. The docking result (Figure 7.B-1) represents the molecular model of ORF8 protein binding to the porphyrin. Figure 7.B-2 is the two-dimensional view of the binding section, where 18 amino acids of the ORF8 are bound to the porphyrin.

Same methods of ORF8 protein were used to analyze the ORF7a protein. The ORF7a's template is 1yo4.1.A (Figure 3.F). The ORF7a protein and the porphyrin had the highest binding energy (37,123.79 kcal/mol). Figure 7.C-1 shows the molecular model of the ORF7a binds to the porphyrin. Fifteen amino acids of the ORF7a are bound to the porphyrin (Figure 7.C-2). The binding part of the ORF7a protein also acts like a clip.

Swiss-model could not provide the template for ORF10. ORF6a and ORF3a are derived from templates 3h08.1.A and 2m6n.1.A, respectively, but the docking of ORF6a (ORF3a) with heme (and porphyrin) failed.
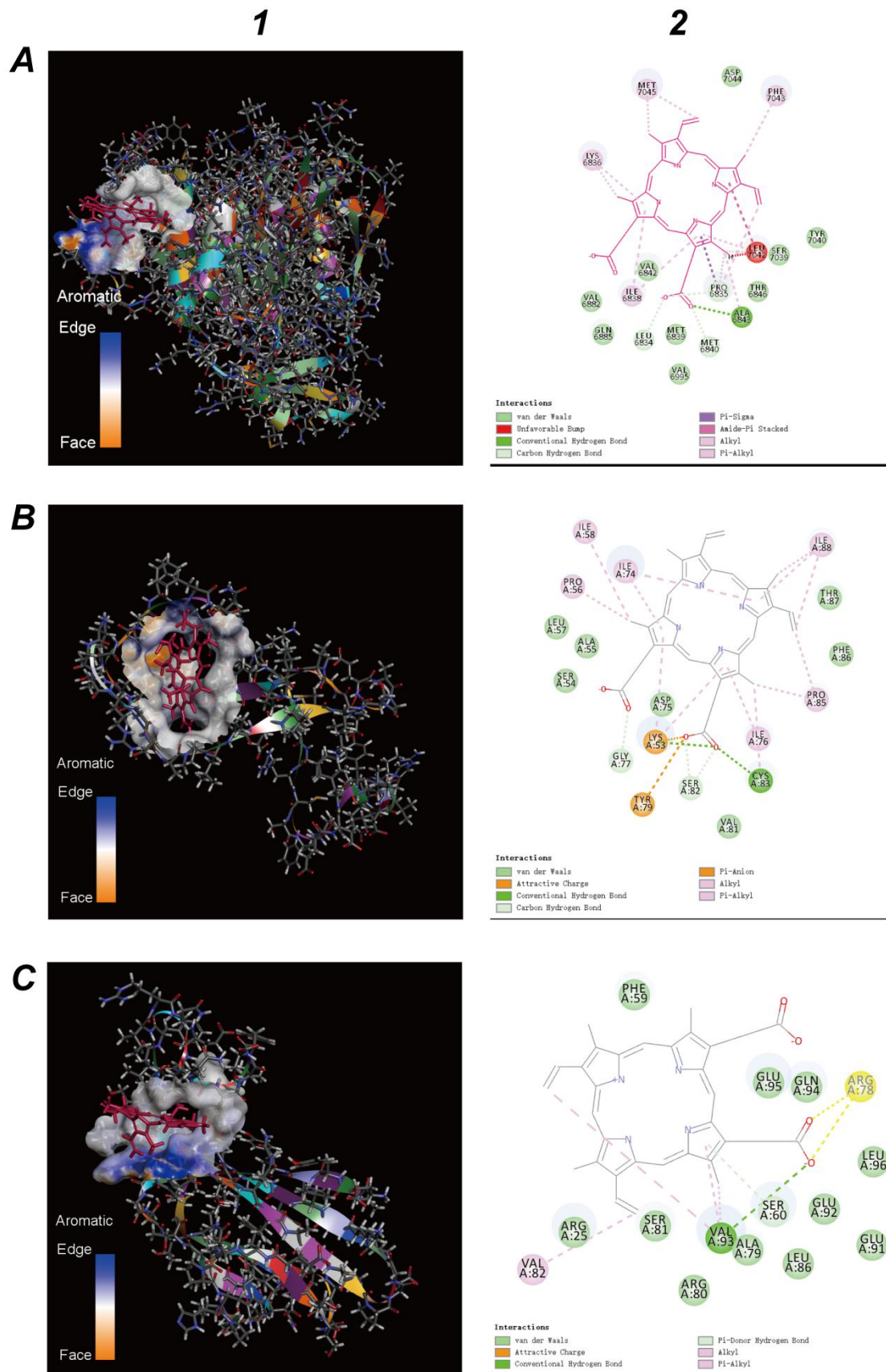
Figure 7. Molecular docking results of viral non-structural proteins and the porphyrin (red). *A*.Molecular docking results of the orf1ab protein and the porphyrin. *B*.Molecular docking results of

the ORF8 protein and the porphyrin. *C.*Molecular docking results of the ORF7a protein and the porphyrin. *1.*Viral non-structural proteins. *2.* View of the binding sections

Finally, the following analysis was performed to find out whether non-structural proteins attacked the heme and dissociated the iron atom to form porphyrins. Here, the same methods as previous structural proteins, MEME's online server, were used to analyze the conserved domains of non-structural proteins and heme oxidase proteins (NP_002124.1: heme oxygenase 1;BAA04789.1: heme oxygenase-2;AAB22110.2: heme oxygenase-2). As shown in Figure 8, ORF10, orf1ab and ORF3a have conserved domains. Combining the results of the previous analysis, it is showed, non-structural proteins: ORF10, orf1ab and ORF3a could attack the heme and dissociate the iron atom to form the porphyrin. However, the *p-Value* of orf1ab and ORF3a is great than 0.1%.Therefore ORF10 may be the primary protein to attack heme, orf1ab and ORF3a capture the heme or the porphyrin.

