COVID-19 Disease: ORF8 and Surface Glycoprotein Inhibit Heme Metabolism by Binding to Porphyrin

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. Due to the limits of the existing experimental tools, many protein roles of novel coronavirus including ORF8 are still unclear. Therefore, in the current scene of an emergency epidemic, it is of high scientific significance to predict the biological role of viral proteins through bioinformatics methods. 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, while orf1ab, ORF10 and ORF3a proteins could coordinately attack heme to dissociate the iron to form the porphyrin. The mechanism seriously interfered with the normal heme anabolic pathway of the human body, being 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. Therefore, this research is of high value to contemporary biological experiments, disease prevention and clinical treatment.

Keywords: COVID-19; novel coronavirus; ORF8; orf1ab; surface glycoprotein; chloroquine

1 Introduction

The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease. Patients with the coronavirus pneumonia have 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 new 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 heme 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 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; 0TX. MERS-CoV nsp10_nsp16 complex--5yn5 was used to homology modeling. HEM and 0TX were used to molecular 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.

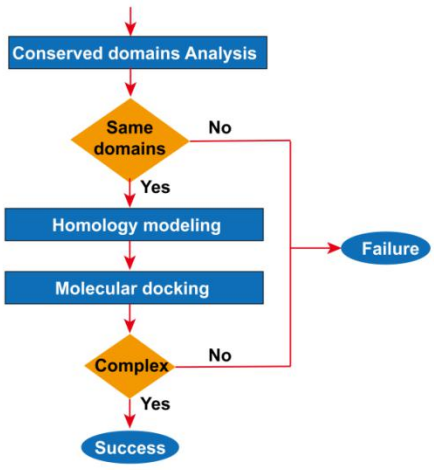

Figure 1. Flow view of Bioinformatics Analysis

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.

3 RESULTS

3.1Virus 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.

