COVID-19: Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit 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.Wenzhong Liu, liuwz@suse.edu.cn.

Abstract

The novel coronavirus pneumonia (COVID-19) is an infectious acute respiratory caused by the novel coronavirus. The virus is the positive-strand RNA one with high homology to bat coronavirus. The pathogenic mechanism of the new coronavirus is still unclear, which is a significant obstacle to the development of drugs and patients' rescue. In this study, conserved domain analysis, homology modeling, and molecular docking were made to compare the biological roles of specific proteins belonging to the novel coronavirus. The conserved domain analysis showed envelope protein (E), nucleocapsid phosphoprotein (N) and ORF3a had heme linked sites, which Arg134 of ORF3a, Cys44 of E, Ile304 of N were the heme-iron linked site, respectively. ORF3a also possessed the conserved domains of human cytochrome C reductases and bacterial EFeB protein. These three domains were highly overlapping so that ORF3a could dissociate the iron of heme to form porphyrin. Heme linked sites of E protein may be relevant to the high infectivity, and the role of heme linked sites of N protein may be related to the virus replication. The docking results showed that orf1ab, ORF10, and ORF3a proteins coordinated to attack the 1-beta chain of hemoglobin, and some structural and non-structural viral proteins could bind porphyrin. Deoxyhemoglobin was more vulnerable to virus attacks than oxidized hemoglobin. But ORF3a was specific and would not attack blue blood protein, normal cytochrome C, and peroxidase. As for the attack, it would cause increasingly less hemoglobin that could carry oxygen and carbon dioxide, thus producing symptoms of respiratory distress and coagulation reaction, damaging many organs and tissues. The mechanism also interfered with the normal heme anabolic pathway of the human body, expecting to cause human diseases. Based on the small molecule drug library, drugbank, we searched for drugs bound to viral proteins by molecular docking. The results showed that some anticancer drugs could attach to the heme-iron linked site of ORF3a and N. Remdesivir was relatively more obvious than Hydroxychloroquine and Chloroquine in terms of the binding capacity of ORF3a, but the combined role of three drugs to ORF3a was lower. Unfortunately, no drug could bind to the heme-iron linked site of E. Besides, these higher binding energies may prevent all screened drugs from binding firmly to viral proteins. Since there were no clinical data, so inhibitory effects on ORF3a and N were still unclear. This theory is only for academic discussion and needed to be verified by other experiments. Please consult a qualified doctor for treatment details. Due to the toxicity and side effects of drugs, do not use medicines yourself. We expect these discoveries to bring more ideas to people to relieve patients' symptoms and save more lives.

Keywords: Novel Coronavirus; ORF3a; Respiratory distress; Ground-glass-Like Lung; Blood; Diabetic; Coagulation; Fibrosis; Cytokine Storm,

1. Background

The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease. Patients with this disease show a fever over 37.3℃, with symptoms such as dry cough, fatigue and difficulty breathing, and frost-glass-like symptoms in the lungs[1-3]. Some patients also have severe diarrhea[4], such as watery stools. Most mild patients get a better prognosis, but for some serious patients, acute respiratory distress syndrome, shock, acidosis and coagulopathy quickly appear, and they even die. A lot of mucus, fibrin, and thrombosis are discovered in the dissected lung tissue[5,6]. The disease is highly transmitted. Now, the number of infected people has exceeded 10 million 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 that a novel coronavirus caused the disease[7,8]. It is noted that the nucleic acid of the novel coronavirus is a kind of positive-stranded RNA[8]. Its structural proteins include: Spike Protein (S), envelope protein (E), membrane protein (M), and nucleocapsid phosphoprotein. Transcribed non-structural proteins consist of 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 the SARS virus[11,12]. Researchers have studied the function of novel coronavirus structural proteins and some non-structural proteins[13,14]. However, the novel coronavirus has potential genomic characteristics, some of which are the main cause of human outbreaks[15,16]. For example, the CoV EIC (Coronavirus envelope protein ion channel) has 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 bring about more serious pathogenicity and transmission differences when compared with SARS-CoV[18]. From earlier studies, it was found that the novel coronavirus enters epithelial cells through the spike protein interacting with the human ACE2 receptor protein on the surface, thus causing human infection. Due to the limitations of existing experimental methods, the specific functions of virtual proteins such as ORF8 and ORF3a are still unclear. The pathogenicity mechanism of the novel coronavirus remains mysterious[19].

Literature[20] revealed biochemical examination indexes of 99 patients with the novel coronavirus pneumonia, and also reflected the abnormal phenomenon of hemoglobin-related biochemical indexes of patients. Besides, it was demonstrated that the hemoglobin and neutrophil counts of most patients decreased, and the index values of serum ferritin, erythrocyte sedimentation rate, C-reactive protein, albumin, and lactate dehydrogenase of many patients increased significantly. This trace implied that patients' hemoglobin was decreasing, and the body would accumulate too many harmful iron ions, which would form inflammation in the body and increase C-reactive protein and albumin. Cells react to stress due to inflammation, thus producing large amounts of serum ferritin to bind free iron ions to reduce damage. Most carbon dioxide is transported in the form of bicarbonate in plasma, while a small portion of carbon dioxide is done by carbamoyl hemoglobin of erythrocytes. Oxygen is transported in the blood by hemoglobin of red blood cells. Hemoglobin is composed of four subunits, 2-α and 2-β, and each subunit has an iron-bound heme[21,22]. The heme is an essential component of hemoglobin. It is a porphyrin containing iron. The structure without iron is called porphyrins. When iron is divalent, hemoglobin can release carbon dioxide and capture oxygen atoms in alveolar cells, and iron is oxidized to the trivalent state. 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 the divalent state. Therefore, here it is thought that viral proteins may attack hemoglobin, causing the heme to dissociate

into iron and porphyrins, and then the viral proteins capture the porphyrin. This interpretation was also consistent with some existing clinical features.

Both lungs hold the role of exchanging carbon dioxide and oxygen. The COVID-19 virus that attacked hemoglobin would yield iron, carbon dioxide, and oxygen, which might put both lung cells in a toxic and inflammatory state. Then, it will form multiple ground glass images and penetration shadows on both sides of the lung[23-25], when ground glass images are often associated with rapid and noticeable hypoxemia.

In the early and severe stage, patients have varying degrees of respiratory distress symptoms. Many doctors have found that ECMO patients have some strange clinical features that are low oxygen, low blood oxygen saturation[26,27] and high dissolved oxygen, which may be a critical reason in the low success rate of rescuing critical patients[28]. ECMO is featured with cardiopulmonary function in vitro, and can help patients exchange oxygen and carbon dioxide in vitro. During this process, drugs such as the anticoagulant can be added at the same time. In this article, it is believed that this phenomenon indicates hemoglobin, which can carry oxygen in the blood of critically ill patients, became extremely low. Due to the deep attack of the virus, there is little normal functioning hemoglobin that can carry oxygen in the blood. Therefore, patients with ECMO have low oxygenation. If the virus can bind a large number of porphyrins, heme synthesis is inhibited. In this case, there is less hemoglobin synthesis. Besides, the virus attacks hemoglobin, which reduces the amount of hemoglobin that is typically oxygenated. Much hemoglobin can not carry oxygen after the attack. As a result, blood oxygen saturation decreases. Oxygenated hemoglobin is red blood cell hemoglobin that binds oxygen. After the virus attacks the oxidized hemoglobin, the lost oxygen atoms accumulate in the blood. As they hardly enter into the tissue cells, the dissolved oxygen in the blood gets higher.

Virus attacks will cause damage to many organs and tissues. For example, capillaries are easily broken due to inflammation[29]. Proteins such as fibrinogen fill the capillaries' cracks through the coagulation reaction. Therefore, a lot of fibrin and thrombi accumulate in the lung tissue of critically ill patients[30]. If the amount of hemoglobin being attacked is much larger than that of the synthetic hemoglobin, the patient will suffer a certain degree of anemia. Different degrees of systemic coagulation occurred as anemia[31-34], vascular damage, and the body's immune response. Vascular injury may be the principal factor for coagulation in COVID-19 patients who then are under a higher D-Dimer content[35,36]. Studies have showed that the COVID-19 virus also infects T-cells[37], which may cause the abnormal function of T-cells[38]. Severe COVID-19 patients often present with mononucleosis[39,40]. Macrophages can engulf free iron, damaged red blood cells and hemoglobin. Of course, macrophages also handle viral inclusions[41], thrombin, fibrin, and foreign matter generated by inflammatory tissues. Thrombin and fibrin may be produced by the coagulation reaction.

If the virus could attack hemoglobin in red blood cells, some prerequisites are necessary here, such as virus infection of red blood cells or hemolysis of red blood cells. There are now some confusing signs. For example, EMMONS, et al. observed that the Colorado tick fever virus existed in red blood cells through electron microscopy[42], which indicates that red blood cells have a risk of infecting some viruses. Mitra, Anupam, et al. found that virus Leukoerythroblastic reaction appeared in a single patient with COVID-19 infection[43]. Bhardwaj, et al. pointed out that pRb and its interaction with Nsp15 affect coronavirus infection[44]. Recently, it has been reported that the COVID-19 virus infects cells through the Spike-CD147 pathway. CD147 plays an essential role in some systems such as the Ok blood group[45], and T lymphocytes[37]. Statistical analysis shows that the susceptibility of COVID-19 has a certain relationship with blood types. The infection rate of blood type A is high, while

that of blood type O is minimal. The blood type is an important antigen on the erythrocyte membrane. The latest report shows that the number of red blood cells in rhesus monkeys infected with the COVID-19 virus is decreasing. However, the mechanism of red blood cell infection by the novel coronavirus remains unclear.

Plasmodium, babesia, and trypanosomes can also infect red blood cells, causing similar symptoms in patients. Trypanosoma extracellular vesicles can also fuse with mammalian red blood cells[46], resulting in the red blood cells to be readily cleared and further leading to anemia. Thrombin sensitization-related adhesion protein and apical membrane antigens are important for babesia invasion of red blood cells[47]. There are numerous types of plasmodium, which can infect red blood cells. Plasmodium usually infects red blood cells in three steps. Firstly, it sticks to red blood cells through proteins such as AMA[48]. Then, the plasmodium ligand and erythrocyte receptor complete a tight connection. For example, the principal receptor for P. falciparum to invade red blood cells is CD147 protein[49]. Finally, dynein, MTRAP (homologous to thrombin-sensitive protein), and other kinetic system proteins finish the process of rapid slide of plasmodium into red blood cells. If the COVID-19 virus also uses thrombin-sensitized protein-mediated sliding into red blood cells, it can account for this phenomenon of systemic coagulation in severe patients to a certain extent. It is useful to that, and there may be a coincidence. Patients infected with plasmodium and babesia are treated with chloroquine and quinine, and those infected with trypanosomiasis can be healed with anisace (quinoline pyrimidyl sulfate). Chloroquine is a drug with serious side effects, and some novel coronavirus pneumonia has been cured by chloroquine. Meanwhile, one detail we can notice is that chloroquine is also a commonly used drug for treating porphyria[51,52].

The main symptoms of porphyria appear in the skin and neurovisceral organs. Existing reports show pigmentation in critically ill patients, which however may be a side effect of drugs such as polymyxin B. Some patients with severe skin diseases may also have fever symptoms, which are easily confused with the symptoms of COVID-19 patients[53]. Researchers have diagnosed COVID-19 infection with erythema rash, extensive urticaria and chickenpox-like vesicles[54]. Some patients with the novel coronavirus pneumonia have similar neurovisceral symptoms. Of course, in the later stage of severe inflammatory infection, the cytokine storm may also induce the failure of many organs. There are symptoms such as diarrhea[55,56], hypotension[57], and electrolyte disturbances[58]. Organs such as hearts[59], livers[60], and kidneys[61] are damaged and develop complications. In terms of neurological symptoms, some early stealth infections experienced the loss of taste and smell[62,63]. Patients with the severe novel coronavirus pneumonia also suffer from rare neurological diseases such as epilepsy[64,65] and encephalitis[66,67]. Therefore, we believed that combining viral proteins and porphyrins interfered with the normal heme anabolic pathway of the human body, causing a series of human pathological reactions.

In short, COVID-19 viral protein may conduct a series of activities by binding porphyrins. The porphyrin in the human body mainly is stored on hemoglobin in the form of the heme. As the virus requires too many porphyrins, it has evolved the function of attacking the heme on hemoglobin and dissociating iron to form porphyrins. 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 with bioinformatics methods. In this study, conserved domain prediction, homology modeling, and molecular docking techniques were used to analyze the functions of virus-related proteins. The conserved domain analysis showed envelope protein (E), nucleocapsid phosphoprotein (N) and ORF3a had heme linked sites. ORF3a also possessed

the conserved domains of human cytochrome C reductases and bacterial EFeB protein, so that ORF3a could dissociate the iron of heme to form porphyrin. This study revealed that some proteins had a function of combining with porphyrins to form a complex, while orf1ab, ORF10 and ORF3a coordinately attacked 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 the heme and made people show symptoms of the disease. Based on the small molecule drug library, drugbank, we searched for drugs bound to viral proteins by molecular docking. The results showed that some anticancer drugs could attach to the heme-iron linked site of ORF3a and N, but no drug could bind to the heme-iron linked site of E. Since there were no clinical data, so their inhibitory effects were still unclear.

2. Methods

2.1 Data set

The protein sequences were downloaded from NCBI: all proteins of the novel coronavirus, heme-binding protein, heme oxidases, cytochrome C oxidase, cytochrome D oxidase, Dechelatase and ORF3/ORF3a proteins. Besides, the sequences were utilized to analyze conserved domains, and some proteins of the 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: HEM; Human Oxy-Hemoglobin 6bb5; DEOXY HUMAN HEMOGLOBIN 1a3n. Some metal porphyrin proteins such as human cytochrome, human erythrocyte catalase, Octopus Hemocyanin. Some SARS-COV-2 proteins such as nsp16-nsp10 6w4h, main protease 6y2e, main proteases in apo form 6m03, Nsp9 RNA binding protein 6w4b, Nsp9 RNA-replicase 6w9q, spike closed state 6vxx, spike open state 6vyb. 6bb5 and 1a3n are employed for protein docking, while HEM is used for molecular docking. SARS-COV-2 proteins are utilized for molecular docking or protein docking.