The results marked that orf1ab, ORF7a, and ORF8 could bind to the porphyrin, while ORF10, ORF3a, and ORF6 could not bind to heme (and porphyrin). ORF10, ORF1ab and ORF3a also have the ability to attack the heme to form a porphyrin. The binding energies of orf1ab, ORF7a, ORF8 and the porphyrin were compared respectively. It was found the binding energy of ORF7a was the lowest, the binding energy of ORF8 was the highest, and the binding energy of orf1ab was medium. This means that binding ORF8 to the porphyrin is the most stable, the binding of orf1ab to the porphyrin is unstable, and binding ORF7a to the porphyrin is the most unstable. The sequences of ORF10 and ORF6 are short, so they should be short signal peptides. Therefore, the mechanism by which non-structural proteins attack heme might be: ORF10, ORF1ab and ORF3a attacked heme and generatedthe porphyrin; ORF6 and ORF7a sentthe porphyrin to ORF8; and ORF8 and the porphyrin formed a stable complex.
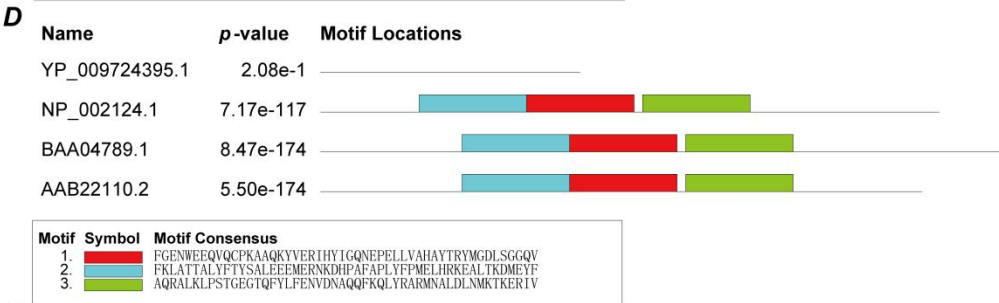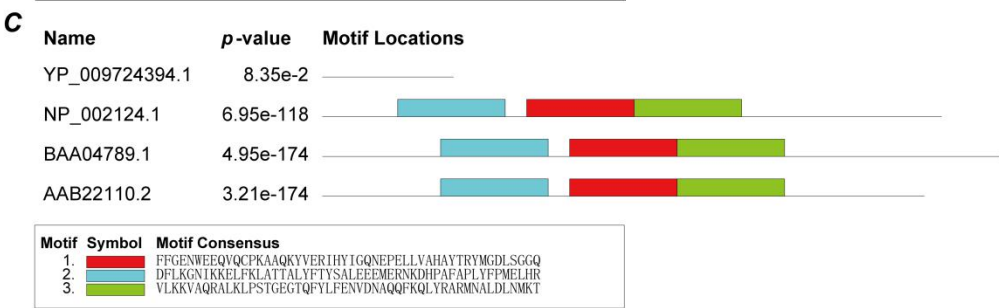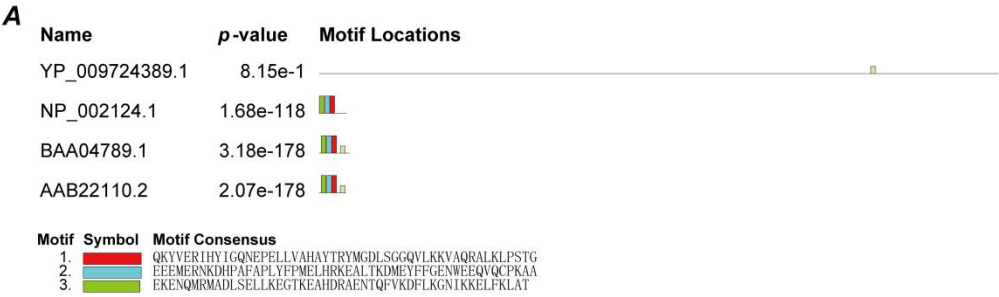
A

Name	*p*-value	Motif Locations
YP_009724389.1	8.15e-1	
NP_002124.1	1.68e-118	
BAA04789.1	3.18e-178	
AAB22110.2	2.07e-178	

Motif	Symbol	Motif Consensus
1.		QKYVERIHYIGQNEPELLVAHAYTRYMGDLSGGQVLKKVAQRALKLPSTG
2.		EEEMERNKDHPAFAPLYFPMELHRKEALTKDMEYFFGENWEEQVQCPKAA
3.		EKENQMRMADLSELLKEGTKEAHDRAENTQFVKDFLKGNIKKELFKLAT

B

Name	*p*-value	Motif Locations
YP_009724391.1	1.43e-1	
NP_002124.1	9.87e-72	
BAA04789.1	8.21e-165	
AAB22110.2	5.36e-165	

Motif	Symbol	Motif Consensus
1.		DMEYFFGENWEEQVQCPKAAQKYVERIHYIGQNEPELLVAHAYTRYMGD
2.		SGGQVLKKVAQRALKLPSTGEGTQFYLFENVDNAQQFKQLY
3.		NKAFEYNMQIFNELDQAGSTLARETLEDGFPVHDGKGDMRKCPFYAAEQD

C

Name	*p*-value	Motif Locations
YP_009724394.1	8.35e-2	
NP_002124.1	6.95e-118	
BAA04789.1	4.95e-174	
AAB22110.2	3.21e-174	

Motif	Symbol	Motif Consensus
1.		FFGENWEEQVQCPKAAQKYVERIHYIGQNEPELLVAHAYTRYMGDLSGGQ
2.		DFLKGNIKKELFKLATTALYFTYSALEEEMERNKDHPAFAPLYFPMELHR
3.		VLKKVAQRALKLPSTGEGTQFYLFENVDNAQQFKQLYRARMNALDLNMKT

D

Name	*p*-value	Motif Locations
YP_009724395.1	2.08e-1	
NP_002124.1	7.17e-117	
BAA04789.1	8.47e-174	
AAB22110.2	5.50e-174	

Motif	Symbol	Motif Consensus
1.		FGENWEEQVQCPKAAQKYVERIHYIGQNEPELLVAHAYTRYMGDLSGGQV
2.		FKLATTALYFTYSALEEEMERNKDHPAFAPLYFPMELHRKEALTKDMEYF
3.		AQRALKLPSTGEGTQFYLFENVDNAQQFKQLYRARMNALDLNMKTKERIV

E

Name	*p*-value	Motif Locations
YP_009724396.1	2.15e-1	
NP_002124.1	5.45e-117	
BAA04789.1	1.26e-173	
AAB22110.2	8.17e-174	