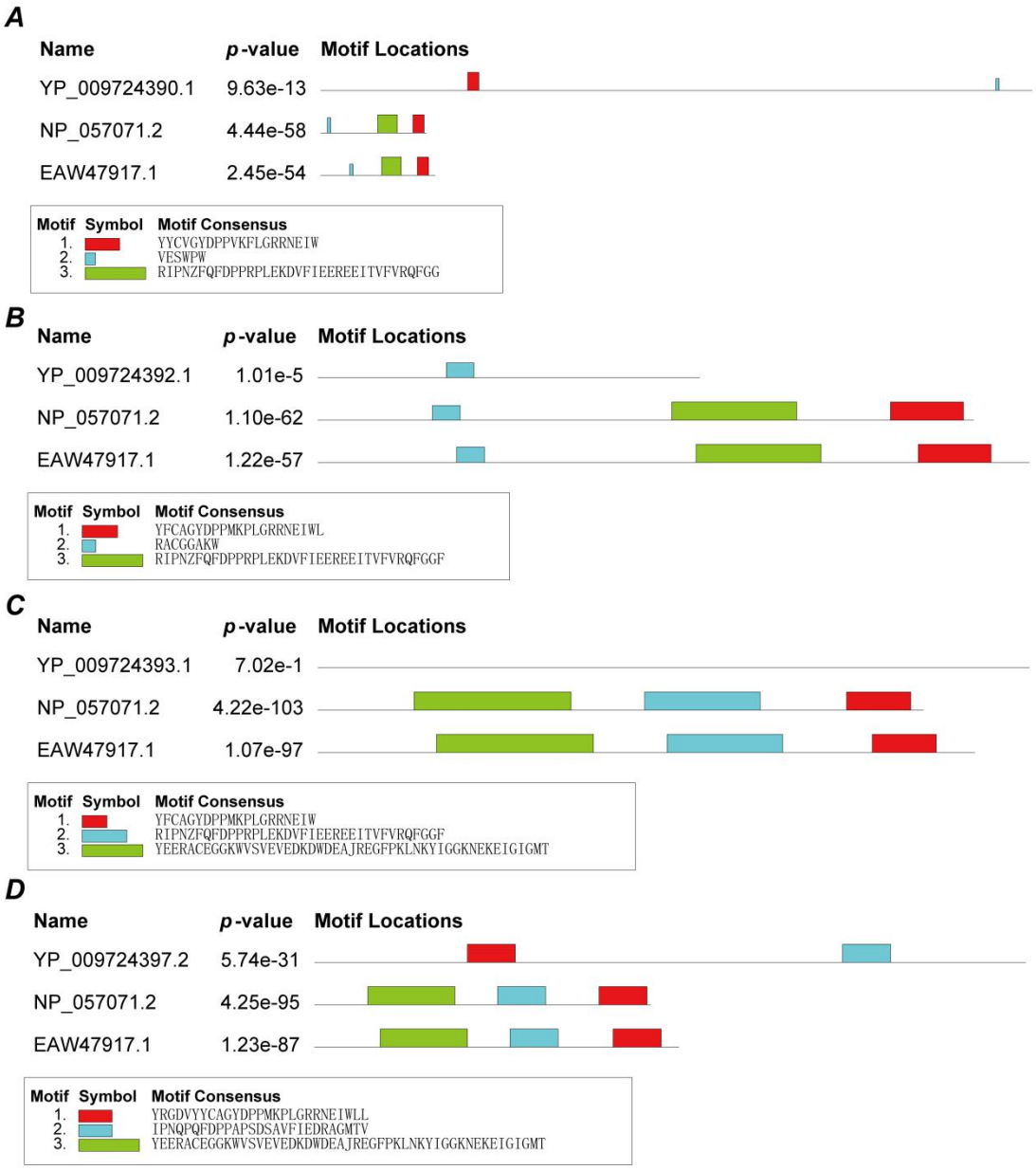

Figure 2. Conserved domains between structure proteins and human heme-binding proteins. ***A***. Conserved domains of surface glycoprotein. ***B***. Conserved domains of envelope protein. ***C***. Conserved domains of membrane glycoprotein. ***D***. Conserved domains of nucleocapsid phosphoprotein.

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 (**7,530,186,265.80**kcal/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 (**219,317.76**kcal/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 (**15,532,506.53**kcal/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.

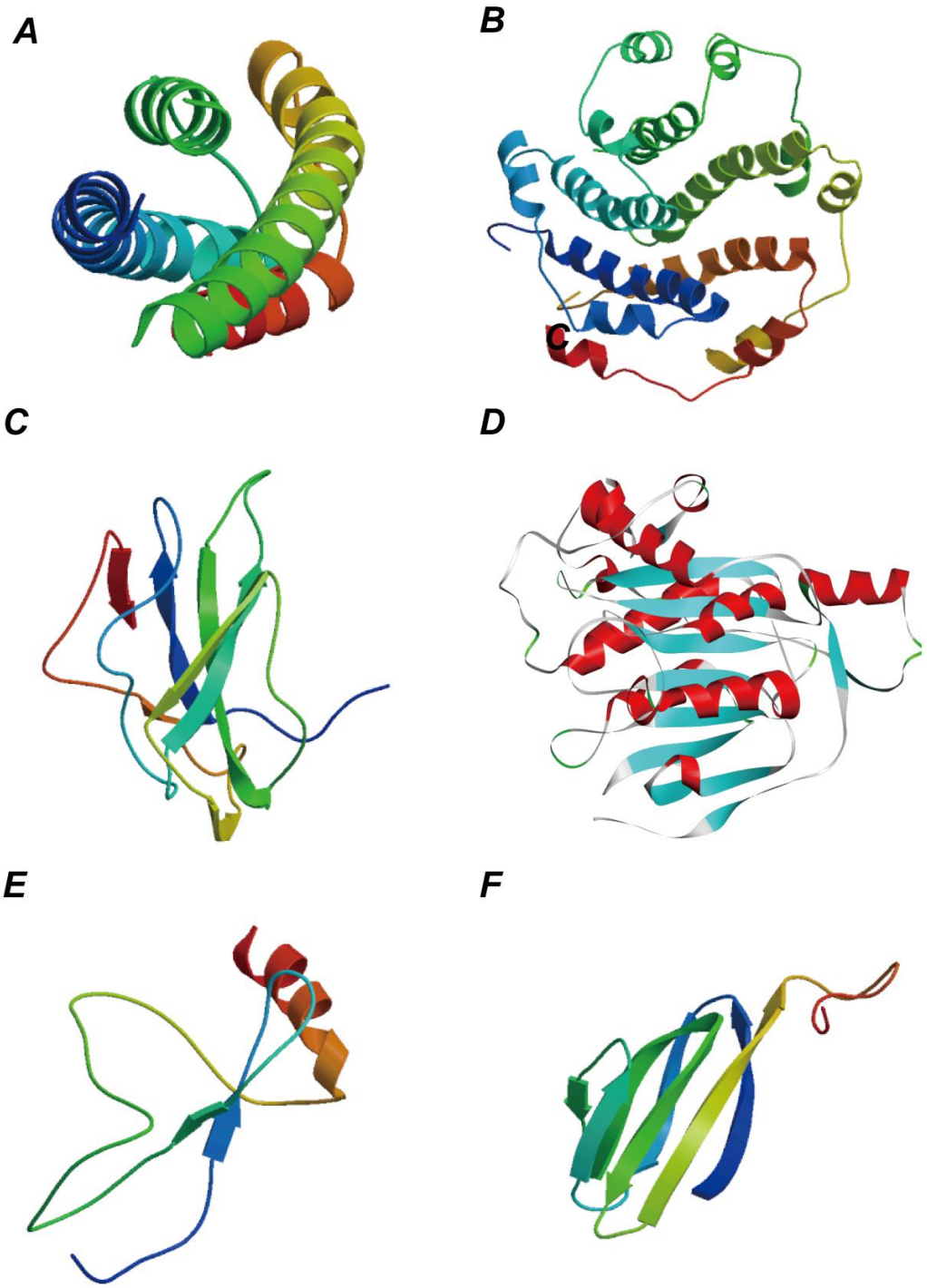

Figure 3.3D structure schematics of the novel coronavirus proteins by the homology modeling. ***A***. E2 glycoprotein of the surface glycoprotein. ***B***. Envelope protein. ***C.***nucleocapsid phosphoprotein. ***D***. orf1ab protein. ***E***. ORF8 protein. ***F***. ORF7a protein