2.2 Flow view of bioinformatics analysis

A series of bioinformatics analysis were performed based on published biological protein sequences in this study. The steps are illustrated: 1. Conserved domains of viral proteins are analyzed by MEME[68-70] Online Server, and were used to predict function differences of viral proteins. The viral proteins could not bind and oxidize the heme, and then were exposed to the conserved domain of heme linked sites. 2.The three-dimensional structure of viral proteins was constructed by homology modeling of SWISS-MODEL[71,72], AlphaFold and Robetta. 3. Docking technology. Utilizing protein docking technology (ZDock tool) of Discovery-Studio 2016[73], the docking of viral proteins with metal porphyrin proteins was performed. Adopting molecular docking technology (LibDock tool) of Discovery-Studio 2016, the receptor-ligand docking of viral proteins with the human heme (or porphyrins) was simulated. Depending on the results of bioinformatics analysis, the related molecular of the disease was proposed.

2.3 Analysis of conserved domains

The MEME Suite is an online website that integrates many tools of prediction and annotation motifs. The maximum expectation (EM) algorithm is the basis for MEM's identification of the motif that 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.

In this study, the MEME was adopted to determine that the viral protein does not have conserved domains of human heme-binding protein and human heme oxidation protein. It is shown that the viral protein cannot bind and oxidize the heme. Then, the MEME was used to find the conserved domains of viral proteins by comparing to some metal porphyrin proteins of the bacteria. It is represented that some viral proteins had the heme-linked sites.

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 SWISS-MODEL, enter the sequence, and click the "Search Template" to perform a simple template search. After the search is completed, a template for modeling can be chosen. A template search will be conducted by clicking the "Build Model", and a template model is selected automatically. As can be seen, several templates were searched, and then numerous models were built. Just a model is chosen here. The model in the PDB format is downloaded and visualized in VMD. SWISS-MODEL is employed to model S, ORF6, proteins

The quality of ROBETTA modeling is excellent; there are many reference templates, and it can also model some viral proteins with low similarity. After the user submits the sequence online, the server gives a modeling queue. After successful modeling, the server will send the 3D structure file to the submitter's mailbox. This online server also has restrictions on the sequence length. In this study, we modeled the E, N, ORF10,ORF7a,ORF6, ORF8 sequence with ROBETTA.

AlphaFold's algorithm significantly outperforms traditional models in the field of protein folding. The AlphaFold project team has constructed 3D files of some viral proteins. The modeling files of ORF3a and M protein were selected for this study.

2.5 Molecular docking technology

The purpose of molecular docking is to show that some viral proteins can bind to porphyrins. 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. Prepare 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. Firstly, delete FE (iron atom) in HEM, and then click the “Prepare Ligands” button; then the porphyrin ligand model will be generated.

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 “DockLigands” submenu. In the pop-up box, select the ligand and the receptor as the newly established receptor model-ALL, and the sites sphere as the sphere coordinates were just established. Finally, click RUN to start docking.

4. Calculate the binding energy and choose the pose with the lowest binding energy. After docking is complete, many locations of the ligand will be displayed. Open the docked view, and click the “Calculate 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, choose the 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 lowest binding energy[74-76].

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, and then the view of the binding section, which can be saved as a picture file pops up.

2.6 Protein docking technology

The main purpose of protein docking technology is to study the attack of viral proteins on metal porphyrin proteins such as hemoglobin. Discovery-Studio's ZDOCK is another molecular docking tool for studying protein interactions. Here, it was used to study the attack of hemoglobin by viral non-structural proteins. The following is the docking of nsp16-nsp10 (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 Complexes" under the "macromolecules" menu. In the pop-up interface, select Human Oxy-Hemoglobin 6bb5 as the receptor, and orf1a as the ligand, and then click the "Run" button. After the computer finishes computing, click on the "protein pose" interface and select the pose and cluster with the highest ZDOCK score. It could obtain the position of orf1ab on Human Oxy-Hemoglobin 6bb5, and Deloxy HUMAN HEMOGLOBIN 1a3n has a similar docking pattern with nsp16-nsp10 protein.

3. RESULTS

3.1 Viral proteins had no conserved domains of human heme-binding and oxidase protein

In humans, hemoglobin can be degraded into globin and the heme, while the latter is composed of porphyrins and an iron ion in the middle of the porphyrin. The heme is insoluble in water, and can be combined with heme-binding proteins to form a complex and be transported to the liver. Besides, it is degraded into bilirubin and excreted from the bile duct, and the body can reuse iron in the molecule. The following bioinformatics methods were adopted to study virus proteins binding to porphyrins.

To understand whether the protein of the virus could bind the heme, a similar domain of heme-binding protein was analyzed. First, MEME's online server was employed to search for conserved domains in each viral protein and human heme-binding protein (ID: NP_057071.2 heme-binding protein 1, ID: EAW47917.1 heme-binding protein 2). Table 1 presents that viral proteins and heme-binding proteins has not any conserved domains; E-values are higher than 0.05, and they are not statistically significant. These results show that the viral proteins are unable to bind the heme. Whether the viral proteins interact with porphyrins is determined by molecular docking.

After that, the following analysis was conducted to find out whether the protein could oxidize the heme. 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). The viral proteins possess domains, but E-values are higher than 0.05 and are not statistically significant. As

a result, conserved domains of proteins are not found. Combining the results of the previous analysis, or viral proteins may not oxidize the heme. It is worth noting that the E value of ORF3a, ORF10, orf1ab, and NP_002124.1 is the smallest among nonstructural proteins, and the E value of ORF3a and ORF10 is less than zero. Whether these three proteins affect the heme of Hemoglobin is determined by molecular docking too.

Table 1. E-value of domains between viral protein and human heme-binding protein

Viral protein	NP_057071.2	EAW47917.1
surface glycoprotein	1.60E-01	3.40E+00
envelope protein	7.00E-01	3.70E+00
membrane glycoprotein	1.10E+01	9.80E+00
nucleocapsid phosphoprotein	2.90E+00	5.20E+00
orf1ab	2.00E+00	2.40E-01
ORF3a	6.00E+00	1.10E+00
ORF6	3.30E+01	1.60E+00
ORF7a	7.20E+00	2.50E+00
ORF8	4.70E+00	1.20E+01
ORF10	1.10E+00	4.70E-01

Table 2. E-value of domains between viral protein and heme oxidase proteins

Viral protein	NP_002124.1	BAA04789
surface glycoprotein	5.4E+000	3.7E-001
envelope protein	3.2E-001	4.4E+000
membrane glycoprotein	6.1E-001	1.3E+000
nucleocapsid phosphoprotein	1.0E+000	1.3E+001
orf1ab	1.2E+001	1.4E+001
ORF3a	1.3E-001	7.0E+000
ORF6	1.4E+000	9.9E+000
ORF7a	2.4E+001	7.1E+000
ORF8	5.9E+000	4.2E+000
ORF10	6.6E-001	2.3E+000

3.2 ORF3a protein catalyzes the dissociation of iron from heme

The previous conclusions show that the viral proteins have no conserved domains of human heme-binding protein and heme oxidation protein, maybe because the mechanism of human heme binding and oxidation is much more complicated, and viral proteins may not have evolved the similar functions of human heme binding and oxidation proteins. We may have to refer to the heme-binding mechanism of lower organisms.

It was noticed that there was a phenomenon, in which some bacteria degrade the heme to form porphyrin rings and iron. These bacteria hunt for iron for survival. There are two primary forms of bacteria degrading the heme. To be specific, one is heme oxidase-mediated, while the other is a heme-degrading enzyme. Both methods have one thing in common, starting with trivalent iron and the heme, and dissociating iron through complex electron transfer, oxidation, reduction, and cracking processes. In the heme oxidase-mediated manner, the process of electron transfer is achieved through cytochrome reductases. If the viral protein dissociates the iron of the heme, it may have a partially conserved domain of cytochrome oxidoreductase. There are tens of thousands of bacteria, and it is impossible to compare them one by one. In this case, only representative bacteria are selected.

If viral proteins attack hemoglobin and the heme, iron, carbon dioxide, and oxygen are produced. The human body may be a hostile environment lacking oxygen, acidity, iron, and oxygen-free radicals. Viral proteins may have similar capabilities to the proteins of some extreme environmental bacteria. The members of Aquificales are the bacterial communities known to have higher growth temperatures so far, and they are widely distributed around the world, mainly growing in volcanic or geothermally heated environments. Here, the Hydrogenobaculum was selected as a reference, and it plays an essential role in the biogeochemical cycle of these hot springs. Hydrogenobaculum (strain Y04AAS1) was a thermophilic acidophilic bacterium, and was isolated from the Earth Hot Springs, such as the obsidian pool of Yellowstone National Park. It grows optimally at 58℃ and pH4, and gains energy through the oxidation of hydrogen (the "knallgas" reaction) or reduced sulfur compounds.

Exposure to the conserved domain of heme linked sites. We downloaded the sequence cytochrome C oxidases (gi | 452882433 | gb | AGG15137 |, AGG15137.1, cytochrome C oxidase monoheme subunit / FixO [Hydrogenobaculum sp. HO]) and its' related structure 3MK7B from NCBI. Then, in this study, the MEME tool was used to compare cytochrome C oxidases with viral proteins to find conserved domains. After that, it was found that ORF3a and cytochrome C oxidases have the conserved domain "IMRAWGCWKCR" (**Figure 1.A**), E value = 2.0E-003<0.05, which has a certain statistical significance. The sequence position of the conserved domain in AGG15137.1 is 111-121 (**Figure 1.B**), which belongs to the front of the cytochrome oxidase domain pfam02433 listed by NCBI. We displayed 3MK7B in the Cn3D tool and looked at the structure area at positions 111-121 (**Figure 2.**). The yellow area in **Figure 2.A** is the functional area of the conserved domain "IMRAWGCWKCR", which connects the heme iron, the carbon at the left end of the alpha position, and the beta position. (**Figure 2.A**).

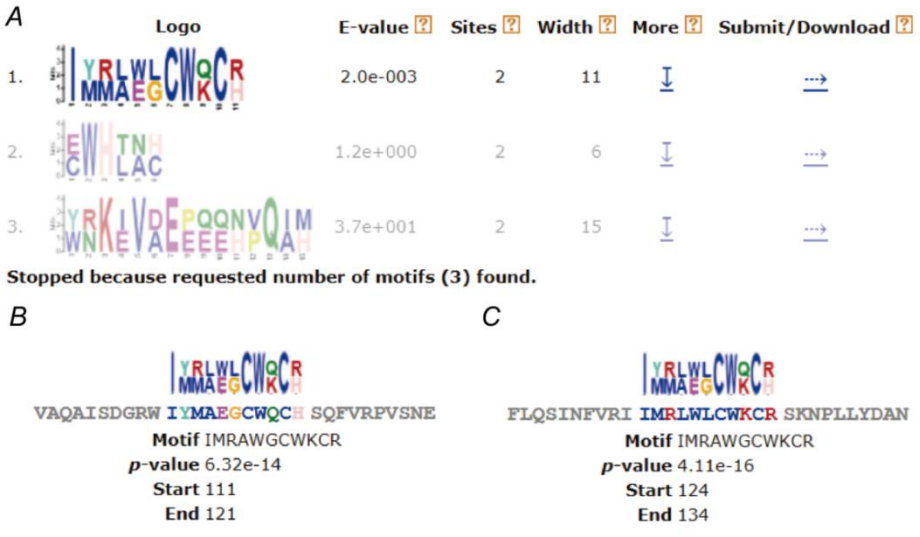

Figure 1. Heme-linked domain of ORF3a**.** ***A***. Heme-linked domain "IMRAWGCWKCR". *B*. heme linked sites at positions 111-121 of AGG15137.1. *C*. heme linked sites at positions 124-134 of ORF3a.

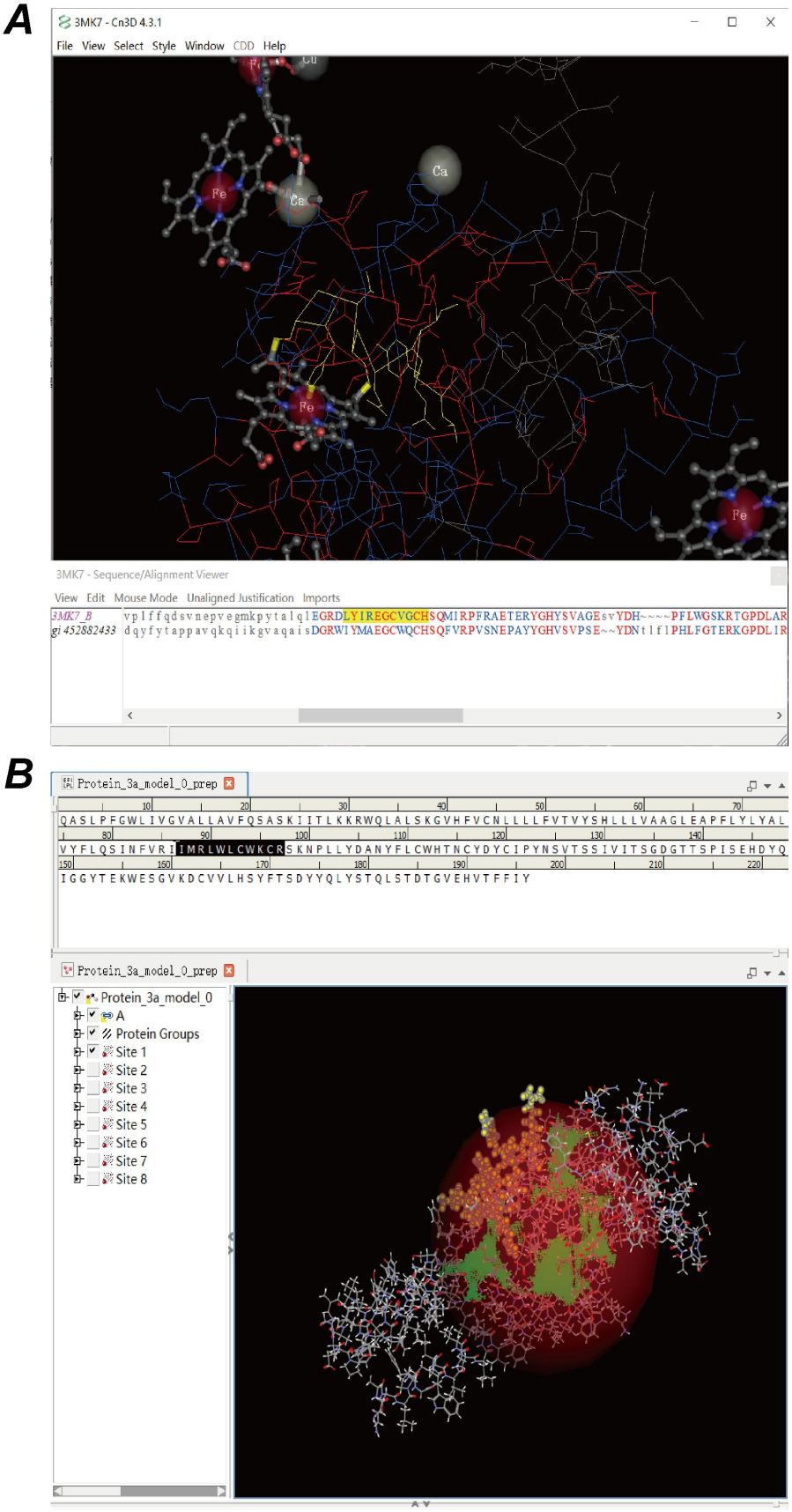

Figure 2. 3D position of heme motif. ***A***. 3D position of heme motif in 3MK7B. ***B***. 3D position of heme motif in ORF3a.