Motif	Symbol	Motif Consensus
1.		FFGENWEEQVQCPKAAQKYVERIHYIGQNEPELLVAHAYTRYMGDLSGGQ
2.		FLKGNIKKELFKLATTALYFTYSALEEEMERNKDHPAFAPLYFPMELHRK
3.		KVAQRALKLPSTGEGTQFYLFENVDNAQQFKQLYRARMNALDLNMKTKER

F

Name	*p*-value	Motif Locations
YP_009725255.1	4.32e-13	
NP_002124.1	2.21e-61	
BAA04789.1	4.97e-147	
AAB22110.2	3.27e-147	

Motif	Symbol	Motif Consensus
1.		DMEYFFGENWEEQVQCPKAAQKYVERIHYIGQNEPELLVAHAYTRYMGD
2.		LYFTYSALEEEMERNKDHPAFAPLYFP
3.		NKAFEYNMQIFNELDQAGSTLARETLEDGFPVHDGKGDMRKCPFYAAEQD

Figure 8. Conserved domains between non-structure proteins and human heme oxygenase proteins. *A*. Conserved domains of orf1ab. *B*. Conserved domains of ORF3a. *C*. Conserved domains of ORF6. *D*. Conserved domains of ORF7a. *E*. Conserved domains of ORF8. *F*. Conserved domains of ORF10.

3.3 Viral non-structural protein attacks the heme on the beta chain of the hemoglobin

Porphyrins in the human body are mostly iron porphyrins, that is, heme. And a lot of heme is not free, but bind to hemoglobin. There was a massive demand of porphyrins for viruses to survive. Therefore, the novel coronavirus targeted hemoglobin, attacked heme and hunted porphyrins. The previous analysis results showed that ORF1ab, ORF3a, and ORF10 have domains similar to heme oxygenase, but only ORF1ab could bind to porphyrin. To study the attack behavior of orf1ab, ORF3a, and ORF10 proteins, we used the ZDOCK molecular docking technology to examine these three proteins. ZDOCK molecular docking technology can analyze protein interactions and find the approximate positions of these three proteins on hemoglobin.

First, we downloaded heme oxygenase 2 (5UC8) from the PDB and used it as a template, and then utilized the homology modeling tool of Discovery-Studio to generate the 3D structure of ORF10 (Figure 9). Since hemoglobin has two forms of oxidation and deoxygenation, the following analysis also performs protein molecular docking in these two cases, taking the posture with the highest ZDOCK score.
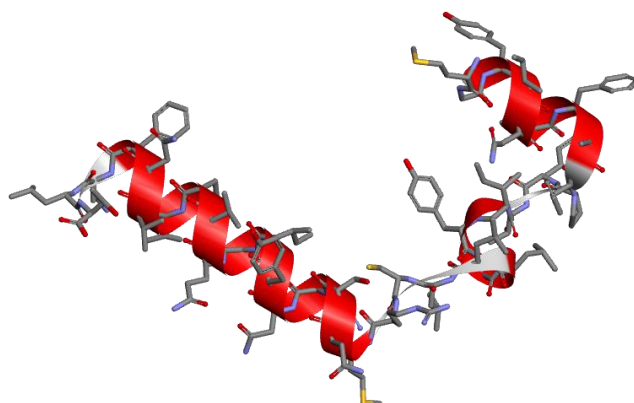

Figure 9. homology modeling of ORF10

For deoxyhemoglobin, orf1ab parked in the middle bottom of the 1-alpha and 2-alpha chain near the 2-alpha chain (Figure 10.A). ORF3a parked in the middle bottom of the 1-alpha and 2-alpha chain near the 2-alpha chain (Figure 10.B). ORF10 moored on the middle bottom of the 1-beta and 2-beta chain near the 1-beta chain (Figure 10.C). The possible mechanism was that orf1ab hit the 2-alpha chain, causing conformations changes in globin protein. ORF3A forced to the 2-alpha chain to attack the 1-beta chain and exposed its heme. ORF10 quickly attached to the 1-beta chain and directly impacted the heme of 1-beta chain. When the iron atom dissociated, the heme changed into porphyrin, and orf1ab finally captured porphyrin. orf1ab played a vital role throughout the attack.
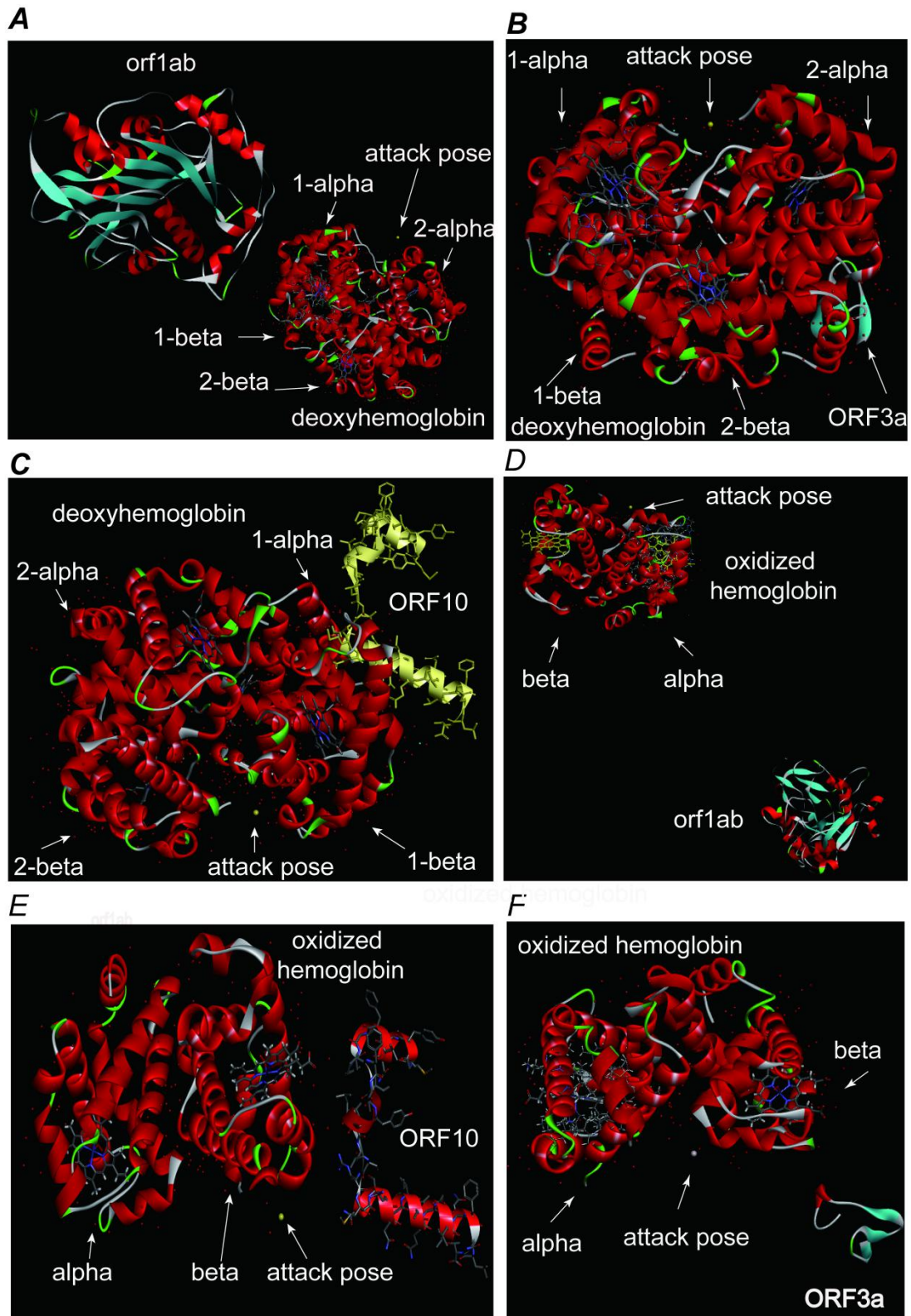
Figure 10. Viral non-structural protein attack hemoglobin. *A*. orf1ab attacks the deoxyhemoglobin. *B*. ORF3a attacks the deoxyhemoglobin. *C*. ORF10 attacks the deoxyhemoglobin. *D*. orf1ab attacks the oxidized hemoglobin. *E*. ORF10 attacks the oxidized hemoglobin. *F*. ORF3a attacks the oxidized hemoglobin.