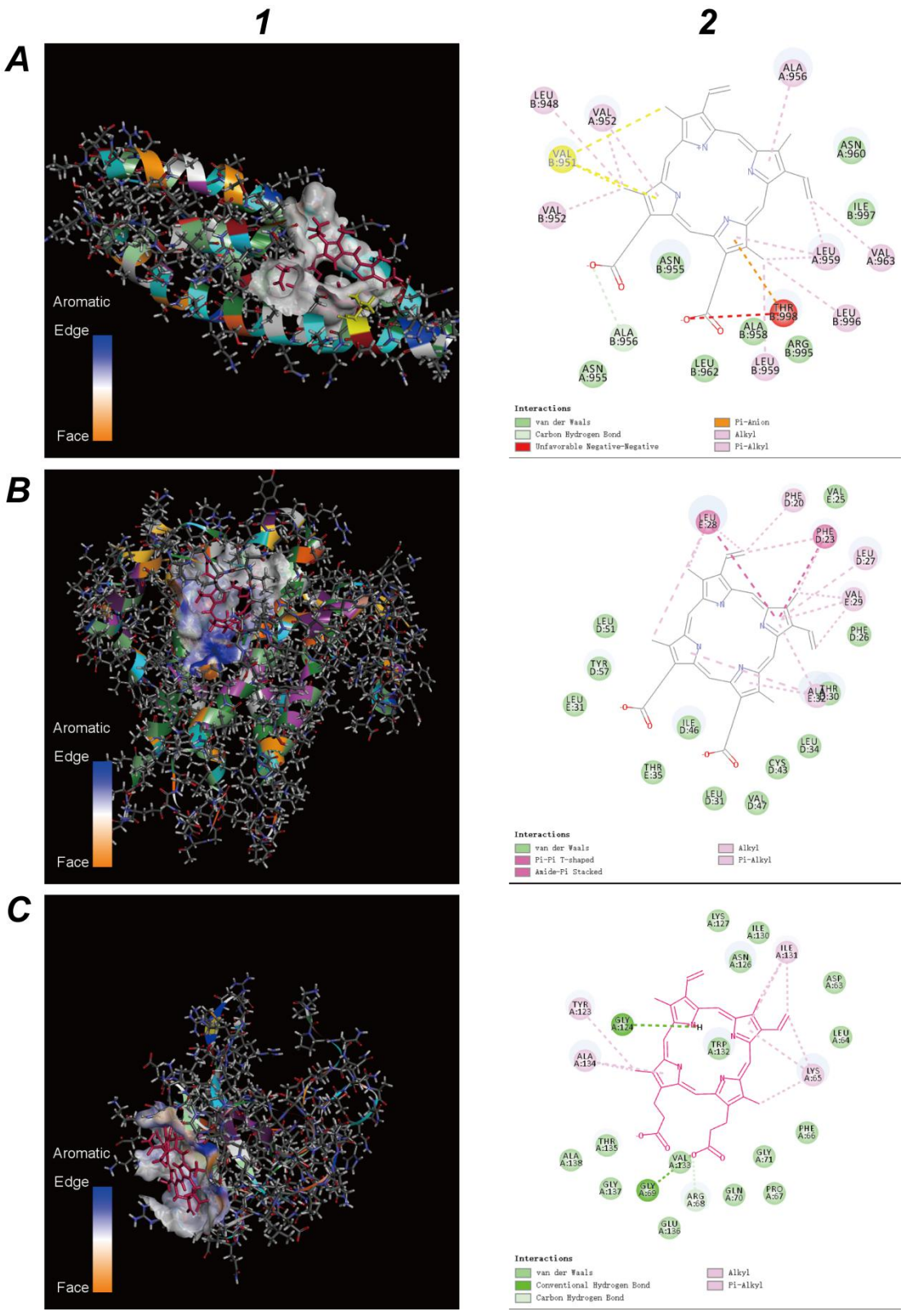

Figure 4. Molecular docking results of viral structure proteins and the porphyrin (red structure). ***A***.Molecular docking results of E2 glycoprotein and the porphyrin. ***B.***Molecular docking results of the envelope protein and the porphyrin. ***C.***Molecular docking results of the nucleocapsid phosphoprotein and the porphyrin. ***1.***Viral structure proteins. ***2.*** View of the binding sections