In other words, the last Arg of the conserved domain corresponds to the His of 3MK7B . His nitrogen atom connected to iron (**Figure 2.A**), and can cooperate with metal ions such as iron. For example, the link of the heme to myoglobin or hemoglobin was formed by histidine and iron as the coordination key. The last two Cys connected the second carbon at the left end of the alpha position and the beta position. The first Cys connected the second carbon at the end of the alpha position, while the second Cys connected the second carbon at the end of the beta position. The sequence position of the conserved domain "IMRAWGCWKCR" in ORF3a is 124-134 (**Figure 1.C**), which belongs to the active area of ORF3a (**Figure 2.B**). The result indicated that Arg134 of ORF3a linked to the iron of the heme, and Cys130 and Cys133 linked to the carbon at the left end of the alpha position and the beta position respectively(**Figure 3)**. It is interesting that cytochrome C oxidases and ORF3a linked the heme at alpha and beta areas with Cys, indicating that this position may be highly conserved. ORF3a's iron linker evolved from His to Arg.

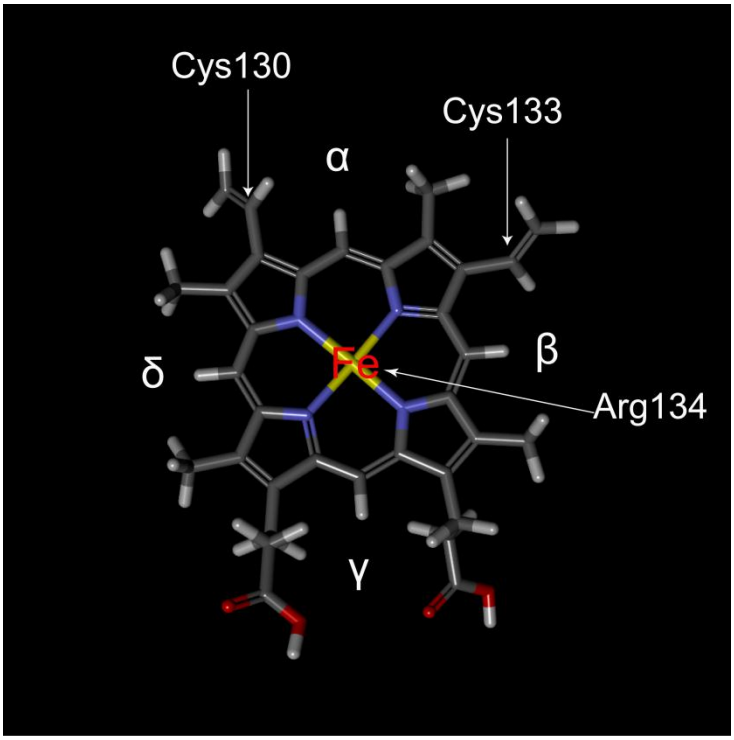

Figure 3. Heme linked positions for ORF3a

The conservative domain of electronic transmission. At present, the detail mechanism of electron transmission in cells is not yet clear. The electron transfer related heme is divided into two parts: from the electron donor to the low-rotation heme and from the low-rotation heme to the binuclear center. The first part can be a multi-step process and varies between different families, while the second part (direct transfer) is consistent throughout the superfamily. It was believed here that ORF3a must have a conserved domain of human cytochrome C oxidases in order to have limited electron transport capabilities. Then, the MEME tool was used to compare ORF3a with human cytochrome C oxidases (AGW78696.1 cytochrome C oxidase subunit I (mitochondrial) [Homo sapiens]). Therefore, it was found that both have a conservative domain "RJWECWKCKRKNPLLEEPNMNLCWHYGCPDPC", E value = 1.0E-002<0.05 (**Figure 4.A**), which has certain statistical significance. The domain of cytochrome C oxidases is 471-502 (**Figure 4.B**), which is the active region for electron transport. The

ORF3a site is 126-157 (**Figure 4.C**), including the conserved domain exposed to the heme, while Cys130, Cys133, and Arg134 are still conserved sites in the electron transport domain.

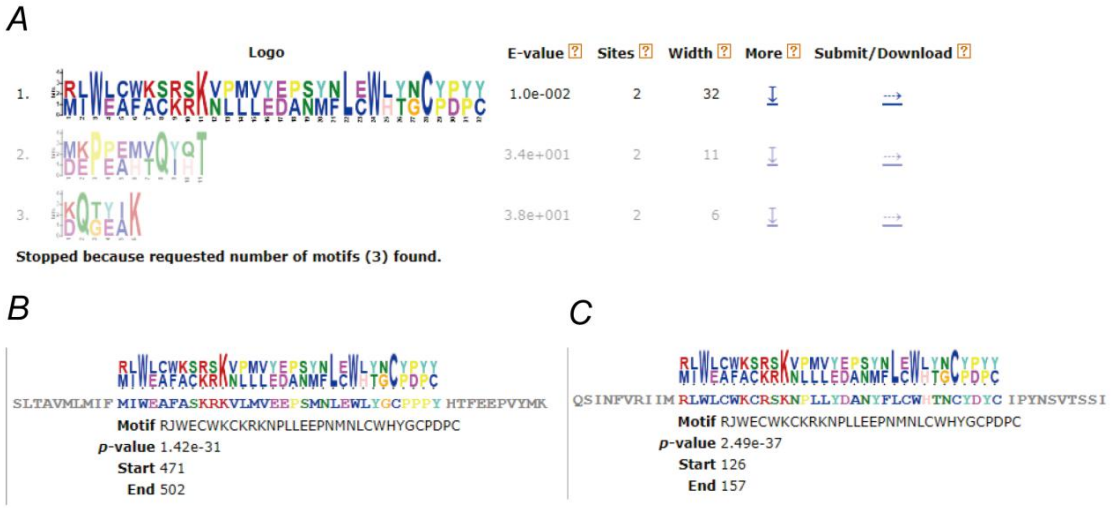

Figure 4. Electronic transmission domain of ORF3a. ***A***. Electronic transmission domain "RJWECWKCKRKNPLLEEPNMNLCWHYGCPDPC". ***B***. Electronic transmission sites at positions 471-502 of AGW78696.1. ***C***. Electronic transmission sites at positions 126-157 of ORF3a.

The conserved domain of EFeB. Researchers found EFeB protein (a type of dechelatase) in bacteria (such as *E. coli*). The conserved domain of EFeB dissociates the iron in the heme by means of heme oxidases, while the porphyrin ring remains intact. It also has the ability of cytochrome reductases to achieve electron transfer. Pseudomonas fluorescens is a bacterium that can multiply in the blood and release endotoxins, which can cause respiratory infections, sepsis or shock in animals. In addition to that, the MEME tool was used to compare ORF3a with Pseudomonas fluorescens dechelase (tr | A0A4Y9TPJ9 | A0A4Y9TPJ9_PSEFL dechelatase/peroxidase OS = Pseudomonas fluorescens OX = 294 GN = efeB PE = 3 SV = 1). Then, we obtained the conservative domain "IJRRPWNCWKCRNKNGQLDDGNLFJCWQADCEDGCI", E-value = 1.3e-2 <0.05 (**Figure 5.A**), with certain statistical significance. The position of dechelatase is 347-382(**Figure 5.B**), which belongs to the conserved domain of EFeB(**Figure 5.C,** Screenshot from Interpro website). The ORF3a site is 123-158(**Figure 5.D**), including the conserved domains both in contact with the heme and in electron transport. Cys130, Cys133 and Arg134 are still EFeB conserved domains. If the conserved domain of electron transport is excluded from that of EFeB of ORF3a, only IIe is retained at both ends. IIe at both ends may play an important role in the dissociation of iron from the heme. Since the mechanism of EFeB is still a mystery, molecular simulation technology can't be adopted to simulate the mechanism of iron dissociation from the heme.

We could not fully simulate the dissociation process of iron from heme. Through molecular simulation of DS, we directly connected Cys130, Cys133 and Arg134 of ORF3a to the corresponding positions of heme, and performed "minimization". The two N bonds connected to Fe were automatically broken, but the other two N bonds were not broken. At present, the mechanism of EFeB is not completely clear. We made many attempts and did not know how to break the other two N bonds and keep the porphyrin ring intact.

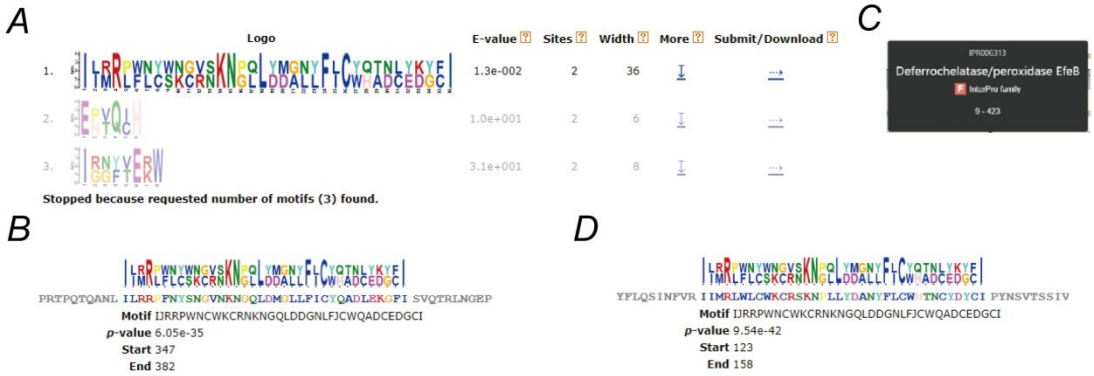

Figure 5. EFeB motif of ORF3a. ***A***. EFeB domain "IJRRPWNCWKCRNKNGQLDDGNLFJCWQADCEDGCI". ***B***. EFeB domain of A0A4Y9TPJ9. ***C***. EFeB domain sites at positions 347-382 of A0A4Y9TPJ9. ***D***. EFeB domain sites at positions 123-158 of ORF3a.

Bacterial control group analysis. To prove that the above analysis is not accidental, a comparative analysis was conducted. Firstly, MEME was used to compare the domains of cytochrome oxidase of Aquificales and viral proteins ORF3a. Then, two types of bacteria, the positive pair of bacteria (**Table 3**) such as Hydrogenophilus and the negative control bacteria (**Table 4**) such as Mycobacterium tuberculosis were randomly selected. For the two types of bacteria, the domains of cytochrome oxidases were searched separately, and the E-value of many results was greater than 0.05, or no conserved heme domain was found. It indicates that ORF3a protein may dissociate the iron from the heme in an ancient way, which is different from the general bacterial protein. The electron transport region partially overlapped with cytochrome c oxidase subunit I of Legionella and Streptococcus pneumoniae, where the viral pneumonia condition was similar to two bacteria infections.

Table 3. conserved domain between Cytochrome oxidase of Aquificales and ORF3a

ID	Species	Protein	E-value	Positions of oxidase	Positions of ORF3a
AAC07899.1	Aquifex aeolicus VF5	cytochrome c oxidase subunit I	3.00E-03	472-481	148-157
BBD76882.1	Hydrogenophilus thermoluteolus	cytochrome b561	3.20E+00	-	-
AAP84291.1	Terricola subterraneus	cytochrome b, partial	3.00E-01	-	-
Q56408.2	COX1_THETH Thermus_thermophilus	Cytochrome c oxidase subunit 1	8.10E-01	-	-
Q5SJ79.1	COX1_THET8 Thermus_thermophilus_HB8	Cytochrome c oxidase subunit 1	1.40E+00	-	-
AOV85240.1	Zelotes subterraneus	cytochrome oxidase subunit 1, partial	4.90E+00	-	-

Table 4. conserved domain between Cytochrome oxidase of bacteria and ORF3a

ID	Species	Protein	E-value	Positions of oxidase	Positions of ORF3a
NP_705912.1	Cryptococcus neoformans var. grubii	Cytochrome c oxidase subunit 1	4.60E-01	-	-
PCQ20475.1	Klebsiella pneumoniae	Cytochrome c oxidase subunit 1	1.30E+00	-	-
WP_027270427.1	Legionella	Cytochrome c oxidase subunit 1	1.80E-02	420-429	148-157
VCU51329.1	Mycobacterium tuberculosis	Cytochrome c oxidase subunit 1	2.70E-01	-	-
AVZ37898.1	Pseudomonas aeruginosa	Cytochrome-c oxidase	5.90E-01	-	-
WP_025904654.1	Staphylococcus	Cytochrome c oxidase subunit 1	1.00E+00	-	-
COP70899.1	Streptococcus pneumoniae	Cytochrome c oxidase subunit 1	2.20E-02	512-536	113-137
AAK33439.1	Streptococcus pyogenes M1 GAS	hypothetical protein SPy_0407	2.20E+00	-	-

Coronavirus control group analysis. To prove that the heme linked sites of ORF3a evolved from bats, the ORF3/ORF3a proteins of several coronaviruses, SARS viruses and MERS viruses were analyzed by comparing with FixO (AGG15137.1 cytochrome C oxidase mono-heme subunit/FixO [Hydrogenobaculum sp. HO]) (**Table 5**). Table 5 shows that ORF3/ORF3a of coronavirus and MERS viruses has no conserved domains of heme linked sites. Table 5 shows that ORF3a of the SARS virus in various animals has a conserved domain of heme linked sites.