For oxidized hemoglobin, orf1ab parked in the middle bottom of the alpha and beta chain and closed to the alpha chain (Figure 10.A). ORF10 parked below of the beta chain and neared to the outer (Figure 10.B). ORF3a parked in the middle bottom of the alpha and beta chain and neared to the beta chain (Figure 10.C). The possible mechanism was that orf1ab bound to the alpha chain and attacked the beta chain, causing configurational changes in the alpha and beta chains; ORF3 attacked the beta chain and exposed heme. ORF10 quickly attached the beta chain and directly impacted the iron atoms on the heme of the beta chain. The heme was dissociated into porphyrin, and orf1ab finally captured porphyrin. orf1ab played a vital role throughout the attack.

Attack of oxidized hemoglobin by viral proteins will lead to less and less hemoglobin that can carry oxygen. The invasion of viral proteins on deoxidized hemoglobin will cause less and less hemoglobin that can carry carbon dioxide and blood sugar. People with diabetes can have unstable blood sugar. The patient is aggravated by carbon dioxide poisoning. The lung cells have extremely intense inflammation due to the inability to exchange carbon dioxide and oxygen frequently, which eventually results in ground-glass-like lung images. Patients with respiratory distress will be made worse.

3.4 Validation for the effect of chloroquine phosphate

The chemical components in chloroquine phosphate compete with the porphyrin and bind to the viral protein, thereby inhibiting the viral protein's attack on heme or binding to the porphyrin. To verify the effect of chloroquine phosphate on the viral molecular mechanism of action, molecular docking technology was accepted. The structure file of 0TX (chloroquine) was downloaded from the PDB database. Then molecular docking technology of Discovery-Studio 2016 was used to test the effects of viral proteins and chloroquine.

Figure 11.A-1 is a schematic diagram of binding chloroquine to a virus surface glycoprotein. Figure 11.A-2 is the binding region of virus surface glycoprotein. 13 amino acids engaged in the binding. The binding energy of chloroquine to the E2 glycoprotein of the virus is 3,325,322,829.64 kcal/mol, which is about half the binding energy of the E2 glycoprotein and the porphyrin. According to the results of Figure 4.A-2, further analysis showed that some amino acids (for example VAL A:952, ALA A:956, ALA B:956, ASN A:955 etc.) of the E2 glycoprotein could bind to not only chloroquine phosphate, but also the porphyrins. In other words, the chloroquine has a one-third chance of inhibiting viral E2 glycoprotein and reducing patient symptoms.

The binding view of the chloroquine and envelope protein is shown in Figure 11.B-1. The binding energy of the chloroquine and envelope proteinis7,852.58 kcal/mol, which is only equivalent to 4% of the binding energy of envelope protein and porphyrin. The binding region is shown in Figure 11.B-2. Figure 4.B-2 and Figure 11.B-2representedsome amino acids (such as LEV E:28, PHE: D:20, VAL E:25) of the envelope protein is not only bound to the chloroquine phosphate, but also to the porphyrin.