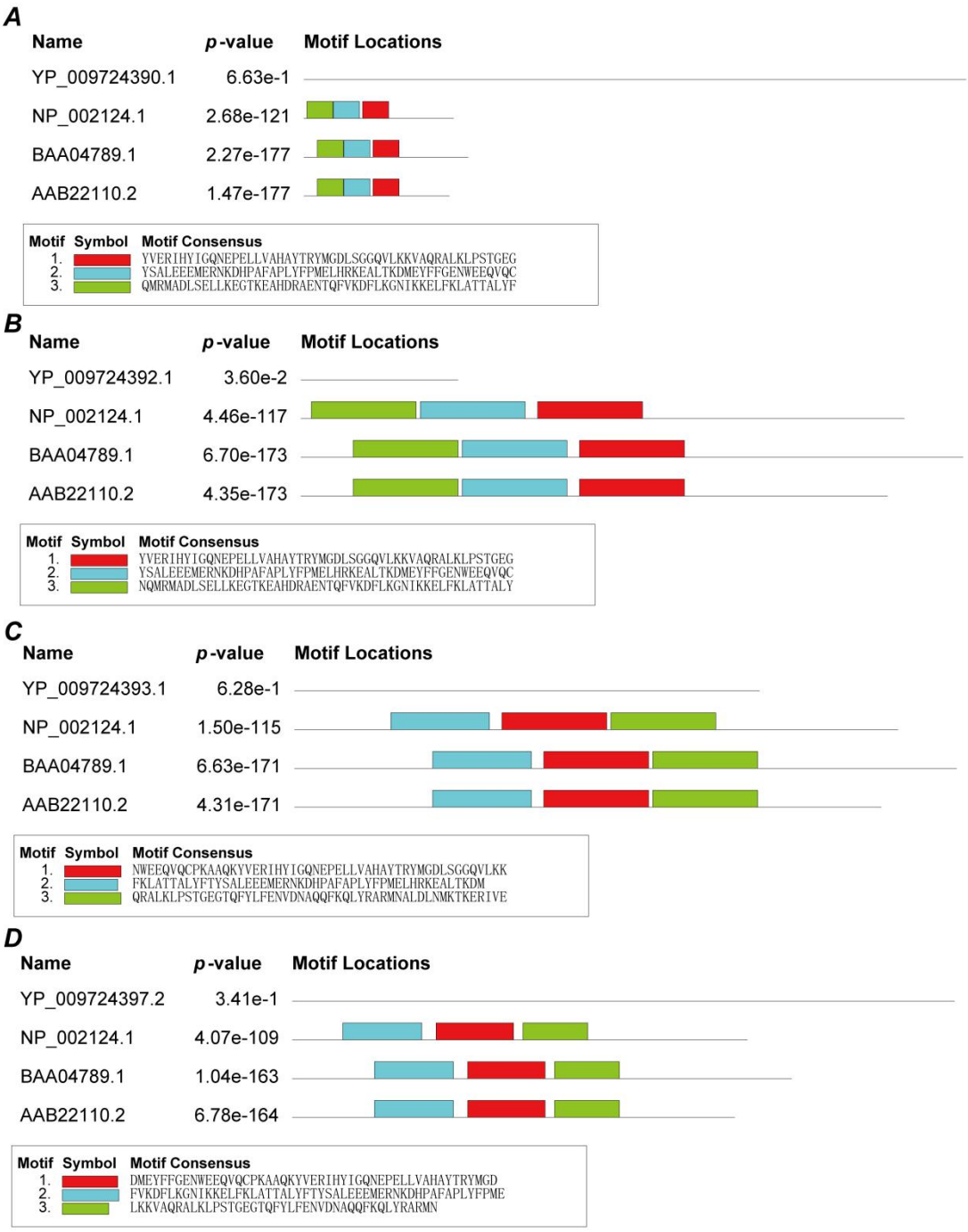

Figure 5. Conserved domains between the structure proteins and human heme oxygenase proteins. ***A***. Conserved domains of surface glycoprotein. ***B***. Conserved domains of envelope protein. ***C***. Conserved domains of membrane. ***D***. Conserved domains of nucleocapsid phosphoprotein.

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 (**561,571.10**kcal/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 (**12,804,859.25**kcal/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.