ORF3a protein may kick off the iron of the heme through electron transfer, oxidation and reduction. The attack of ORF3a on the heme is consistent with the research report of ORF3a causing apoptosis. the heme can regulate apoptosis through cytochrome C and other molecules. We checked the conserved domain of the ORF3a protein from NCBI, and the results showed that the ORF3a protein belongs to the APA3_viroporin super family (number: pfam11289) that is a pro-apoptotic protein. The severe acute respiratory syndrome coronavirus (SARS-CoV) causes apoptosis of infected cells, which is one of the culprits. Therefore, the iron dissociated from the heme by ORF3a may be the direct cause of apoptosis.

Table 5. Heme linked motifs in ORF3/ORF3a protein of some coronavirus

No.	ID	Species	Protein	E-value	Positions of FixO	Positions of ORF3/ORF3a	Motif
1	AYV99804.1	SARS coronavirus Urbani	ORF3a	1.40E-03	111-121	124-134	IMRCWGCWKCK
2	AAP72975.1	SARS coronavirus HSR 1	ORF3a	1.40E-03	111-121	124-134	IMRCWGCWKCK
3	AAP33698.1	SARS coronavirus Frankfurt 1	ORF3a	1.40E-03	111-121	124-134	IMRCWGCWKCK
4	AAU04650.1	SARS coronavirus civet010	ORF3	7.50E-04	111-121	124-134	IMRCWGCWKCK
5	AAQ94061.1	SARS coronavirus AS	ORF3a	1.40E-03	111-121	124-134	IMRCWGCWKCK
6	AAV91632.1	SARS coronavirus A022	ORF3	5.10E-04	111-121	124-134	IMRCWGCWKCK
7	AAU04635.1	Civet SARS CoV 007/2004	ORF3	5.10E-04	111-121	124-134	IMRCWGCWKCK
8	ATO98232.1	Bat SARS-like coronavirus	ORF3a	1.50E-03	111-121	124-134	IMRCWGCWKCK
9	YP_003858585.1	Bat coronavirus BM48-31/BGR/2008	ORF3	5.10E-04	111-121	124-134	IMRCWGCWKCK
10	QIA48642.1	Pangolin coronavirus	ORF3a	2.80E-03	111-121	124-134	IMRAWGCWKCR
11	APD51508.1	229E-related bat coronavirus	ORF3	2.50E-02	108-121	104-117	GRWFHMAWGCWQCK
12	AZF86131.1	Alpha coronavirus Bat-CoV/P.kuhlii/Italy/206679-3/2010	ORF3	1.40E-02	109-121	128-140	CWIYNFEGCWCCR
13	AAZ67053.1	Bat SARS CoV Rp3/2004	ORF3	1.00E-05	108-121	121-134	CRWIMRCWGCWKCR
14	ABD75326.1	Bat SARS CoV Rm1/2004	ORF3a	1.30E-05	108-121	121-134	CRWIMRCWGCWKCR
15	ABD75316.1	Bat SARS CoV Rf1/2004	ORF3a	9.8E-05	108-121	121-134	CRWIMRCWGCWKCK
16	AAT76153.1	SARS coronavirus TJF	ORF3	1.60E+00	-	-	-
17	ATN23890.1	Rhinolophus bat coronavirus HKU2	ORF3	4.20E+00	-	-	-
18	AWH65922.1	Pipistrellus bat coronavirus HKU5	ORF3	4.00E+00	-	-	-
19	YP_009328936.1	NL63-related bat coronavirus	ORF3	1.70E-01	-	-	-
20	DX59483.1	Miniopterus bat coronavirus/Kenya/KY27/2006	ORF3	1.30E+00	-	-	-
21	ACA52172.1	Miniopterus bat coronavirus HKU8	ORF3	2.80E+00	-	-	-
22	AVN89325.1	MERs coronavirus	ORF3	2.50E-01	-	-	-
23	AFO70498.1	Human coronavirus NL63	ORF3	2.50E-01	-	-	-
24	AHY21470.1	Human betacoronavirus 2c Jordan-N3/2012	ORF3	8.30E-02	-	-	-
25	ADX59467.1	Eidolon bat coronavirus/Kenya/KY24/2006	ORF3	2.60E+00	-	-	-
26	YP_009380522.1	Coronavirus AcCoV-JC34	ORF3	3.90E+00	-	-	-
27	ADX59459.1	Chaerephon bat coronavirus/Kenya/KY41/2006	ORF3	2.60E+00	-	-	-
28	ADX59452.1	Cardioderma bat coronavirus/Kenya/KY43/2006	ORF3	1.40E+00	-	-	-
29	ACA52158.1	Bat coronavirus 1B	ORF3	9.90E-01	-	-	-
30	YP_001718606.1	Bat coronavirus 1A	ORF3	2.10E-01	-	-	-
31	QBP43280.1	Bat coronavirus	ORF3a	7.80E-01	-	-	-

3.3 Heme linked sites of envelope and nucleocapsid phosphoprotein

It was found that the envelope and nucleocapsid phosphoprotein have domains similar to heme linked sites with the same methods of ORF3a protein, respectively. The conserved domain of the envelope is "CWQCCN" (**Figure 6.A**), E-value=1.1e-2<0.05, which is featured with certain statistical significance. The position of AGG15137.1 is 117-122 (**Figure 6.B**), and the sequence fragment is "CWQCHS". The position at the envelope is 40-45 (**Figure 6.C**), and the sequence fragment is "CAYCCN". Through the above analysis, it can be seen that the first two Cys40,Cys43 on this domain are connected to the carbon at the left end of the alpha position and the beta position, respectively, and the last Cys44 is related to the iron of the heme. Besides, the MEME tool was used to compare the envelope with cytochrome d oxidases (AGH93087.1 cytochrome d oxidase cyd, subunit II [Hydrogenobaculum sp. SN]). Then, it was found that the envelope has a conserved domain (**Figure 7**), which does not overlap with the heme linked site.The conserved domain of the envelope is "WRHTFD" (**Figure 7.A**), E-value=2.7e-002<0.05, which is featured with certain statistical significance. The position of AGH93087.1 is 118-123 (**Figure 7.B**), and the sequence fragment is "WRHTFD". The position at the envelope is 80-85 (**Figure 7.C**), and the sequence fragment is "QIHTID".

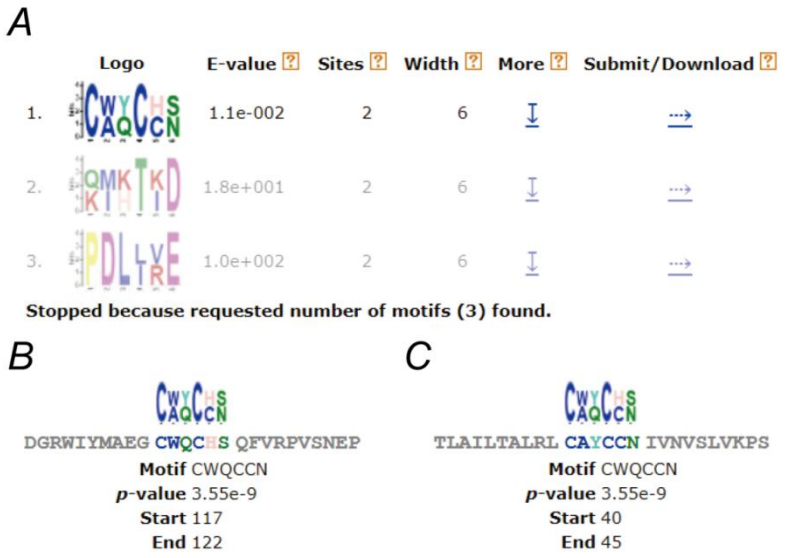

Figure 6. Heme motif of Envelope. *A*. Heme-linked domain "CWQCCN". *B*. heme linked sites at positions 117-122 of AGG15137.1. *C*. heme linked sites at positions 40-45 of Envelope.

The conserved domain of nucleocapsid phosphoprotein is "CWPCIAQF" (**Figure 8.A**), E-value=7.6e-3<0.05, which has certain statistical significance. The position of AGG15137.1 is 117-124 (**Figure 8.B**), and the sequence fragment is "CWQCHSQF". The position of nucleocapsid phosphoprotein is 300-307 (**Figure 8.C**), and the sequence fragment is "HWPQIAQF". Through the above analysis, it can be found that the first His300 and the first Gln303 on this domain connect the carbon at the left end of the alpha and beta positions, respectively, and the last IIe304 connects the iron of the heme. Confusingly, N's heme linked site has no Cys. Due to the limitation of computing resources and methods, it is unable to analyze the role of N heme linked sites.

However, further domain analysis showed that the envelope and nucleocapsid phosphoprotein did not have similar conserved domains as human cytochrome C oxidases (AGW78696.1). It may be that the electron transmission of the envelope and nucleocapsid is special. Meanwhile, nucleocapsid

phosphoprotein has no similar domain compared with AGH93087 too. Further domain analysis also showed that the envelope and nucleocapsid phosphoprotein did not have a conserved domain similar to Dechelatase (A0A4Y9TPJ9) of Pseudomonas fluorescens. It is probable that the envelope and nucleocapsid phosphoprotein cannot dissociate the iron from the heme at all.

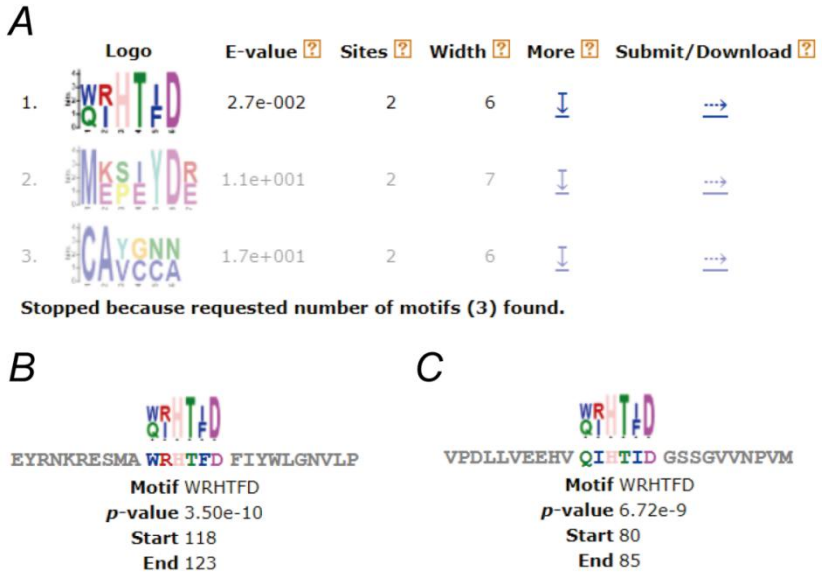

Figure 7. Electronic transmission motif of Envelope. ***A***. Electronic transmission domain "WRHTFD". *B*. Electronic transmission sites at positions 118-123 of AGG15137.1. *C*. Electronic transmission sites at positions 80-85 of Envelope.

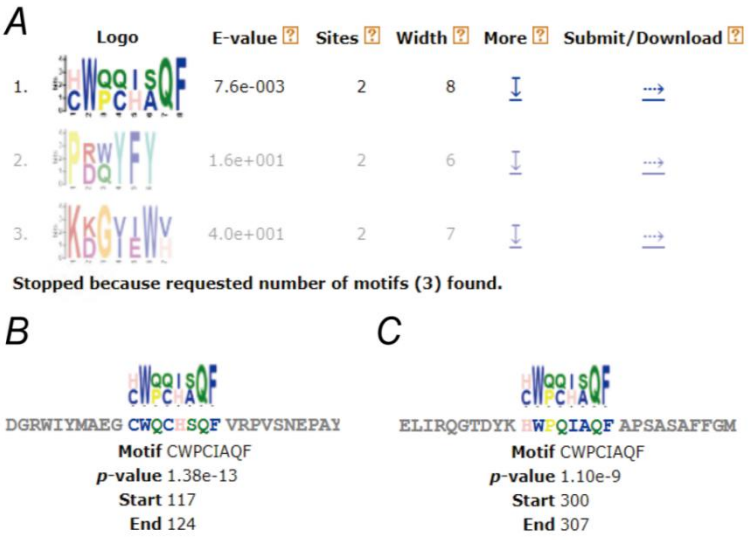

Figure 8. Heme motif of nucleocapsid phosphoprotein. ***A***. Heme domain "CWPCIAQF". *B*. heme linked sites at positions 117-124 of AGG15137.1. *C*. heme linked sites at positions 300-307 of nucleocapsid phosphoprotein.