Figure 11.C-1 is a schematic diagram of binding the chloroquine to the nucleocapsid phosphoprotein. The binding energy of chloroquine to the nucleocapsid phosphoprotein is 198,815.22 kcal/mol, which is only equivalent to the 1.4% of the binding energy of the nucleocapsid phosphoprotein and the porphyrin. ALA A:50 etc.of nucleocapsid phosphoprotein are involved in binding (Figure 12.C-2). Figures 4.C-2 and Figures 11.C-2 declared that amino acids of nucleocapsid phosphoprotein could bind the porphyrin, but could not bind chloroquine.
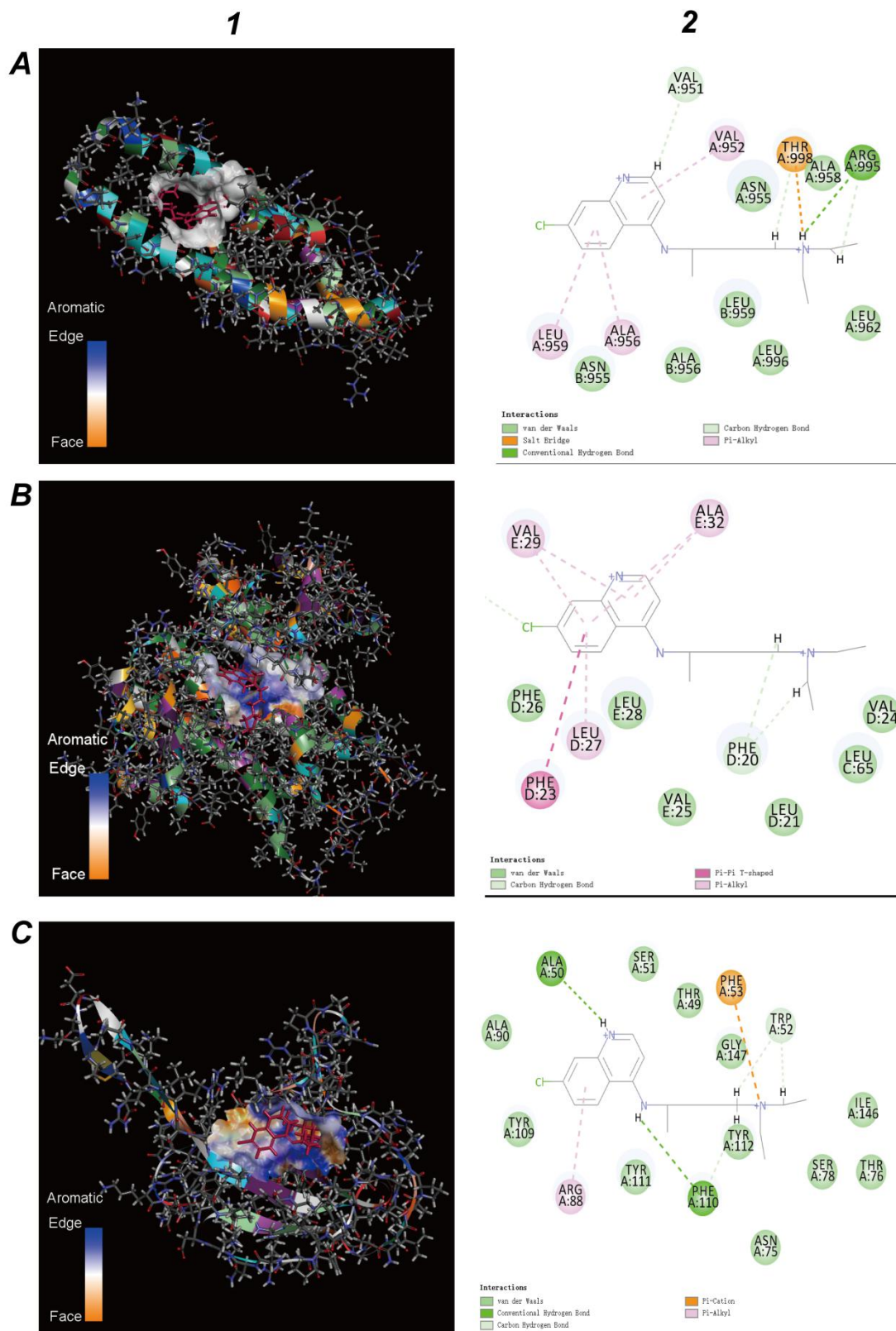
The docking of membrane protein with chloroquine failed.

Figure 11. Molecular docking results of viral structure proteins and the chloroquine (red). *A*.Molecular docking results of the E2 glycoprotein and the porphyrin. *B*.Molecular docking results of the envelope protein and the porphyrin. *C*.Molecular docking results of the nucleocapsid phosphoprotein and the porphyrin. *1*.Viral structure proteins. *2.* View of the binding sections

A schematic diagram of binding chloroquine to the orf1ab protein is shown in Figure 12.A-1. Binding section of the orf1ab protein is plotted in Figure 12.A-2. The binding energy of chloroquine and the orf1ab protein is 4,584,302.64 kcal/mol, which equals 8-fold of the binding energy between the orf1ab and the porphyrin. According to the results of Figure 7.A-2, it was shown that some amino acids such as MET 7045, PHE 7043, LYS 6836 of orf1ab protein could be not only bound to chloroquine phosphate, but also to porphyrin.

A schematic diagram of binding chloroquine to the ORF8 protein is shown in Figure 12.B-1. Figure 12.B-2 shows the binding section of the ORF8. The binding energy of chloroquine to the ORF8 protein is 4,707,657.39 kcal/mol, which is only equivalent to 37% of the binding energy of the ORF8 protein to the porphyrin. According to the result of Figure 7.B-2, it showed the amino acids like as ILE A:74, ASP A:75,LYS A:53 of ORF8 could not only bind to chloroquine phosphate, but also to the porphyrin.

A schematic diagram of binding chloroquine to the ORF7a protein is shown in Figure 12.C-1. Figure 12.C-2 is the view of the binding section. The binding energy of chloroquine to the ORF7a protein is 497,154.45 kcal / mol, which equals 13-fold of the binding energy of the ORF7a protein to the porphyrin. According to the results of Figure 7.C-2, it was shown the amino acids such as GLN A:94, ARG A:78 and LEU A:96 of ORF7aprotein could be not only bound to chloroquine phosphate, but also to the porphyrin.

The docking of ORF3a, ORF6 and ORF10 proteins with chloroquine failed.