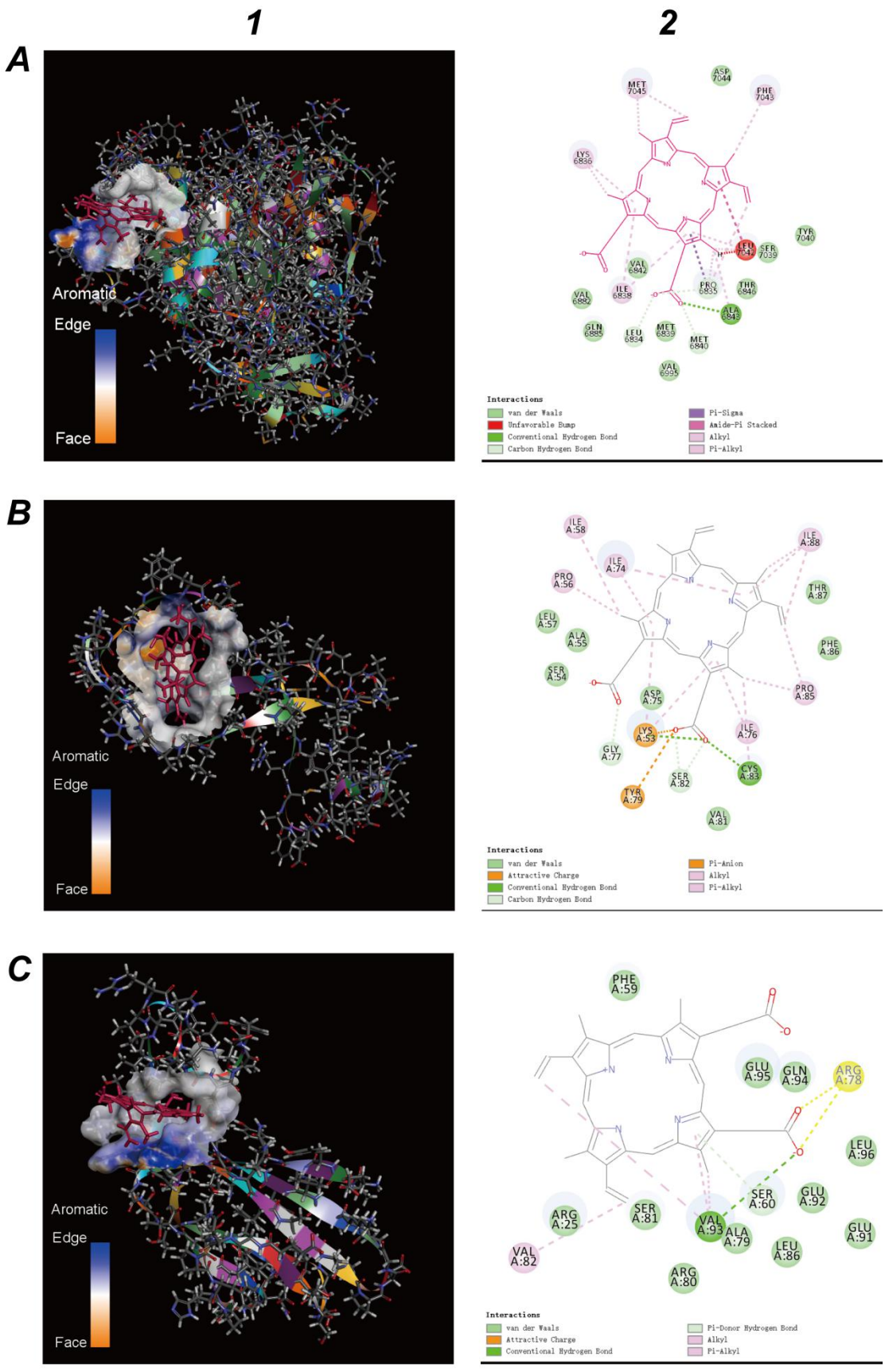

Figure 7. Molecular docking results of viral non-structural proteins and the porphyrin (red). ***A***.Molecular docking results of the orf1ab protein and the porphyrin. ***B.***Molecular docking results of

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.