3.4 Viral non-structural protein attacks the 1-beta chain of the hemoglobin

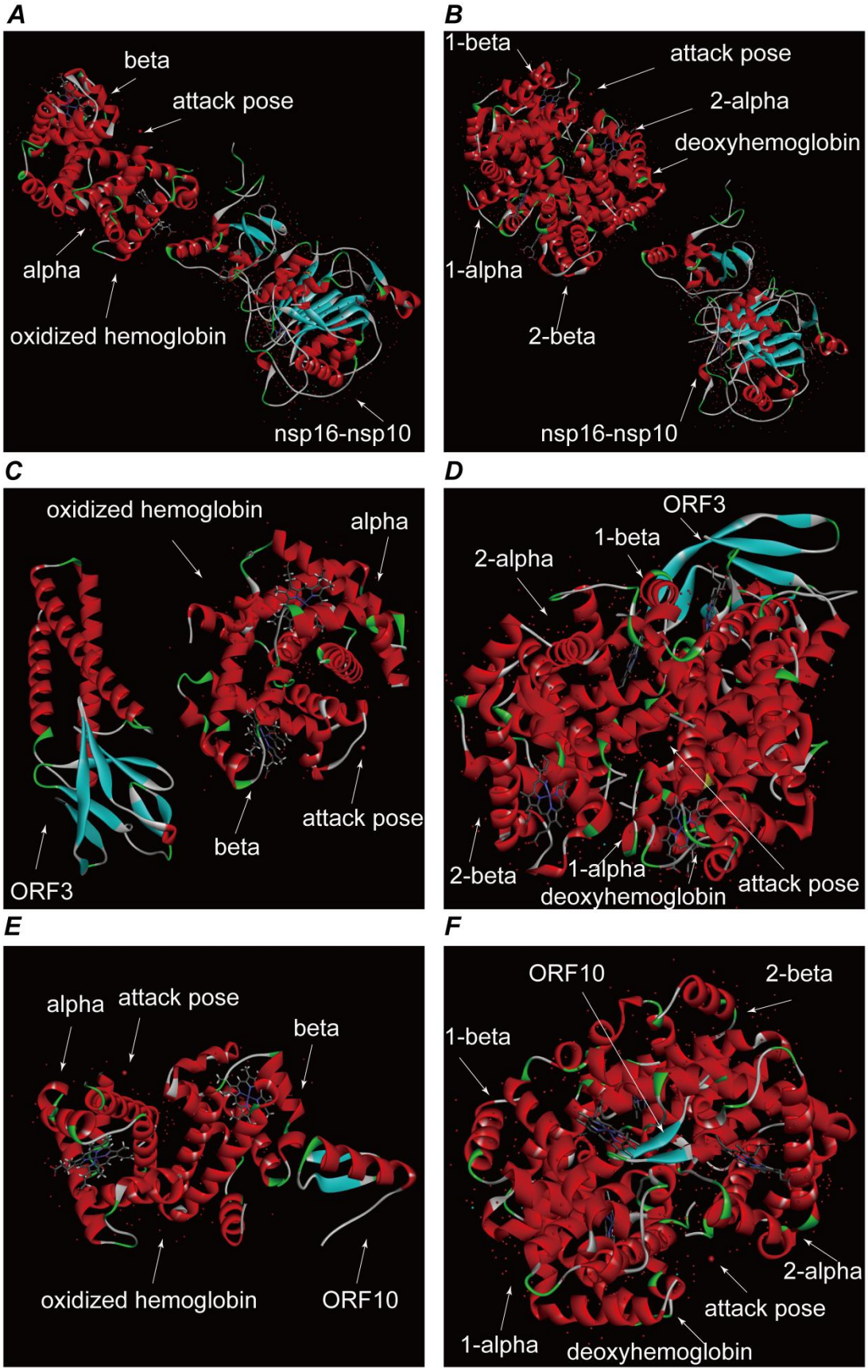

Figure 9. Viral non-structural protein attacks hemoglobin. ***A***. nsp16-nsp10 attacks the oxidized hemoglobin. ***B***. nsp16-nsp10 attacks deoxyhemoglobin. ***C.*** ORF3a attacks the oxidized hemoglobin. ***D***.

ORF3a attacks deoxyhemoglobin. ***E***. ORF10 attacks the oxidized hemoglobin. ***F***. ORF10 attacks deoxyhemoglobin.

Attacks the 1-beta chain of the hemoglobin. The iron porphyrin, namely the heme is the main porphyrin in the human body. Many the heme are not free but bound to hemoglobin. The novel coronavirus may attack hemoglobin, target the heme, and acquire porphyrins. In order to study the attack behavior of nsp16-nsp10 (orf1ab), ORF3a, and ORF10 proteins, ZDOCK molecular docking technology was adopted to examine these three proteins, and it could also analyze protein interactions and find the approximate location of these three proteins on hemoglobin.

For oxidized hemoglobin, nsp16-nsp10 acted on the middle bottom of the alpha and beta chains and was close to the alpha chain (**Figure 9.A**), while ORF3a acted at the bottom of the beta chain (**Figure 9.C**). ORF10 acted below the alpha chain **(Figure 9.E**). The possible mechanism was that nsp16-nsp10 first attacked the alpha chain, and then, ORF3a and ORF10 successively attacked the beta chain.

For deoxyhemoglobin, nsp16-nsp10 acted on the top of the 1-beta (**Figure 9.B**), and ORF3a acted at the bottom of the 1-beta (**Figure 9.D**), while ORF10 acted on the top of 1-beta (**Figure 9.F**). ORF3a and ORF10 have embedded themselves in deoxyhemoglobin and directly docked to the heme of the beta chain, which indicates that the viral protein can attack the heme on hemoglobin. The possible mechanism is that nsp16-nsp10 first attacks the 1-beta chain, and then, ORF3a and ORF10 successively attack the 1-beta chain.

It is challenging to perform molecular simulations. Due to the close distance of the attack postures of some proteins, the order of three proteins attacked is still unclear. The nsp16-nsp10 molecule may be an essential protein, playing a vital role throughout the attack. It is worth noting that the above simulation shows that deoxyhemoglobin is more vulnerable than oxidized hemoglobin. Attack of oxidized hemoglobin by viral proteins will lead to less and less hemoglobin that can carry oxygen. The invasion of viral proteins on deoxyhemoglobin will make increasingly less hemoglobin that can carry carbon dioxide and blood sugar. People with diabetes can have unstable blood sugar, and body cells have extreme inflammation due to excess iron, carbon dioxide and oxygen. Patients with respiratory distress will be worse, and organs and tissues of the whole body experience different degrees of damage.

Control analysis of ORF3a protein attack specificity. ORF3a was a vital protein, which played an important role in attacking hemoglobin and dissociate heme. To determine ORF3a attack specificity, we docked several erythrocyte catalase, cytochrome oxidases, and blue blood proteins to ORF3a, respectively (**Table 6**). **Table 6** showed that the ORF3a protein also effectively attacked cytochrome C G41S (**Figure 10**). **Figure 10** showed that the ORF3a protein completely embedded in Human cytochrome C G41S, and heme was blocked. Cytochrome C is a cytochrome oxidase, the only peripheral protein in the electron transport chain located in the outer membrane inside the mitochondria. Cytochrome C is distributed more in cardiomyocytes. Cytochrome C plays an essential role in cellular respiration and participates in transferring electrons to oxygen. The attack of cytochrome C G41S caused hypoxia and poisoning of tissue cells and hindered energy metabolism. The result indicated that people with cytochrome C mutations were at significant risk. Fortunately, ORF3a did not affect standard cytochrome C.Since the structure of human oxidized and deoxymyoglobin was not found in the PDB database for protein docking, it was unclear whether ORF3a can attack myoglobin.

ORF3a plays an essential role in attacking hemoglobin, which may be specific. In other words, the existing calculation model is based on the docking results of the earlier ORF3a sequence (without

considering variation) and the standard hemoglobin structure. If the structure of ORF3a or hemoglobin changes, it may get uncertain results: either hemoglobin is more likely to be attacked; or hemoglobin is less likely to be attacked.

Table 6. Docking results between ORF3a and metal porphyrin proteins

No.	Name	Portein	Fully Embed
1	3nwv	Human cytochrome C G41S	Yes
2	5z62	Human cytochrome C	None
3	1dgf	Human erythrocyte catalase	None
4	1umk	Human Erythrocyte NADH-cytochrome b5 Reductase	None
5	1w0g	Human cytochrome P450 3A4	None
6	1OG5	P450 CYP2C9	None
7	1v4w	Deoxygenated form of bluefin tuna hemoglobin	None
8	3qjo	Functional unit (KLH1-H) of keyhole limpet hemocyanin	None
9	1lla	Deoxygenated limulus polyphemus subunit II hemocyanin	Extremely weak
10	3ixv	Static pseudoatomic model of scorpion hemocyanin	None
11	1nol	Oxygenated hemocyanin (subunit type ii)	None
12	1js8	Octopus Hemocyanin Functional Unit	None
13	1lnl	deoxygenated hemocyanin from Rapana thomasiana	None

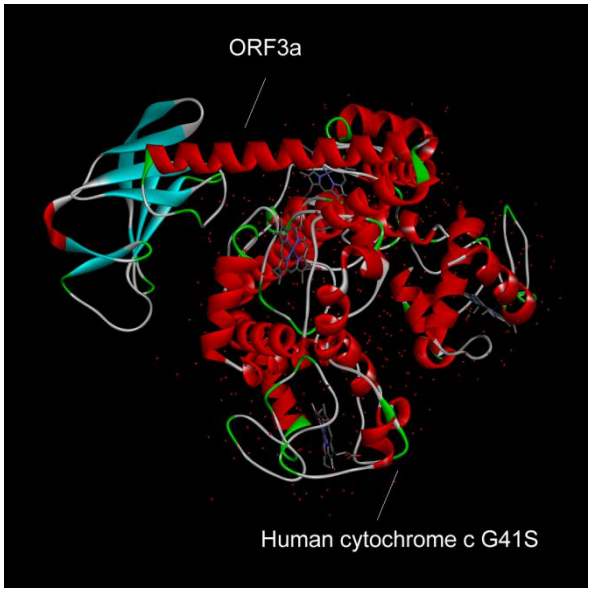

Figure 10. ORF3a protein fully embedded in human cytochrome C G41S

3.5 Virus structural proteins binding porphyrins

The binding of four structural proteins to porphyrins was studied through molecular docking. The 3D-structural file of heme was downloaded from the PDB database.

The SWISS-MODEL online server modeled the surface glycoproteins to produce a three-dimensional structure, and the E2 templates were selected. The three-dimensional structure file of spike protein is downloaded from the PDB database website. Discovery-Studio realized molecular docking of surface glycoproteins and the porphyrin. Besides, the spike protein was used to dock with porphyrins. Some sites can dock successfully, but the calculation of binding energy is not successful. E2 glycoprotein is derived from template 1zva.1.A. The docking of E2 glycoprotein and the heme was also fruitless. When the iron ion was removed, the heme became a porphyrin, and many kinds of docking were finalized between the E2 glycoprotein and the porphyrin. Calculating the binding energy, the docking pose with the lowest binding energy 8.93 kcal/mol (**Table 7**) was accepted.

When analysing envelope protein, the same methods were adopted. The envelope protein was modeled by Robetta. Discovery-Studio found several kinds of docking of the envelope protein and the porphyrin, where the docking pose with the lowest binding energy -27.43 kcal/mol (**Table 7**) was chosen. The homology modeling file of the membrane protein comes from a part of the project on COVID-19 viral protein modeling published by the AlphaFold group. The molecular docking technology of Discovery-Studio's LibDock was adopted to achieve the docking of membrane proteins and porphyrins. The posture with the lowest binding energy -288.46 kcal/mol (**Table 7**) was selected. The same methods were adopted to analyze the nucleocapsid phosphoprotein. The nucleocapsid phosphoprotein was modeled by Robetta too. Discovery-Studio provides the docking between the nucleocapsid phosphoprotein and the porphyrin with the lowest binding energy -295.26 kcal/mol (**Table 7**). .

As mentioned before, it is impossible to assess whether the surface glycoprotein can bind porphyrins. It was found that the binding energy of nucleocapsid phosphoprotein was the lowest, while the binding energy of E2 glycoprotein was the highest, and the binding energy of envelope protein and membrane glycoprotein was medium, which means that binding E2 glycoprotein to the porphyrin is the most unstable; the binding of binding envelope protein to the porphyrin is stable, and that of membrane glycoprotein to the porphyrin is stable too. In addition, nucleocapsid phosphoprotein to the porphyrin is the most stable. However, the specific mechanism concerning binding of most viral structural proteins to porphyrins is not clear.

Table 7. Binding energy of structure proteins and the porphyrin

Viral structureprotein	Libdockscore	Binding Energy (kcal/mol)
surface glycoprotein (E2) (modeling)	87.94	8.93
envelope protein (modeling)	110.77	-27.43
membrane glycoprotein (modeling)	83.43	-288.46
nucleocapsid phosphoprotein (modeling)	112.63	-295.26

3.6 Virus non-structural proteins bind to the porphyrin

Molecular docking technique was used to analyze the binding of non-structural proteins to porphyrins. orf1ab can be cleaved into many sub-proteins, but the crystal structure of many sub-proteins has not been determined, and the detailed mechanism of action is imprecise. The crystal structure of the nsp16-nsp10 complex, nsp9 and main protease proteins of orf1ab have been determined. In this article, the docking of these proteins with porphyrins was mainly studied. Then, molecular docking of nsp16-nsp10 protein and porphyrins was finished by Discovery-Studio. nsp16-nsp10

protein and the heme could not complete the docking experiment, but by removing iron ions, they could make the heme into a porphyrin, and the radius of action increased. Then, several types of docking were completed. Besides, the binding energy was calculated. Through inspection, it was considered that nsp16-nsp10 could effectively bind to porphyrins with binding energy -197.41 kcal/mol. Similarly, we completed the docking of Nsp9 RNA-replicase, Nsp9 RNA binding, main protease in the apo form, main protease free and four forms of protein, and calculated the binding energy. The minimum binding energy of this form of proteins is -189.81kcal/mol, -201.78kcal/mol, -120.35kcal/mol and -120.18 kcal/mol (**Table 8**).

Table 8. Binding energy of non-structural proteins and the porphyrin

Viral non-structuralprotein	Libdockscore	Binding Energy (kcal/mol)
nsp16 - nsp10 (orf1ab)	97.97	-197.41
Nsp9 RNA-replicase (orf1ab)	60.14	-189.81
Nsp9 RNA binding (orf1ab)	75.58	-201.78
main protease in apo form (orf1ab)	120.86	-120.35
main proteasefree (orf1ab)	73.50	-120.18
ORF3a (modeling)	73.74	-234.85
ORF6a (modeling)	-	-
ORF7a (modeling)	90.01	-211.14
ORF8 (modeling)	72.42	-229.20
ORF10 (modeling)	-	-

The same analysis steps were followed to the binding of ORF3a protein to porphyrins. The structure file of ORF3a was downloaded from the AlphaFold project. The docking result represents the molecular model of ORF3a protein binding to the porphyrin, where the docking poses with the lowest binding energy -234.85 kcal/mol (**Table 8**). To study the binding properties of ORF7a protein to porphyrins, the same analysis steps as the structural protein method were used. The structure file of ORF7a was modeling and downloaded from the ROBETTA server. The ORF7a protein and the porphyrin had the lowest binding energy -211.14 kcal/mol (**Table 8**). The same methods of ORF7a protein were adopted to analyze the ORF8 protein. There were several kinds of docking of the ORF8 protein and the porphyrin, where the docking pose with the lowest binding energy -229.20 kcal/mol (**Table 8**) was selected.

The docking of ORF6a with porphyrins fails. The docking of ORF10 with porphyrins also fails, which means that ORF10, ORF6 cannot bind porphyrins effectively.

The results show that orf1ab, ORF3a, ORF7a, and ORF8 can bind porphyrins, while ORF10, and ORF6 cannot. The binding energies of orf1ab, ORF3a, ORF7a, ORF8, and porphyrins were compared respectively. The orf1ab protein plays many roles, and the detailed mechanism of its binding to porphyrins is unclear. As the sequences of ORF10 and ORF6 are extremely short, they should be short signal peptides.