These results marked the chloroquine could inhibit E2 and ORF8 bind to the porphyrin to form a complex respectively to a certain extent. Meanwhile, chloroquine could prevent orf1ab, ORF3a and ORF10 to attack the heme to form the porphyrin.
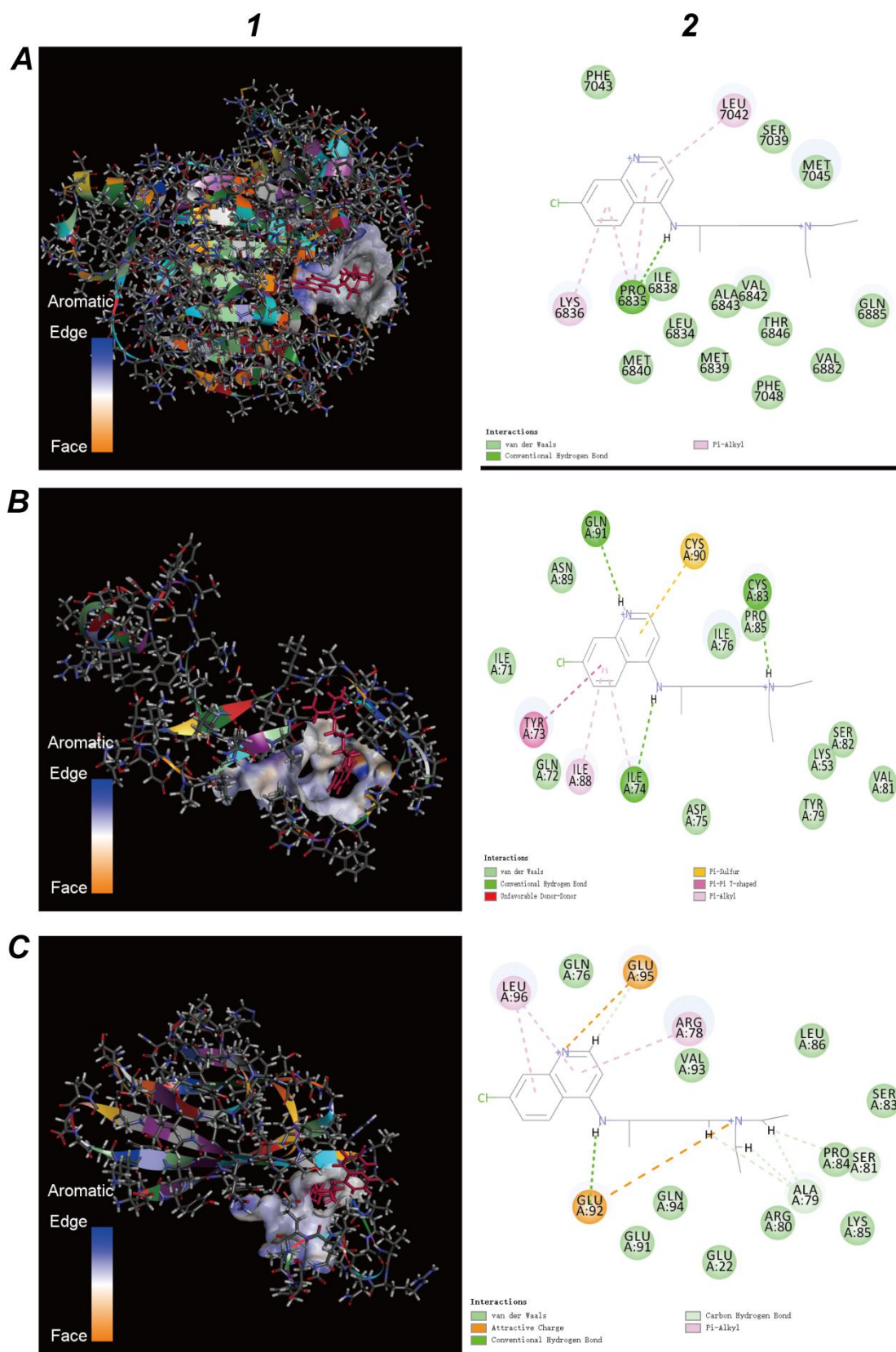
Figure 12. Molecular docking results of viral non-structural proteins and the chloroquine (red structure). *A*.Molecular docking results of the orf1ab protein and the chloroquine. *B*.Molecular docking results of the ORF8 protein and the chloroquine. *C*.Molecular docking results of the ORF7a protein and the chloroquine. *1*.Viral non-structural proteins. *2*. View of the binding sections.

3.5 Validation for the effect of Favipiravir

Favipiravir is the latest anti-novel coronavirus drug with specific therapeutic effects. In Favipiravir, the most critical ligand is 1RP, which is 6 - fluoro - 3 - oxo - 4 - (5 - O - phosphono - beta - D - ribofuranosyl) - 3, 4 - dihydropyrazine - 2 - carboxamide. Following the same method used to examine chloroquine, we investigated the efficacy of Favipiravir. As can be seen from Table 1, Favipiravir cannot be bind to E2 glycoprotein and Nucleocapsid, and its binding energy to Envelope protein, ORF7a, orf1ab is higher than that to porphyrin. It is useful to note that the binding energy of Envelope protein and Favipiravir is more than 2700 times the binding energy of porphyrin. The primary function of Envelope protein is to help the virus enter host cells, which shows that Favipiravir can effectively prevent the virus from infecting human cells. The binding energy of ORF7a to Favisiravir is 450 times higher than that of porphyrin, indicating that it can effectively avoid the non-structural protein of the virus capturing porphyrin. The binding energy of orf1ab and Favisiravir is 1.8 times higher than that of porphyrin, which shows that Faifiravir can prevent virus unstructured protein from attacking heme on hemoglobin. According to previous studies, the binding energy of orf1ab and Favisiravir is much smaller than that of chloroquine, so Favisiravir's ability to improve respiratory distress is lower. In summary, the primary role of Favipiravir is to prevent the virus from entering host cells and catching free porphyrins.

Table 1. Effect of Favipiravir

Virus potein	Porphyrin (kcal/mol)	Favipiravir (kcal/mol)	Has Identical residues	Target	Target Rate (Favipiravir /Porphyrin)
E2 glycoprotein	7,530,186,265.80	-	-	-	-
Envelope protein	219,317.76	597,814,480.55	Yes	Yes	2,725.79
Nucleocapsid	15,532,506.53	-	-	-	-
orf1ab	561,571.10	1,052,489.88	Yes	Yes	1.87
ORF8	12,804,859.25	348,589.80	Yes	-	-
ORF7a	37,123.79	17,034,560.60	Yes	Yes	458.86

4 Discussion

4.1 The novel coronavirus originated from an ancient virus

For the most primitive-life viruses, it isn't very easy to see their role in binding the porphyrin. The porphyrin compounds are widely present in photosynthetic or non-photosynthetic organisms, and they are associated with critical physiological processes such as catalysis, oxygen transfer, and energy transfer. The porphyrin is also an ancient compound widely present on the earth. The porphyrin is first found in crude oil and asphalt rock in 1934. The porphyrin has unique photoelectronic properties and excellent thermal stability and has broad application prospects in materials chemistry, medicine, biochemistry, and analytical chemistry. It is excellent performance in two-photon absorption, fluorescence effect, energy transfer, and other aspects. Fluorescence resonance energy transfer (FRET)

is a non-radiative process in which a donor in an excited state transfers energy to a receptor in the ground state through a long-range dipole effect. The FRET characteristics of the porphyrin may be the primary survival mode on which the original virus relied.