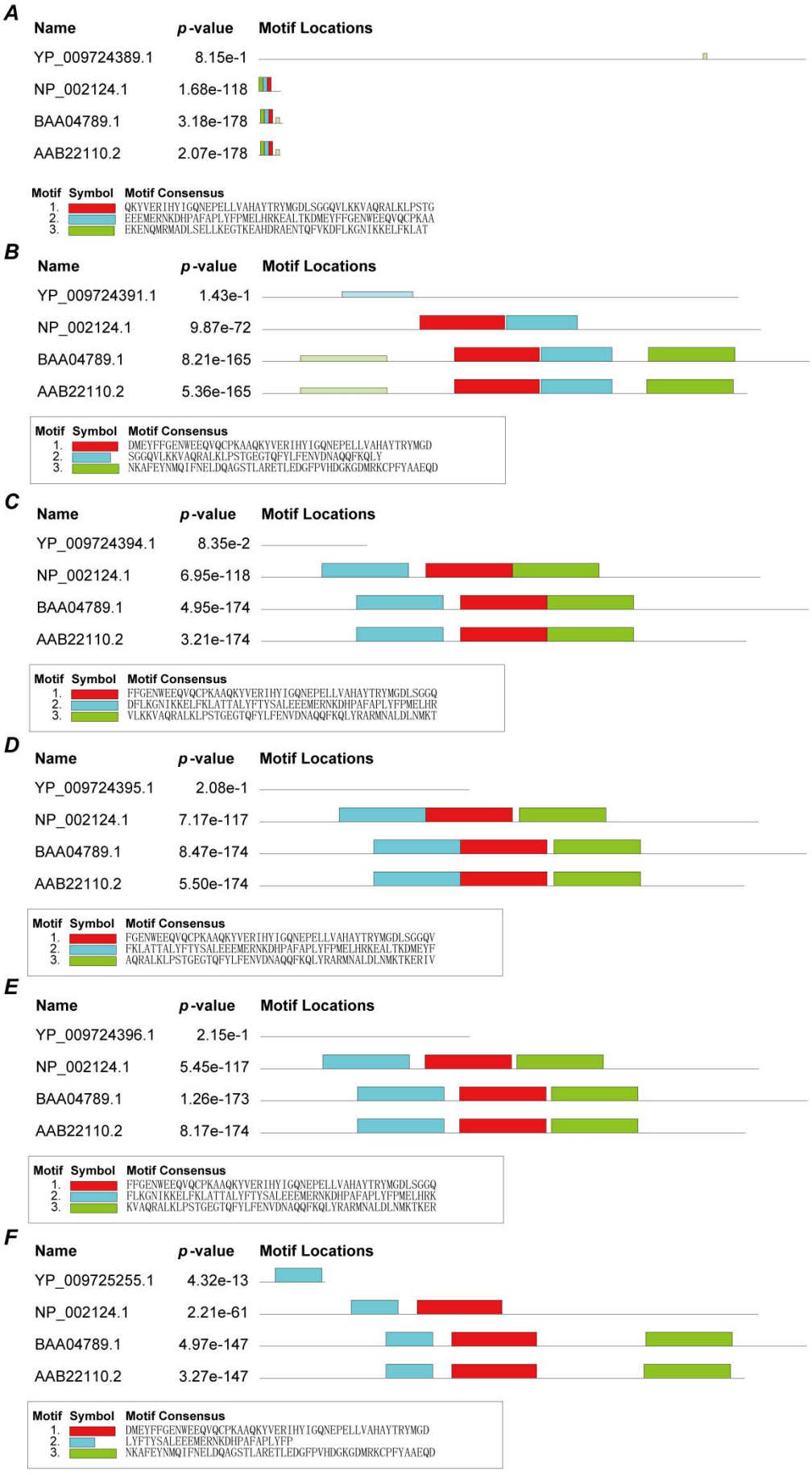
A
Name
p-value
Motif Locations
YP_009724389.1
8.15e-1
NP_002124.1
1.68e-118
BAA04789.1
3.18e-178
AAB22110.2
2.07e-178
Motif Symbol Motif Consensus
1. QKYVERIHYIGQNEPELLVAHAYTRYMGDLSGGQVLKKVAQRALKLPSTG
2. EEEMERNKDHPAFAPLYFPMELHRKEALTKDMEYFFGENWEEQVQCPKAA
3. EKENQMRMADLSELLKEGTKEAHDRAENTQFVKDFLKGNIKKELFKLAT
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 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 9.A-1 is a schematic diagram of binding chloroquine to a virus surface glycoprotein. **Figure 9.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 9.B-1**. The binding energy of the chloroquine and envelope proteinis**7,852.58** kcal/mol, which is only equivalent to 4% of the binding energy of envelope protein and porphyrin. The binding region is shown in **Figure 9.B-2**. **Figure 4.B-2** and **Figure 9.B-2**representedsome 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 9.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 9.C-2**). **Figures 4.C-2** and **Figures 9.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.