We will analyze the function of heme-linked sites of ORF3a, N, and E proteins through the verification analysis of drugbank's drugs later. Since the existing calculation methods have not found the heme-linked sites of S, ORF1ab, ORF8, and other proteins, we will not do a comparative analysis

on them for the time being.

3.7 Drugs bound to the heme-iron linked site of ORF3a

From the above analysis, ORF3a had domains similar to the respiratory chain (electron transport): heme-linked domain, electron transport domain, and EFeB domain kicked off the iron. It was necessary to block ORF3a's respiratory chain. That is, the drug is bound to the respiratory chain domain and blocked the electron transfer process. Because the mechanism of EFeB was unclear, it was not yet possible to determine the therapeutic effect of the drug bound to the respiratory chain site by calculation methods: the impact of inhibiting or promoting attack. However, if the drug linked to the heme iron-linked site of Arg134, it could directly prevent the iron of heme linked to ORF3a . Therefore, it was difficult for ORF3a to produce redox reactions related to electron transfer. In this study, the analysis of drugs' binding effect mainly based on this principle, that was, linked to the heme-iron-site of the viral protein.

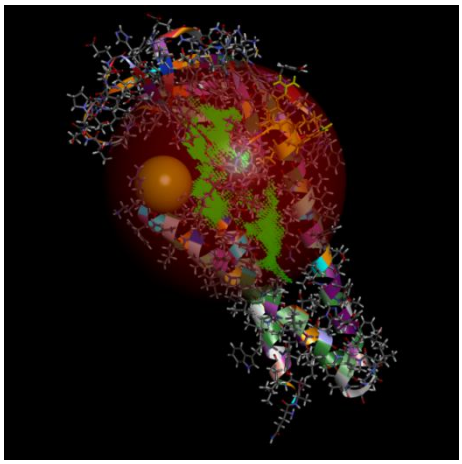

Figure 11. Hematoporphyrin linked region of ORF3a protein. The active region of Site 1 was the most significant red ball region. In the upper right corner of the red sphere area (marked in yellow) was the heme-linked sites of ORF3a ("CWKCR").The lower-left corner of the red ball zone was the Site6 active zone.

Through the "prepare protein" and "from receptor cavities" operations of DS, 8 sites---active areas were obtained from ORF3a. The previous analysis results showed that ORF3a bound to hematoporphyrin through the active area (**Figure 11**). In Figure 11, the active region of Site1 was the most significant red ball region, and the binding energy of hematoporphyrin was -51.542 kcal/mol (threshold T1). The active area of Site1 also included ORF3a-like respiratory chain sites. In the upper right corner of the red sphere area (marked in yellow) was the heme-binding sites of ORF3a ("CWKCR"). The lower-left corner of the red ball zone was the Site6 active zone; the hematoporphyrin binding energy was -234.85 kcal/mol (T2 value). There was a cavity between Site6 and the "CWKCR" area. ORF3a mainly captured heme through the active area of Site6 and then sent it to the "CWKCR" site through this cavity. Site1 activity area also directly obtained heme and sent it to the "CWKCR" location, but this was not the primary way. If the drug had a particular inhibitory effect, its binding energy was less than the threshold T1. For complete suppression, the binding energy was less than the threshold T2. When ORF3a saturated attacked hemoglobin and produced a large amount of heme, drugs with binding energy higher than T2 could not compete with heme for binding.

We downloaded more than 8,000 small molecule compounds from drugbank, and after passing

“prepare ligand” of DS, we got more than 7000 drugs. For of these ligands and ORF3a, it took us for half of a month to dock and compute the binding energy. In the molecular docking test of site1 active area, the binding energies of 3366 drugs were less than 0, while the binding energies of only 1,292 drugs were less than the threshold -51.542 kcal/mol, and the binding energies of 168 drugs were less than the threshold -234.85 kcal/mol. These batches of directories were analyzed.

Table 9. Binding ability of several drugs on the heme-iron linked site of ORF3a protein

No.	Formula	Generic name	Arg134-Fe site	Binding Energy	Binding Ability
1	C12H28O7P2	Lauryl alcohol diphosphonic acid	Yes	-267.94	Medium
2	C16H38O6P3	(2,2-diphosphonoethyl) (dodecyl) dimethylphosphonium	Yes	-242.01	Medium
3	C5H12O7P2	isopentenyl pyrophosphate	Yes	-189.17	Low
4	C21H28N7O17P3	Nicotinamide adenine dinucleotide phosphate	Yes	-183.90	Low
5	C22H21N8O8PS4	Ceftaroline fosamil	Yes	-177.58	Low
6	C29H32Cl2N2O5S	Lusutrombopag	Yes	-174.08	Low
7	C14H26N4O11P2	Citicoline	Yes	-173.97	Low
8	C15H28O6P2S	Farnesyl thiopyrophosphate	Yes	-173.39	Low
9	C12H17N2O4P	Psilocybin	Yes	-172.36	Low
10	C5H12O7P2	Dimethylallyl Diphosphate	Yes	-162.36	Low
11	C5H15N2O3PS	Amifostine	Yes	-150.92	Low
12	C15H27N3O12P2	6-Aminohexyl-uridine-C1,5'-diphosphate	Yes	-149.30	Low
13	C15H30N5O5	N-[N-[1-Hydroxycarboxyethyl-Carbonyl]Leucylamino-Butyl]-Guanidine	Yes	-148.11	Low
14	C17H30NO5P	[(3,7,11-trimethyl-dodeca-2,6,10-trienyloxycarbamoyl)-methyl]-phosphonic acid	Yes	-146.79	Low
15	C30H41N3O10P2	AP-22408	Yes	-145.64	Low
16	C7H15N2O6P	N-(Phosphonoacetyl)-L-Ornithine	Yes	-142.67	Low
17	C20H36O7P2	Geranylgeranyl diphosphate	Yes	-139.08	Low
18	C30H34N3O7P	RU81843	Yes	-137.28	Low
19	C19H18F3N5O2S	Elenbecestat	Yes	-136.94	Low
20	C15H24N2O17P2	Galactose-uridine-5'-diphosphate	Yes	-136.19	Low
21	C6H8O6	Ascorbic acid	Yes	-81.48	Low
22	C27H35N6O8P	Remdesivir	Yes	-73.39	Low
23	C18H26ClN3O	Hydroxychloroquine	Yes	-41.65	Weak
24	C18H26ClN3	Chloroquine	Yes	-18.69	Weak
25	C13H20N2O3S	Articaine	None	-21.64	None
26	C27H37N3O7S	Darunavir	None	23.59	None
27	C5H4FN3O2	Favipiravir	None	35.52	None
28	C20H26N2O2	Hydroquinine	Yes	392.10	None
29	C20H24N2O2	Quinine	None	835.39	None

Note:

1. Arg134-Fe site is whether the drug can bind to the Arg134 site of ORF3a, which is the iron linked site of heme.

2. The threshold value T1=-51.5424 kcal/mol, the threshold value T2=-234.85 kcal/mol. The binding ability of the binding energy less than the threshold T2 is medium, the binding ability of the binding energy greater than the threshold T1 is weak, and the binding ability of the binding energy between T2 and T1 is low. The binding ability “None” is that drugs cannot bind Arg134-Fe site, or the binding energy is positive .

The DS tool was used to analyze the binding area of drugs with binding energy less than T one by one, to find the medicines bound Arg134-Fe site. After several days of analysis, we screened 21 drugs with Arg134-Fe site (**Table 9**) . Among these drugs, the most binding effective were Lauryl alcohol diphosphonic acid and (2,2-diphosphonoethyl) dimethylphosphonium, the GGPPS inhibitor, which is an anti-cancer drug. Interestingly, in addition to the GGPPS inhibitor, most drugs are also related to cancer treatment. For example, Ceftaroline fosamil was often used for antibacterial treatment of cancer patients. Farnesyl thiopyrophosphate had an inhibitory effect on Ras. Amifostine was a cytoprotective agent, mainly used for the adjuvant treatment of various tumors. Besides, a lower binding effect drugs was nicotinamide adenine dinucleotide phosphate (NADPH, Coenzyme II), an analog of NADH. NADPH was a reduction reaction molecule that transferred hydrogen and participated in biosynthesis; NADH was a molecule that participated in the respiratory chain reaction.

We also mapped the binding areas of several controversial drugs **(Table 9)**. The results show that Remdesivir, Hydroxychloroquine, and Chloroquine could all bind Arg134-Fe site. But Remdesivir binding energy was slightly lower than T1, and Hydroxychloroquine, Chloroquine binding energy was higher than T1. It showed that Remdesivir was relatively more obvious than Hydroxychloroquine and Chloroquine in terms of inhibitory effect. It was worth noting that Remdesivir was an adenosine analog. Besides, Articaine did not have Arg134-Fe site and may not inhibit ORF3a. The binding energy of Darunavir, Favipiravir, Hydroquinine, and Quinine were all greater than 0, will not generate stable complexes, and may not inhibit ORF3a too. If they were clinically useful, they also were through other mechanisms.

We used DS further to perform pharmacophore analysis on drugs such as GGPPS inhibitors, but the tool could not complete the analysis usually. Therefore, we did not yet know the effect of these drugs on pharmacophore.

3.8 Drugs bound to the heme-iron linked site of Nucleocapsid phosphoprotein

The spatial structure of nucleocapsid phosphoprotein was like a hippocampus (**Figure 12**). The large red circle in **Figure 12** was the Site 1 active region of N, and the binding energy of hematoporphyrin was T1 = -264.82 kcal/mol. The small red circle below the big red circle was the Site 2 active region of N, and the binding energy of hematoporphyrin was T2 = -295.26 kcal/mol. The yellow structure above the big red circle was a part of the heme linking area of N --- "HWPQI". His300 and Gln303 were the regions where heme carbon was connected, and Ile304 was the region where heme iron was connected. There was a cavity under the big red circle. Heme may be mainly hunted through Site 2 and then sent to the heme linked area through Site 1. The site 1 active area may not be the primary hunting way.

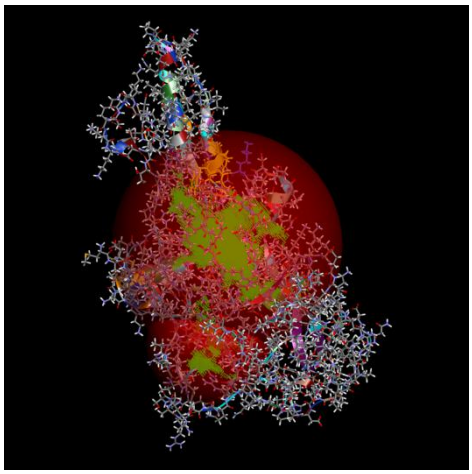

Figure 12. Hematoporphyrin linked region of Nucleocapsid_phosphoprotein. The active region of Site1 was the most significant red ball region. In the upper corner of the Site1 (marked in yellow) was the heme-linked sites of Nucleocapsid_phosphoprotein ("HWPQI"). The lower corner of the Site1 was the Site2 active zone.

Table 10. Binding ability of several drugs on the heme-iron linked site of N protein

No.	Formula	Generic name	Ile304-Fe site	Binding Energy	Binding Ability
1	C27H30N4O3	RPR128515	Yes	-216.918	Weak
2	C28H30N6OS	Masitinib	Yes	-195.326	Weak
3	C22H25N5O3S	Ziresovir	Yes	-180.206	Weak
4	C25H22N6O2	Sotrastaurin	Yes	-170.412	Weak
5	C21H29N7O13P2	3-Aminomethyl-Pyridinium-Adenine-Dinucleotide	Yes	-163.012	Weak
6	C10H15N3	N-[3-(aminomethyl)benzyl]acetamidine	Yes	-161.573	Weak
7	C19H22N2OS	Acepromazine	Yes	-157.806	Weak
8	C30H31F3N8O	Bafetinib	Yes	-148.752	Weak
9	C17H15N5O	6-amino-4-(2-phenylethyl)-1,7-dihydro-8H-imidazo[4,5-g]quinazolin-8-one	Yes	-136.444	Weak
10	C17H19N5O2	Piritrexim	Yes	-130.676	Weak
11	C27H26N4O2	3-(1H-indol-3-yl)-4-(1-{2-[(2S)-1-methylpyrrolidinyl]ethyl}-1H-indol-3-yl)-1H-pyrrole-2,5-dione	Yes	-125.283	Weak
12	C21H23N3OS	Periciazine	Yes	-122.43	Weak
13	C27H35N6O8P	Remdesivir	None	-41.5106	None
14	C18H26ClN3O	Hydroxychloroquine	None	183.024	None
15	C18H26ClN3	Chloroquine	None	105.354	None
16	C13H20N2O3S	Articaine	Yes	-4.1339	Extremely weak
17	C27H37N3O7S	Darunavir	None	13.6733	None
18	C5H4FN3O2	Favipiravir	Error	Error	Error
19	C20H26N2O2	Hydroquinine	None	12996.2	None
20	C20H24N2O2	Quinine	None	437.791	None

Note:

1. Ile304-Fe site is whether the drug can bind to the Ile304 site of N, which is the iron linked site of heme.

2. The threshold value T1=-264.82 kcal/mol, the threshold value T2=-295.26 kcal/mol. The binding ability of the binding energy less than the threshold T2 is medium, the binding ability of the binding energy greater than the threshold T1 is weak, and the binding ability of the binding energy between T2 and T1 is low. The binding ability “None” is that drugs cannot bind Ile304-Fe site, or the binding energy is positive.

We used similar methods to dock drugs with the N molecule, thereby obtaining 2160 drugs with a binding energy of less than 0, included 152 drugs with a binding energy of less than T1. We plotted docking results with lower binding energy one by one by DS 2016. As a result, there was no drug with binding energy less than T1 could be bind to the heme iron linked site. We continued to screen 12 medicines bound to the heme iron linked site with a binding energy of less than 0 (**Table 10**). Confusingly, some drugs were also anticancer drugs, such as marcetinib, sostatin, bafitinib, and pirixin.

Table 10 also lists the analysis results of some drugs in use. Drugs such as Remdesivir, Chloroquine, Hydroxychloroquine, Darunavir, Hydroquinine, and Quinine cannot bind the heme iron connection site. Remdesivir can participate in the heme carbon connection site, but it is unknown whether it can prevent N from binding heme iron. Favipiravirr's calculation simulation failed, and the docking failed. Articaine can bind the heme iron connection site, but the binding energy is too high. The drugs listed in Table 10, because the binding energy is more senior than T1, may be challenging to compete with heme and bind to N to form a stable complex.