There are numerous theories about the origin of viruses, one of which is called co-evolution theory, which viruses can evolve from complexes of the protein and the nucleic acid . Various methods do not explain that a virus survived independently of non-appearing cells at the beginning of life, so the origin of a virus remains a mystery. This paper proposes that a virus could be bind to the porphyrin, which could explain the survival problem of an original virus. Because the porphyrin has the energy transfer characteristic of fluorescence resonance, viruses that bind to porphyrins could obtain energy through this light-induced method. A virus that gained power could achieve minimal displacement movements, or wake itself from hibernation, or enter hibernation from an active state. Depending on the research results in this study, the novel coronavirus was a life form dependent on the porphyrin. Therefore, we could believe that the novel coronavirus originated from an ancient virus that may have evolved over countless generations in bats.

4.2 Higher permeability of porphyrins into cell membranes leads to high infections

Highly evolving of the novel coronavirus also displays some paradoxical characteristics. The current theory suggests that the novel coronavirus binds to the human ACE2 receptor through a spike protein. It enters human cells in the form of phagocytosis. Infectious disease models indicated that the novel coronavirus pneumonia is highly contagious. Therefore, the spike protein and human ACE2 protein should have a strong binding ability, but there are reports in the literature that this binding ability is weak. What causes the high infectivity of the novel coronavirus? We believe that in addition to the invasive method of spike-ACE2, it should maintain the original invasive pattern.

Medical workers have detected the novel coronavirus from urine, saliva, feces, and blood. The virus can also live in body fluids. In such media, porphyrin is a prevalent substance. Porphyrin compounds are a class of nitrogen-containing polymers, and existing studies have found that they have a strong ability to locate and penetrate cell membranes. At the beginning of life, virus molecules with porphyrins directly moved into the original membrane structure by porphyrin permeability. This study showed that the E2 glycoprotein and Envelope protein of the novel coronavirus could bind well to porphyrins. Therefore, the coronavirus may also directly penetrate the human cell membrane through porphyrin, so the infection is robust. Our validation analysis showed that Favipiravir could only prevent the binding of Envelope protein and porphyrin. Meanwhile, chloroquine could effectively prevent the binding of E2 glycoprotein to porphyrin to a certain extent. Therefore, the infectivity of the novel coronavirus pneumonia was not completely prevented by the drugs, because of the the binding of E2 glycoprotein and porphyrin was not inhibited.

4.3 Higher hemoglobin caused higher morbidity

The therapeutic effect of chloroquine phosphate on novel coronavirus pneumonia shows that novel coronavirus pneumonia might be closely related to abnormal hemoglobin metabolism in humans. The number of hemoglobin is a major blood biochemical indicator, and the content is different in different genders. The number of normal men is significantly higher than that of normal women, which might also be a reason why men are more likely to be infected with the novel coronavirus pneumonia than women. Besides, patients of novel coronavirus pneumonia are most of the middle-aged and older adults. Many of these patients have underlying diseases such as diabetes. Diabetic patients have higher

glycated hemoglobin. Glycated hemoglobin is deoxyhemoglobin. Glycated hemoglobin is a combination of hemoglobin and blood glucose, which is another reason for the high infection rate for older people.

This study has confirmed that orf1ab, ORF3a, and ORF10 could coordinately attack heme on the beta chain of hemoglobin. Both oxygenated and deoxygenated hemoglobin are attacked. During the attack, the positions of orf1ab, ORF3, and ORF10 are slightly different. It showed that the higher the hemoglobin content, the higher the risk of disease. However, it is not sure that the disease rate caused by abnormal hemoglobin (structural) is relatively low. The hemoglobin of patients and recoverers should be detected for further research and treatment.

4.4 Inhibiting the heme anabolic pathway and causing the disease

This article considered the virus directly interfered with the assembly of human hemoglobin. The main reason was the normal heme was too low. Heme joins in critical biological activities such as regulation of gene expression and protein translation. Porphyrin is an important material for the synthesis of heme. Because the existing traces show there is too much free iron in the body, it should be that virus-producing molecule competes with iron for the porphyrin. Inhibiting the heme anabolic pathway and causing symptoms in humans.

It is not clear whether the spatial molecular structure of heme and porphyrin in patients with porphyria is the same as that in healthy people. If there is an abnormal structure, it is not clear whether this porphyrin can bind to a viral protein to form a complex, or if a viral protein can attack this heme. It should be proved by clinical and experimental research.

4.5 Viral protein infects hemoglobin by the immune hemolysis of red blood cells

Red blood cells mainly contain hemoglobin. During hemolysis, hemoglobin escapes from the cells and dissolves in the plasma. At this time, the ability of hemoglobin to carry oxygen is lost. Hemolysis occurs due to rupture of red blood cell membranes and dissolution of the matrix. Or the expansion of the red cell membrane pores allows hemoglobin to escape, leaving behind a double concave disc-shaped cell membrane --- "hematocyte". Immune hemolysis is a definite hemolysis brought about by the antigen-antibody reaction. The non-specific hemolysis is caused by physical, chemical or biological factors. After hemolysis of red blood cells, viral proteins may infect hemoglobin. Considering that some researchers have calculated that people with some type O bloods are not easily infected with COVID-19, we speculate that immune hemolysis may be the main method of viral protein infection of hemoglobin. Viral proteins attack the hemoglobin after the hemoglobin is infected. Resulting from limited computational tools, we cannot simulate whether viral proteins attack hemoglobin outside or inside red blood cells.

4.6 The complexity of individual immunity

Some theories suggest that an immune response occurs in the body after a patient becomes ill. Some patients develop immune antibodies after recovery. According to this study, E2 glycoprotein, envelope protein, nucleocapsid phosphoprotein, orf1ab, ORF7a, and ORF8 of the virus could bind to porphyrin. But from the current research, it is unclear which immune antibodies have been raised against viral proteins.

Besides, some patients may be killed by their cytokine storm. Comparing to patients with SARs, the anatomical characteristics of the dead are different. The complex of virus proteins and the porphyrin may be little soluble. Too much mucus in the tissues of the deceased patients was the cause

of too much mucin protein. Mucin could turn loosely connected cells into tightly adhered cells and increases lubrication between cells. It suggests the compound leads to reduced cell connectivity, and cells need mucin to consolidate tissue-cell connectivity and lubricity. Also, when a patient enters a severe infection period, viral structural proteins were mainly used for virus assembly. Therefore, we cannot find noticeable virus inclusions in tissue cells of the dissected patient.