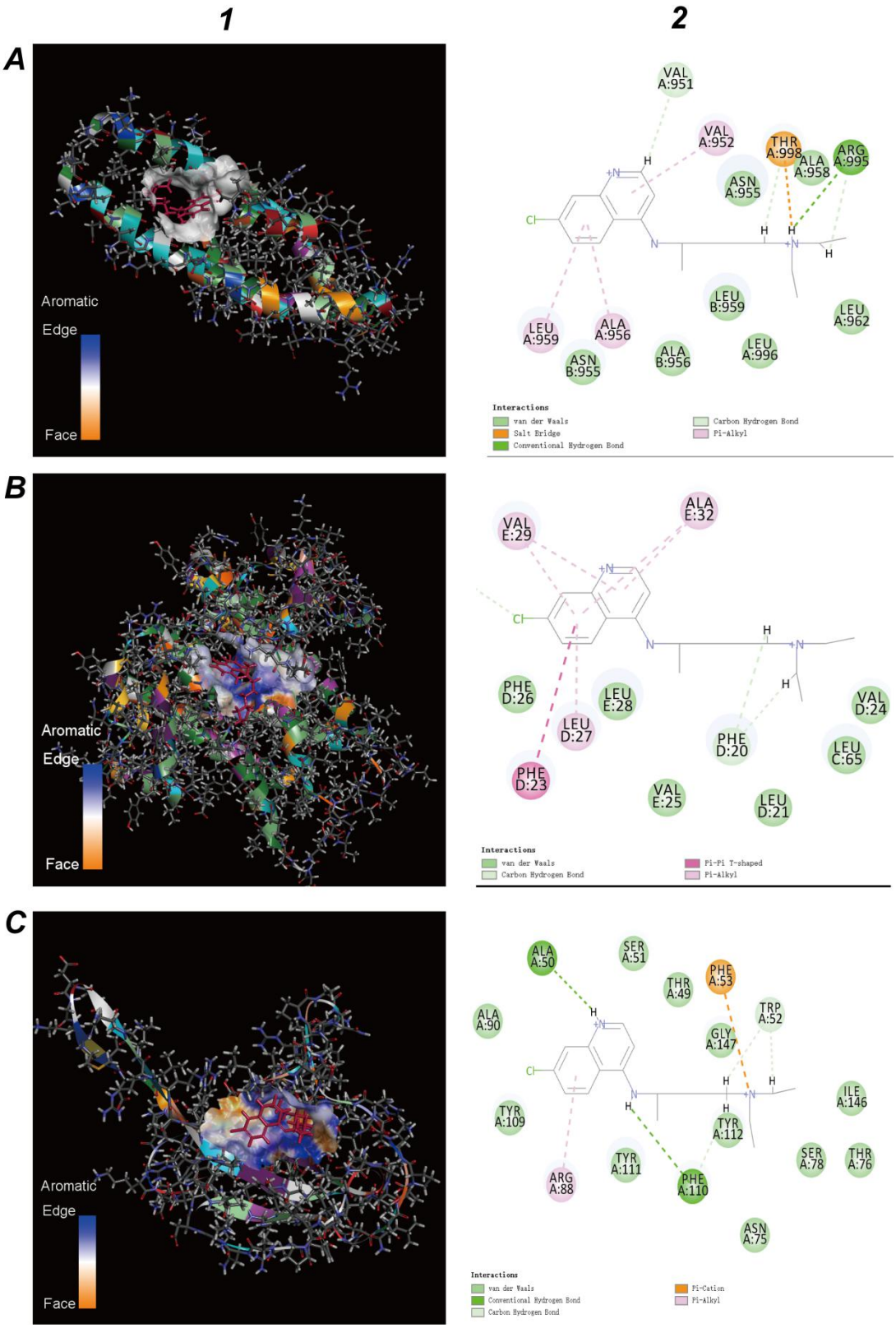

Figure 9. Molecular docking results of viral structure proteins and the chloroquine (red). ***A***.Molecular docking results of the E2 glycoprotein and the porphyrin. ***B.***Molecular docking results of the envelope protein and the porphyrin. ***C.***Molecular docking results of the nucleocapsid phosphoprotein 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 10.A-1**. Binding section of the orf1ab protein is plotted in **Figure 10.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 10.B-1**. **Figure 10.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 10.C**-1. **Figure 10.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.