3.9 No drug could bind to the heme-iron linked site of E protein

The spatial structure of E protein was like a shrimp (**Figure 13**)。 The structure marked by the red box as similar to the gathered clamp of the shrimp. It protected the yellow and red labeled structures at the top --- the part of heme linked domain,”CAYCC”. The heme linked sites were like a shrimp head, the yellow labeled structure was the heme carbon connection region (Cys40-Cys43), and the red labeled structure was the heme iron linked site (Cys44). The structure of the lower right corner of the head was like a shrimp body. Under the shrimp clamp was a cavity, which is the only active area Site 1. The connection from the cavity to the heme iron linked site was just a small hole, and some molecules blocked in the hole. The possible mechanism was that after the heme was attached to the cavity, the clamp was properly released and the heme was sent to the upper --- heme linked domain. Then, the clamp gathered again to protect the heme linked domain. Similar to previous analysis, a drug bound to the heme iron linked site to prevent some redox or catalytic reactions, and could have a certain inhibitory effect. It was impossible to determine role of a drug bound to the heme carbon linked sites by calculated methods.

We carried out molecular docking of drugs and E protein according to the previous method. The default cavity radius of DS software was very small. We increased the radius of the cavity to dock, but based on the results of the docking, the binding energy could not always be calculated. We used the default cavity radius to dock and calculate the binding energy, and obtained 847 drugs with a binding energy less than 0. For these drugs, we used DS to draw one by one. The results showed that all drugs were docked in and around the cavity behind the red box. There were many drugs that bound to the heme carbon linked sites, but no drugs bound to the heme iron linked site. To confirm this unfortunate result, we analyzed the 179 active sites of the site 1 region one by one. The analysis results showed that there was indeed no active site near the heme iron linked site. The result indicated that it was because

the hole from the cavity to the heme iron linked site was too small, and the existing drugs only bound to the cavity and its vicinity, and not reached the heme iron linked site. In other words, no drugs was found that inhibit heme iron linked sites. It may be the reason why the virus is highly contagious.

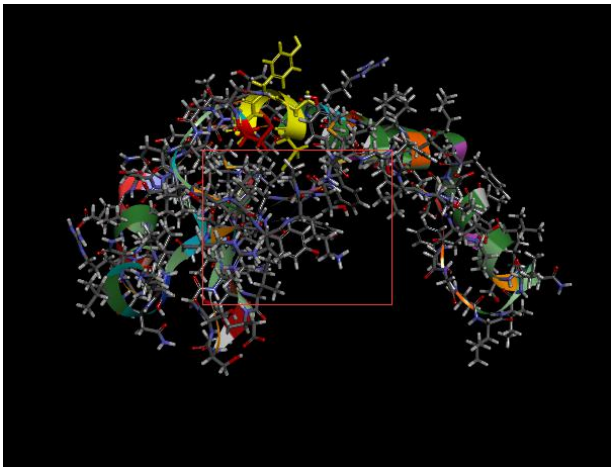

Figure 13. The spatial structure of E protein was like a shrimp. The red box marked the structure as the gathered clamp. It protects the shrimp heads above --- the part of heme linked area (yellow and red marked structures,"CAYCC"). Behind the clamp is an active cavity. The yellow marked structure was the heme carbon linked region (Cys40-Cys43), and the red marked structure was the heme iron linked region (Cys44).

4 Discussion

4.1 Obtaining conserved domains to dissociate iron from the heme through gene recombination

For the most primitive life viruses, it is not so easy to see their role in binding the porphyrin, an ancient compound that widely exists on the earth. It should be noticed that the porphyrin was first found in crude oil and asphalt rock in 1934, and it has unique photoelectronic properties, excellent thermal stability and broad application prospects in materials chemistry, medicine, biochemistry, and analytical chemistry. It is also featured with excellent performance in two-photon absorption, fluorescence effect, energy transfer, and other aspects. Of course, 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. 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. Depending on the research results in this study, the novel coronavirus was a life form

dependent on the porphyrin. Therefore, it was concluded that the novel coronavirus originated from an ancient virus that evolved over countless generations in bats.

This study discovered the heme linked sites of the ORF3a protein belonging to the SARS coronavirus. It was also indirectly confirmed that the novel coronavirus originated from an ancient virus. As we all know, Hydrogenobaculum was acidophilic bacteria in hot springs of karst caves. The overflowing hot spring water formed a river in the cave, which was the water source of various creatures in the cave. There are burrowing animals in the cave, such as civet cats and pangolins. Bats hang upside down, and mainly feed on mosquitoes, while bat feces accumulate on the ground, ferment and acidify.

There may be an ancient virus that obtained the heme linked sites of cytochrome C oxidoreductases in Hydrogenobaculum through genetic recombination. The ancient recombinant virus was released from Hydrogenobaculum and attached to the bat food. Then, bats eat food and are infected with the viruses. It should be noted that there are many viruses in bats. Through the genetic recombination of bats, the ORF3 protein of the coronavirus in the bat gained the heme linked sites on the ancient virus, cytochrome reductases and EFeB domains. In this way, coronaviruses evolved into SARS coronaviruses in bats whose feces with the SARS coronavirus also contaminated the environment of other burrowing animals, such as civet cats and pangolins. Then, the civet cats and pangolins infected with the SARS coronavirus polluted humans' living environment. In this way, humans were infected with the SARS coronavirus. The novel coronavirus SARS-CoV-2 is also a SARS coronavirus, and it may be a new breed of reorganization and evolution in bats.

4.2 heme linked sites of structural protein E may be associated with high viral infection

From the study, it was found that ORF3a, E and N had heme linked sites, while other structural proteins had no heme linked sites. Therefore, in this article, this was not considered accidental. However, it failed to find the domain where the iron of the structural protein was dissociated. In this case, it is unclear whether E and N can dissociate heme iron. Therefore, E and N in contact with the heme, should not form stable compounds. In addition to that, the linked site of heme iron of E was Cys, the sulfur-containing amino acid. Sulfur and heme iron could constitute an active region—like the cytochrome d oxidase of Hydrogenobaculum sp. SN. E protein had the conserved domain, which can realize the functions of electron transport. Cytochrome d oxidases usually combined with cytochrome b oxidases to form a complex. Cytochrome oxidase b catalyzed the synthesis and degradation of endogenous substances (such as fatty acids), which were the main components of cell membranes. It was speculated that the heme linked sites of the E might be related to the role of viruses and cell membranes.

The further evolution of the novel coronavirus also displays some paradoxical characteristics. The current theory reveals that the novel coronavirus binds to the human ACE2 receptor through a spike protein. The novel coronavirus enters human cells in the form of phagocytosis. The novel coronavirus pneumonia is highly contagious. 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 since the virus can live in body fluids. In such media, the porphyrin is a prevalent substance, and porphyrin compounds are a class of nitrogen-containing polymers. Existing studies have shown that they have a strong ability to locate and penetrate cell membranes. Therefore, the novel coronavirus may also directly penetrate the human cell membrane through linking heme. In addition, E was related to the

process of viral infection and had a special spatial structure, where the existing drugs could hardly reach the heme iron linked site of E. Therefore, it was difficult to effectively inhibit the activity of E protein. In this case, it can be seen that the infection is robust.

4.3 Immune cells are infected and secrete antibodies and viral proteins

Some theories hold that an immune response occurs in the body after a patient becomes ill. Some patients develop immune antibodies after recovery. According to this study, the virus protein could bind to porphyrins. However, 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. Compared with patients who had 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 factor of too much mucin protein that could turn loosely connected cells into tightly adhered cells and increase lubrication between cells. It is suggested that 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 are mainly used for virus assembly. Therefore, it is hard to find noticeable virus inclusions in tissue cells of the dissected patient.

Immune cells, such as plasma cells, are also known as effector B cells. Plasma cells are mainly observed in the connective tissue of the intrinsic membrane in both the digestive tract and the respiratory tract, and they are also antibody-secreting cells. This kind of cell has the function of synthesizing and storing antibodies, namely immunoglobulins, and participates in humoral immune responses. Depending on the source of antibody production, antibodies include natural antibodies, such as anti-A and anti-B antibodies in blood group ABO. According to the agglutination state of antigen reaction, antibodies are divided into complete antibody IgM and incomplete antibody IgG. The detection of IgM and IgG in the blood can determine whether the human body is infected with the virus. There is a great amount of IgM in the blood of suspected patients with the novel coronavirus pneumonia. Through treatment, the amount of IgM in the patient is reduced, and the amount of IgG is raised, indicating that their body has resistance and immunity. There are reports that plasma cells also have ACE2 receptors, or in other words, it could be a Spike-ACE2 infection pathway. As the reports show that the spleen, bone marrow, and lymph nodes of severe patients are also significantly damaged, we speculate that plasma cells are also closely related to the infection and recovery of patients with the coronavirus.

Plasma cells can secrete various antibodies, which also explains the release of viral proteins in the body. Viral proteins orf1ab, ORF3a, and ORF10 are synthesized in cells and attack hemoglobin and the heme outside the cells. Viral proteins were possible outside the cell through secreted protein pathways, while secreted proteins mainly consist of digestive enzymes, antibodies, and some hormones. Based on the above viewpoint that disease infection is linked to plasma cells, we believed that viral proteins were secreted mainly from the inside to the outside of the cell through the secretory pathway of antibodies. One possible mechanism is that after the plasma cell was infected, the viral transcription and translation processes were launched, and then viral proteins such as orf1ab, ORF3a, and ORF10 were secreted out of the cell. However, it is not clear whether the viral proteins are secreted outside the cell by binding to blood group antibodies.

We planned to simulate this mechanism, but the amount of calculation was extremely large. After we input the "blood antibody" in the search box of the PDB database, the web page showed nearly 160,000 records, when nearly 47,000 records were associated with humans. Besides, the molecular docking simulation of antibodies and proteins such as orf1ab is the docking of proteins, and the

calculation process is complicated to the greatest extent. Therefore, it failed to simulate this mechanism, but it was suggested that other laboratories can use supercomputers to simulate this mechanism.

4.4 The manner of virus proteins attached and attacked hemoglobin

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 the rupture of red blood cell membranes and the dissolution of the matrix. Alternatively, 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 specific hemolysis brought about by the antigen-antibody reaction. Physical, chemical or biological factors generate non-specific hemolysis. After hemolysis of red blood cells, viral proteins may attach to hemoglobin. Considering that some researchers have calculated that people with some O types of blood are not easily infected with COVID-19, it is inferred that immunohemolysis may be the first method for viral proteins to attach to hemoglobin. The virus has a Spike-CD147 pathway, which may be one way for the virus to infect red blood cells. The virus or virus protein complex enters into the red blood cell by thrombin-sensitive protein and dynein system protein, which may be the second method for viral proteins to attach to hemoglobin. Due to limited computational tools, whether viral proteins attack hemoglobin outside or inside red blood cells can’t be simulated.

4.5 Higher hemoglobin causes higher morbidity

The novel coronavirus pneumonia might be closely related to abnormal hemoglobin metabolism in humans. The number of hemoglobin is a significant blood biochemical indicator, and the content varies with 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, most patients with the novel coronavirus pneumonia are the middle-aged and older adults, while many of these patients have underlying diseases such as diabetes. Diabetic patients have higher glycated hemoglobin which is deoxyhemoglobin and is also a combination of hemoglobin and blood glucose, which is another reason for the high infection rate for the elder people.

This study has confirmed that orf1ab, ORF3a, and ORF10 could coordinately attack the heme on the beta chain of hemoglobin. Both oxygenated hemoglobin and deoxygenated hemoglobin are attacked, but the latter is more attacked by the virus. During the attack, the positions of orf1ab, ORF3a, and ORF10 are slightly different, which shows that the higher the hemoglobin content, the higher the risk of disease. However, it is not sure that the disease rate incited by abnormal hemoglobin (structural) is relatively low. The hemoglobin of patients and rehabilitees could be detected for further research and treatment.

4.6 Interfering with the normal heme anabolic pathway

This article held that the virus directly interfered with the assembly of human hemoglobin. The main reason was that the normal heme was too low. Heme joins in critical biological activities such as regulation of gene expression and protein translation, and the porphyrin is an essential material for the synthesis of the heme. As the existing traces show there is too much free iron in the body of critically ill patients, it could be that the 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 the heme and porphyrins in patients with

porphyria is the same as that in healthy people. If there is an abnormal structure, it is unobvious whether this porphyrin can bind to a viral protein to form a complex, or whether a viral protein can attack this heme. It could be proved by clinical and experimental research.

4.7 Novel coronavirus pneumonia may be a particular type of lung cancer

This study found that ORF3a protein had an incomplete respiratory chain domain. Screening and verification of ORF3a inhibitor drugs showed that nucleotides, nucleoside analogs, or fructose related to the growth of cancer cells could bind to the active region of hematoporphyrin. Nucleotides similar to NADPH and nucleoside analogs similar to reducivir bound to the respiratory chain of ORF3a. Besides, we noticed that cancer cells also had an incomplete respiratory chain system. This type of anaerobic energy metabolism was characterized by high sugar intake, active glycolysis, and high metabolite lactic acid content (Warburg effect). We speculated that the SARS-CoV-2 viral proteins was similar to the respiratory chain system of cancer cells. Therefore, it was not surprising that some anti-cancer drugs could bound ORF3a protein.

Severe novel coronavirus pneumonia was similar to advanced lung cancer. Both types of severely ill patients had severe intravascular coagulation reactions. All tissues and organs were irreversibly damaged, and there were multiple organ infections and complications. For example, pulmonary fibrosis caused a lot of mucus in the lungs and tissue necrosis. There was myocarditis with pericardial effusion in the heart. There was acute tubular necrosis in the kidney. The cells of the liver degenerated and accompanied by disease of the leaflets. Bleeding under the skin, bleeding spots, ecchymosis appeared on the surface. Headache, hemiplegia, epilepsy would be in both types of patients. Some problems occurred in the immune mechanism. There was a cytokine storm in two kinds of patients who deteriorate. Also, the genetic signals were abnormal, such as multinucleated cells. **The only thing unclear was whether COVID-19 patients had genetic mutation** (such as p53, cytochrome C, and peroxidase) **through the action of viral proteins and whether this mutation caused a similar "cancer" feature.**

The primary pathogenic mechanism of this virus was inferred depending on the computational simulation and discussion analysis in this study. The virus might first infect cells with ACE2 receptors, including immune cells, which produced antibodies and viral proteins. The Spike-CD147 pathway infected red blood cells or red blood cells generated immune hemolysis. Hemoglobin was attached and then attacked. The attack produced toxicity and inflammation due to derivatives. The virus also captured porphyrins and inhibited heme metabolism, similarly carcinogenesis. Then, organs had complications, when cytokine storm caused multiple organ failure.