5 Conclusions

Since the emergency epidemic, it is of high scientific significance to use bioinformatics to analyze the roles of novel coronavirus proteins (such as ORF8 and surface glycoproteins). In this study, domain prediction methods were applied to search for conserved domains. The structure of protein molecules such as ORF8 and surface glycoproteins were obtained using homology modeling methods. Molecular docking technology was used to analyze the binding part of viral proteins to the heme and the porphyrin. The study results show that ORF8 and surface glycoproteins could combine to the porphyrin to form a complex, respectively. At the same time, orf1ab, ORF10, and ORF3a proteins could coordinate attack the heme on the 1-beta chain of hemoglobin to dissociate the iron to form the porphyrin. The attack will lead to less hemoglobin to carry oxygen and carbon dioxide. The lung cells have extremely intense inflammation due to the inability to exchange carbon dioxide and oxygen frequently, which eventually results in ground-glass-like lung images. Patients with respiratory distress will be made worse. Diabetic patients and older people have higher glycated hemoglobin. Glycated hemoglobin was reduced by the attack, which made patients' blood sugar unstable. Since the porphyrin complexes of the virus produced in the human body inhibited the heme anabolic pathway, they caused a wide range of infection and disease.

With these findings in mind, further analysis revealed that chloroquine could prevent orf1ab, ORF3a, and ORF10 from attacking the heme to form the porphyrin, and inhibit the binding of ORF8 and surface glycoproteins to porphyrins to a certain extent, effectively relieve the symptoms of respiratory distress. Since the ability of chloroquine to inhibit structural proteins is not particularly obvious, the therapeutic effect on different people may be different. Favipiravir could inhibit the envelope protein and ORF7a protein bind to porphyrin, prevent the virus from entering host cells, and catching free porphyrins. Due to the side effects and allergic reactions of drugs such as chloroquine, please consult a qualified doctor for treatment details, and do not take the medicine yourself.

This paper is only for academic discussion, the correctness needs to be confirmed by other laboratories. We look forward to a laboratory that can prove whether this theory is wrong or correct from the following experiments: 1. Use X-RAY to detect hemoglobin in critically ill patients to find out if there is any abnormality. 2. It is proved by virus experiments: viral proteins can bind porphyrin; viral proteins can attack heme; viral proteins can attack hemoglobin in the blood.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets and results supporting the conclusions of this article are available at https://pan.baidu.com/s/1v8kP0zAyvnACXm-vJHWJuQ, code: rhnb.
Or : https://mega.nz/folder/ciYywQbZ#KjwG5OflNrXTqDpCjWLd1g

Competing interests

The authors declare that they have no competing interests.

Funding

Author's contribution

Funding was obtained by WZL. Design, analysis, writing: WZL. Data curation, check manuscript: HLL All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

Author details

[1] School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong, 643002, China. [2] School of Life Science and Food Engineering, Yibin University, Yibin, 644000, China.

References

1   Diao, K., Han, P., Pang, T., Li, Y. & Yang, Z. HRCT Imaging Features in Representative Imported Cases of 2019 Novel Coronavirus Pneumonia. *Precision Clinical Medicine* (2020).

2   Chang, D. *et al.* Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China. *JAMA* (2020).

3   Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* (2020).

4   Li, X., Zai, J., Wang, X. & Li, Y. Potential of large 'first generation'human-to-human transmission of 2019-nCoV. *Journal of Medical Virology* (2020).

5   Wang, D. *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama* (2020).

6   Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).

7   Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* (2020).

8   Wu, F. *et al.* A novel coronavirus associated with human respiratory disease in China. *Nature*, 1-8 (2020).

9   Lu, H., Stratton, C. W. & Tang, Y. W. Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle. *Journal of Medical Virology*.

10   Zhu, N. *et al.* China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* (2020).

11 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* (2020).

12 Wang, M. *et al.* A precision medicine approach to managing Wuhan Coronavirus pneumonia. *Precision Clinical Medicine* (2020).

13 Schaecher, S. R. & Pekosz, A. in *Molecular Biology of the SARS-Coronavirus*   153-166 (Springer, 2010).

14 McBride, R. & Fielding, B. C. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* **4**, 2902-2923 (2012).

15 Wu, A. *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe* (2020).

16 Paraskevis, D. *et al.* Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, Genetics and Evolution*, 104212 (2020).

17 Li, S. *et al.* Regulation of the ER Stress Response by the Ion Channel Activity of the Infectious Bronchitis Coronavirus Envelope Protein Modulates Virion Release, Apoptosis, Viral Fitness, and Pathogenesis. *Frontiers in Microbiology* **10**, 3022 (2020).

18 To, K. K.-W. *et al.* Consistent detection of 2019 novel coronavirus in saliva. *Clinical Infectious Diseases* (2020).

19 Dong, N. *et al.* Genomic and protein structure modelling analysis depicts the origin and pathogenicity of 2019-nCoV, a novel coronavirus which caused a pneumonia outbreak in Wuhan, China. *F1000Research* **9**, 121 (2020).

20 Rothe, C. *et al.* Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine* (2020).

21 Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* (2020).

22 Das, R. & Sharma, P. in *Clinical Molecular Medicine*   327-339 (Elsevier, 2020).

23 Kazazian Jr, H. H. & Woodhead, A. P. Hemoglobin A synthesis in the developing fetus. *New England Journal of Medicine* **289**, 58-62 (1973).

24 Liu, J. *et al.* Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *Journal of Medical Virology* (2020).

25 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research*, 1-3 (2020).

26 Bernardo-Seisdedos, G., Gil, D., Blouin, J.-M., Richard, E. & Millet, O. in *Protein Homeostasis Diseases* 389-413 (Elsevier, 2020).

27 Lameda, I. L. P. & Koch, T. R. in *Liver Diseases*   107-116 (Springer, 2020).

28 Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic acids research* **43**, W39-W49 (2015).

29 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**, W202-W208 (2009).

30 Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic acids research* **34**, W369-W373 (2006).

31 Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research* **31**, 3381-3385 (2003).

32 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42**, W252-W258 (2014).

33 Studio, D. Discovery Studio. *Accelrys [2.1]* (2008).