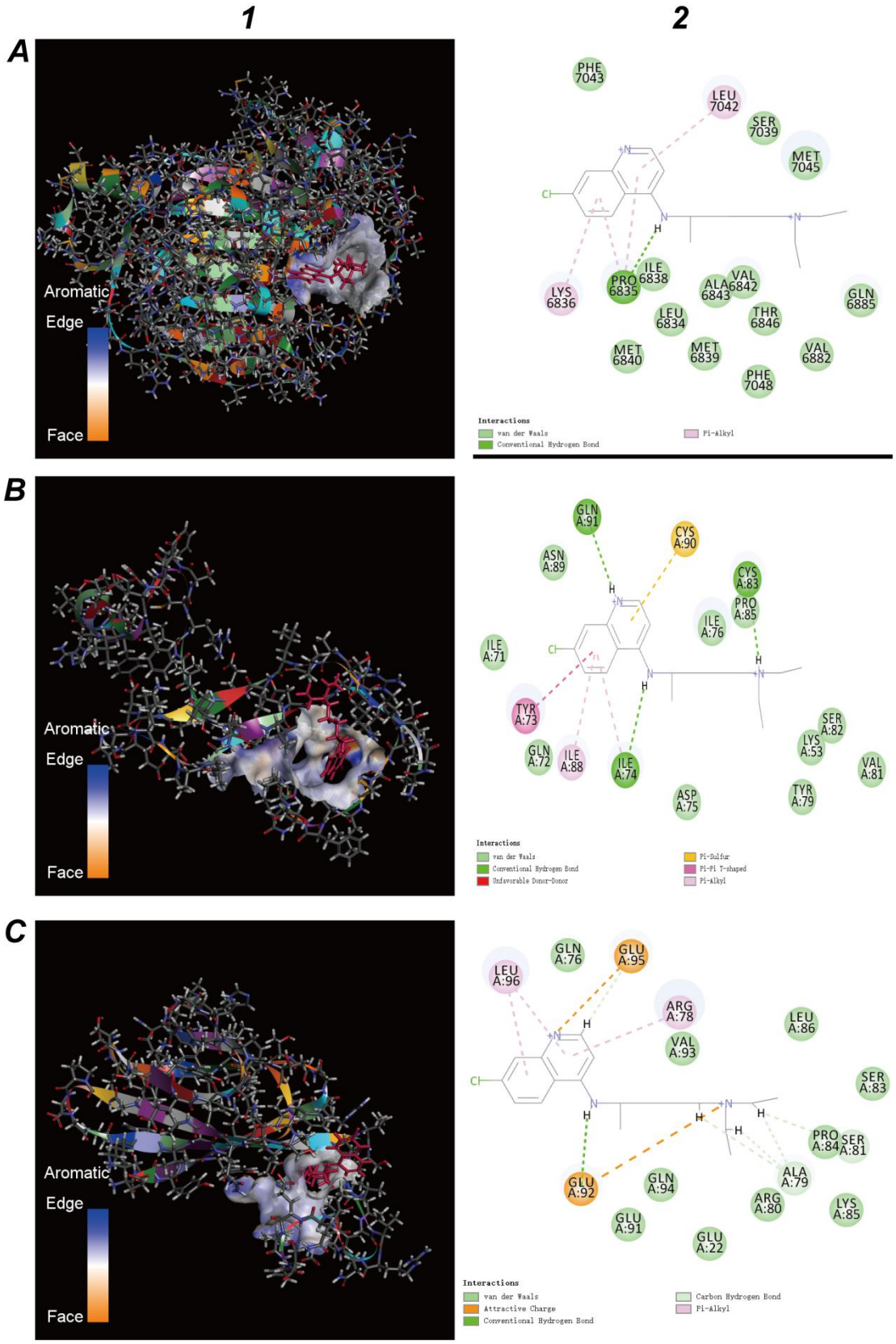

Figure 10. Molecular docking results of viral non-structural proteins and the chloroquine (red structure). ***A***.Molecular docking results of the orf1ab protein and the chloroquine. ***B.***Molecular docking results of the ORF8 protein and the chloroquine. ***C.***Molecular docking results of the ORF7a protein and the chloroquine. ***1.***Viral non-structural proteins. ***2.*** View of the binding sections.

4 Discussion

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. In addition, patients of novel coronavirus pneumonia are most of the middle-aged and elderly people. Many of these patients have underlying diseases such as diabetes. Diabetic patients have higher glycated hemoglobin. Glycated hemoglobin is a combination of hemoglobin and blood glucose, which is another reason of the high infection rate for older people.

Therefore, 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 important biological activities such as regulation of gene expression and protein translation. Protease ALAS1 is a rate-limiting enzyme in heme anabolic metabolism, which is inhibited by the heme. When the heme is too low, the synthesis of protease ALAS1 is inhibited, and the synthesis of heme is blocked again. 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. Complex of virus proteins and the porphyrin may be poorly soluble. Too much mucus in the tissues of the dead patients was the cause of too much mucin protein. Mucin could turn loosely connected cells into tightly adhered cells, and increases lubrication between cells.This suggests the complex leads to poor cell connectivity and cells need mucin to consolidate tissue-cell connectivity and lubricity. In addition, when a patient enters a severe infection period, viral structure proteins were mainly used for virus assembly. Therefore, we cannot find obvious virus inclusions in tissue cells of 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 then 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 combined tothe porphyrin to form a complex respectively, while ORF10, orf1ab and ORF3a attack the heme to dissociate the iron to form the porphyrin. Since the two porphyrin complexes 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 to attack the heme to form the porphyrin, and inhibit the binding of ORF8 and surface glycoproteins to porphyrins to a certain extent. Given the current epidemic, it is believed the results of this study are of high value in preventing the spread of novel coronavirus pneumonia, developing drugs and vaccines, and clinical treatment.

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/1YQNGoN6L9rPU8K5Bnh3EuQ, code: ry25.

Competing interests

The authors declare that they have no competing interests.

Author Contributions:

Design, analysis, writing: Wenzhong Liu. Data curation, check manuscript: Hualan Li. All authors have read and agreed to the published version of the manuscript.

Funding

This work was partially supported by the Natural Science Foundation for Talent Introduction Project of Sichuan University of Science & Engineering (No. 2018RCL20).

Acknowledgements

Not applicable.

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 new 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 new 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).