Therefore, we speculated that the novel coronavirus pneumonia might be a particular type of lung cancer. For this disease, the sooner it is prevented and treated, the sooner it may be cured. Without timely isolation, intervention, and treatment, there was a high possibility of infection and rapid deterioration.

5. Conclusion

Since the outbreak of the epidemic, it is of great scientific significance to use bioinformatics to analyze novel coronavirus proteins' roles. In this study, conserved domain analysis, homology modeling, and molecular docking was made to compare the biological functions of specific proteins belonging to the novel coronavirus.

We found E, N, and ORF3a have heme linked motif by conserved domain analysis. Arg134 of

ORF3a, Cys44 of E, Ile304 of N were the heme-iron linked sites, respectively. ORF3a also possessed the conserved domains of human cytochrome C reductases and bacterial EFeB protein. These three domains were highly overlapping so that ORF3a could dissociate the iron of heme to form porphyrin. Heme linked sites of E protein may be relevant to the high infectivity, and the role of heme linked sites of N protein may link to the virus replication. The heme linked motif may originate from cytochrome C oxidase monoheme subunit / FixO [Hydrogenobaculum sp. HO]. SARs virus may be obtained these conserved domains through gene recombination in bat body.

Homology modeling methods obtained the viral protein molecules. Molecular docking technology was adopted to analyze the viral proteins' role. The study results showed orf1ab, ORF3a, and ORF10 proteins could coordinately attack 1-beta chain of hemoglobin. Deoxyhemoglobin was more vulnerable to virus attacks than oxidized hemoglobin. The attack will lead to less hemoglobin to transport oxygen and carbon dioxide. ORF3a was specific and would not attack blue blood protein, normal cytochrome C, and peroxidase. But ORF3a can attack the mutant cytochrome C. Cytochrome C is involved in the respiratory chain process of cells and most distributed in the cardiomyocytes. Therefore, if the structure of ORF3a (such as mutation) or hemoglobin (such as Genotype) changes, it may get uncertain results: either hemoglobin is more likely to be attacked; or hemoglobin is less likely to be attacked. At the same time, some viral proteins could combine with the porphyrin to form a complex. As the porphyrin complexes of the virus produced in the human body, the mechanism also may interfere with the normal heme anabolic pathway of the human body, and they made a wide range of infection and disease.

At present, patients with the novel coronavirus pneumonia almost face the situation that no specific drugs are available. Based on the small molecule drug library, drugbank, we searched for drugs bound to the heme-iron linked site of viral proteins by molecular docking. The drugs with the highest binding capacity to ORF3a were Lauryl alcohol diphosphonic acid, (2,2-diphosphonoethyl) dimethylphosphonium, Isopentenyl pyrophosphate, Nicotinamide adenine dinucleotide phosphate, Ceftaroline fosamil, Lusutrombopag,Citicoline, Farnesyl thiopyrophosphate, Psilocybin, Dimethylallyl Diphosphate, Amifostine, 6-Aminohexyl-uridine-C1,5'-diphosphate etc.. Remdesivir was relatively more obvious than Hydroxychloroquine and Chloroquine in terms of the binding capacity of ORF3a. Still, the combined role of three drugs to ORF3a was lower. The drugs with highest binding capacity to nucleocapsid phosphoproteinis were RPR128515, Masitinib, Ziresovir, Sotrastaurin and 3-Aminomethyl-Pyridinium-Adenine-Dinucleotide, N-[3-(aminomethyl)benzyl]acetamidine, Acepromazine, Bafetinib, 6-amino-4-(2-phenylethyl)-1,7-dihydro-8H-imidazo[4,5-g]quinazolin-8-one, Piritrexim, 3-(1H-indol-3-yl)-4-(1-{2-[(2S)-1-methylpyrrolidinyl]ethyl}-1H-indol-3-yl)-1H-pyrrole-2,5-dione, Periciazine etc. .Unfortunately, no drug could bind to the heme iron linked site of E. Confusingly, some of the drugs in the above list are anti-cancer drugs. Besides, these higher binding energies may prevent all screened drugs from binding firmly to viral proteins. Finally, **it is worth noting that the screened drugs could dock with specific viral proteins, but their inhibitory effects on ORF3a and N protein are still unclear because of no clinical data.**

This theory is only for academic discussion and needed to be verified by other experiments. Please consult a qualified doctor for treatment details. Due to the toxicity and side effects of drugs, do not use medicines yourself. We expect these discoveries to bring more ideas to people to relieve patients' symptoms and save more lives.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets and results supporting the conclusions of this article are available at https://pan.baidu.com/s/1WF9D0AyZYY3tEHBIXnvFSg , code: 5z4x.
Or: https://mega.nz/folder/4yI31K7D#4hLWJcYkrX5RFRdRaK-D0A

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by a grant from the National Natural Science Foundation for the Talent Introduction Project of Sichuan University of Science and Engineering (award number: 2018RCL20, grant recipient: WZL).

Author's contribution

Funding was obtained by WZL. Besides, design, analysis and writing are finished by WZL, while data curation and manuscript check are undertaken by HLL. Both authors have read and agreed to the published version of the manuscript.

Acknowledgements

Thanks readers for free review and suggestions.

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 Song, Y. *et al.* SARS-CoV-2 induced diarrhoea as onset symptom in patient with COVID-19. *Gut* (2020).

5 Yao, X. *et al.* A pathological report of three COVID-19 cases by minimally invasive autopsies. *Zhonghua bing li xue za zhi= Chinese journal of pathology* **49**, E009-E009 (2020).

6 Dolhnikoff, M. *et al.* Pathological evidence of pulmonary thrombotic phenomena in severe COVID-19. *Journal of Thrombosis and Haemostasis* (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 Rothe, C. *et al.* Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine* (2020).

20 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).

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

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

23 Pan, F. *et al.* Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*, 200370 (2020).

24 Poggiali, E. *et al.* Can lung US help critical care clinicians in the early diagnosis of novel coronavirus (COVID-19) Pneumonia? *Radiology*, 200847 (2020).

25 Lei, J., Li, J., Li, X. & Qi, X. CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology* **295**, 18-18 (2020).

26 Hong, X., Xiong, J., Feng, Z. & Shi, Y. Extracorporeal membrane oxygenation (ECMO): does it have a role in the treatment of severe COVID-19? *International Journal of Infectious Diseases* (2020).

27 Ñamendys-Silva, S. A. ECMO for ARDS due to COVID-19. *Heart & Lung: The Journal of Cardiopulmonary and Acute Care* (2020).

28 Henry, B. M. & Lippi, G. Poor survival with extracorporeal membrane oxygenation in acute respiratory distress syndrome (ARDS) due to coronavirus disease 2019 (COVID-19): Pooled analysis of early reports. *Journal of Critical Care* (2020).

29 Zhang, W. *et al.* The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The experience of clinical immunologists from China. *Clinical Immunology*, 108393 (2020).

30 McGonagle, D., Sharif, K., O'Regan, A. & Bridgewood, C. Interleukin-6 use in COVID-19 pneumonia related macrophage activation syndrome. *Autoimmunity reviews*, 102537 (2020).

31 Leichtman, D. A. & Brewer, G. J. Elevated plasma levels of fibrinopeptide A during sickle cell anemia pain crisis–evidence for intravascular coagulation. *American journal of hematology* **5**, 183-190 (1978).

32 Rojanasthien, S., Surakamolleart, V., Boonpucknavig, S. & Isarangkura, P. Hematological and coagulation studies in malaria. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet* **75**, 190-194 (1992).

33 Ogata, T. *et al.* Cerebral venous thrombosis associated with iron deficiency anemia. *Journal of Stroke and Cerebrovascular Diseases* **17**, 426-428 (2008).

34 Alder, L. & Tambe, A. in *StatPearls [Internet]* (StatPearls Publishing, 2019).

35 Gao, Y. *et al.* Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *Journal of Medical Virology* (2020).

36 Tang, N., Li, D., Wang, X. & Sun, Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *Journal of Thrombosis and Haemostasis* (2020).

37 Wang, X. *et al.* SARS-CoV-2 infects T lymphocytes through its spike protein-mediated membrane fusion. *Cellular & Molecular Immunology*, 1-3 (2020).

38 Wang, F. *et al.* Characteristics of peripheral lymphocyte subset alteration in COVID-19 pneumonia. *The Journal of infectious diseases* (2020).

39 Borghetti, A., Ciccullo, A., Visconti, E., Tamburrini, E. & Di Giambenedetto, S. COVID-19 diagnosis does not rule out other concomitant diseases. *European Journal of Clinical Investigation*, e13241.

40 Li, S.-r., Tang, Z.-j., Li, Z.-h. & Liu, X. Searching therapeutic strategy of new coronavirus pneumonia from angiotensin-converting enzyme 2: the target of COVID-19 and SARS-CoV. *European Journal of Clinical Microbiology & Infectious Diseases*, 1 (2020).

41 Luo, W. *et al.* Clinical pathology of critical patient with novel coronavirus pneumonia (COVID-19). *Pathology & Pathobiology* **2020020407** (2020).

42 Emmons, R., Oshiro, L., Johnson, H. & Lennette, E. Intra-erythrocytic location of Colorado tick fever virus. *Journal of General Virology* **17**, 185-195 (1972).

43 Mitra, A. *et al.* Leukoerythroblastic reaction in a patient with COVID-19 infection. *American Journal of Hematology* (2020).

44 Bhardwaj, K., Liu, P., Leibowitz, J. L. & Kao, C. C. The coronavirus endoribonuclease Nsp15 interacts with retinoblastoma tumor suppressor protein. *Journal of virology* **86**, 4294-4304 (2012).

45 Spring, F. A. *et al.* The Oka blood group antigen is a marker for the M6 leukocyte activation antigen, the human homolog of OX-47 antigen, basigin and neurothelin, an immunoglobulin superfamily molecule that is widely expressed in human cells and tissues. *European journal of immunology* **27**, 891-897 (1997).

46 Szempruch, A. J. *et al.* Extracellular vesicles from Trypanosoma brucei mediate virulence factor transfer and cause host anemia. *Cell* **164**, 246-257 (2016).

47 Howell, S. A. *et al.* A single malaria merozoite serine protease mediates shedding of multiple surface proteins by juxtamembrane cleavage. *Journal of Biological Chemistry* **278**, 23890-23898 (2003).

48 Mitchell, G., Thomas, A., Margos, G., Dluzewski, A. & Bannister, L. Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. *Infection and immunity* **72**, 154-158 (2004).

49 Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature* **480**, 534-537 (2011).

50 Baum, J. *et al.* A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *Journal of Biological Chemistry* **281**, 5197-5208 (2006).

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

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

53 Zhang, H. *et al.* Consensus on pre-examination and triage in clinic of dermatology during outbreak of COVID-19 from Chinese experts. *International Journal of Dermatology and Venereology* (2020).

54 Recalcati, S. Cutaneous manifestations in COVID-19: a first perspective. *Journal of the European Academy of Dermatology and Venereology* (2020).

55 Liang, W. *et al.* Diarrhoea may be underestimated: a missing link in 2019 novel coronavirus. *Gut* (2020).

56 Chan, J. F.-W. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **395**, 514-523 (2020).

57 Hassan, S., Sheikh, F. N., Jamal, S., Ezeh, J. K. & Akhtar, A. Coronavirus (COVID-19): a review of clinical features, diagnosis, and treatment. *Cureus* **12** (2020).

58 Lippi, G., South, A. M. & Henry, B. M. ANNALS EXPRESS: Electrolyte Imbalances in Patients with Severe Coronavirus Disease 2019 (COVID-19). *Annals of Clinical Biochemistry*, 0004563220922255 (2020).

59 Xiong, T.-Y., Redwood, S., Prendergast, B. & Chen, M. Coronaviruses and the cardiovascular system: acute and long-term implications. *European Heart Journal* (2020).

60 Zhang, C., Shi, L. & Wang, F.-S. Liver injury in COVID-19: management and challenges. *The Lancet Gastroenterology & Hepatology* (2020).

61 Naicker, S. *et al.* The Novel Coronavirus 2019 epidemic and kidneys. *Kidney International* (2020).

62 Wu, Y. *et al.* Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain, Behavior, and Immunity* (2020).

63 Mungroo, M. R., Khan, N. A. & Siddiqui, R. Novel Coronavirus: Current Understanding of Clinical Features, Diagnosis, Pathogenesis, and Treatment Options. *Pathogens* **9**, 297 (2020).

64 Lu, L. *et al.* New-onset acute symptomatic seizure and risk factors in Corona Virus Disease 2019: A Retrospective Multicenter Study. *Epilepsia* (2020).

65 Wang, H.-Y. *et al.* Potential neurological symptoms of COVID-19. *Therapeutic Advances in Neurological Disorders* **13**, 1756286420917830 (2020).

66 Moriguchi, T. *et al.* A first Case of Meningitis/Encephalitis associated with SARS-Coronavirus-2. *International Journal of Infectious Diseases* (2020).

67 Liu, K., Pan, M., Xiao, Z. & Xu, X. Neurological manifestations of the coronavirus (SARS-CoV-2) pandemic 2019–2020. *Journal of Neurology, Neurosurgery & Psychiatry* (2020).

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

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

70 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).

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

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

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

74 Liu, X., Shi, Y., Deng, Y. & Dai, R. Using molecular docking analysis to discovery dregea sinensis hemsl. potential mechanism of anticancer, antidepression, and immunoregulation. *Pharmacognosy magazine* **13**, 358 (2017).

75 Kandhaswamy, A., Saravanan Rangan, I. A. & KS, M. Synthesis, in silico docking and admet studies of arylacetic acid derivatives as prostaglandin endoperoxide H synthase-2 inhibitors. *SYNTHESIS* **10** (2017).

76 Zhang, Q. *et al.* In silico screening of anti-inflammatory constituents with good drug-like properties from twigs of Cinnamomum cassia based on molecular docking and network pharmacology. *Tropical Journal of Pharmaceutical Research* **18** (2019).