The unignorable N protein of COVID-19

Zehua Zeng
University of Science and Technology Beijing
Zhi Luo
University of Science and Technology Beijing
Yubang Shan
University of Science and Technology Beijing
Jingjie Yang
University of Science and Technology Beijing
Chuan Chen
University of Science and Technology Beijing
Luna Wang
University of Science and Technology Beijing
Hongwu Du (hongwudu@ustb.edu.cn)
University of Science and Technology Beijing

Research Article

Keywords: COVID-19, SARS-CoV-2, clinical manifestations, nucleocapsid protein

Posted Date: June 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34518/v1

Abstract

With more and more people infected with COVID-19, it was found that the SARS-CoV-2 virus was quite different from SARS-CoV. Most researches have focused on ACE2 and S protein, however, the known variations of these proteins are not enough to explain why there were so many different clinical manifestations between patients infected with such two viruses. Here, the N protein of the two coronaviruses was parallelly analyzed. Through the analysis of N protein structure, protein conserved binding domain, binding site and ligand, and CTL epitope, it was found that N protein may have a unique expression profile in the SARS-CoV-2. For example, mirror structure and RNA binding tendency.

Introduction

In December 2019, a new public health event broke out in Wuhan, China, and swept the globe for the next three months. Scientists identified the cause of the disease as a pneumonia infection caused by a novel coronavirus (1). So far, WHO has announced that COVID–19 is a pandemic. Besides, most of the victims died from cytokine storms caused by the autoimmune response (2, 3). Finding the cause of the cytokine storm is particularly essential. At present, the general research direction of academia is spike protein (S protein). Same as other strains, S protein is the key for this novel coronavirus to enter human cells (4), its unique receptor angiotensin-converting enzyme 2 (ACE2) is distributed in the epithelial cells of the lung, the epithelial cells of the small intestine and the eyelids (5–7). The infection in these cells will lead to an inflammatory response. Different from SARS-associated coronavirus (SARS-CoV), persons infected with SARS-CoV–2 displayed mild at an early stage, but when their symptoms got worse, this process was relatively rapid (8, 9). This novel coronavirus has many unknown features, and further researches are needed.

Currently, most researches have focused on ACE2 and S protein of SARS-CoV–2. Still, the known variations of these proteins are not enough to explain why there were many different clinical manifestations between patients infected with SARS-CoV and SARS-CoV–2. In this study, we hope to figure out this issue from another perspective, to suggest a new research strategy focus on the nucleocapsid protein (N protein, NP). The N protein is a structural protein whose primary function is to recognize a stretch of RNA that serves as a packaging signal and leads to the formation of the ribonucleoprotein (RNP) complex during assembly (10). It is also involved in the immune response, and the formation of the RNP is vital for maintaining the RNA in an ordered conformation suitable for replication and transcription of the viral genome. In a retrospective study of SARS-CoV, we found that N protein plays an important role in the coronavirus like inducing a human immune response (11). In addition to assembling new viral molecules, one of its important functions is to inhibit IFN in infected cells (12), and such mechanism is used by many different viruses, including SARS-CoV (12).

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction(13). I-TASSER (as ‘Zhang-Server’) was ranked as the No one server for protein structure prediction in recent community-wide CASP7-CASP13 experiments. After the outbreak of COVID–

19, I-TASSER provided 24 predictive models of potential open reading frames for SARS-CoV–2 (14). We hope to find out more information about the characteristics of COVID–19, such as from the prediction of N protein structure.

In order to better understand the N protein, we analyzed the variation of the virus, the conserved domain of the protein, the spatial structure of the N protein, and predicted its binding sites and possible binding ligands, and made some inferences about the new crown. There have been few reports on the N protein of SARS-CoV–2, so this study hopes to call on people to invest more research into N protein through this study in that as a more conservative antigen. Besides, as a conservative antigen, N protein is also of great significance for vaccine development.

Materials And Methods

Sequence Data Collection and Variation Analysis

The Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu–1, complete genome（Genbank： MN908947.3）and Severe acute respiratory syndrome-related coronavirus isolates Tor2, complete genome （Genbank：AY274119.3）. 958 nucleotide sequences of COVID 19 were acquired from GISAID (due on March 20, 2020). 851 out of 958 sequences were complete and were conducted the following analysis (the detailed information of 851 sequences were demonstrated in supplement table 1). Every 100 sequences were merged in 1 file and aligned by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) due to the limitation of file size. Mega X was used to translate DNA sequences into amino acid sequences and observe the mutation of the N protein.

Similarity calculation by Basic Local Alignment Search Tool

To compare the SARS-CoV and SARS-CoV–2 N protein similarity, we use the blast(33) online tools (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for SARS-CoV (Genbank: AY274119.3) and SARS-CoV–2 (Genbank: MN908947.3) to calculate. Blastp is selected for the algorithm.

Conservative prediction by Conserved Domains Database

This study were using NCBI CDD(34) online tools (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), for SARS-COV (Genbank: AY274119.3) and SARS-CoV–2 (Genbank: MN908947.3) nucleotide sequence of protein structure domain conservative projections, respectively. Select CDD v3.17–52910 PSSMs for search database; Expect the Value threshold = 0.010000; Apply low complexity filter, Force lives to search and Rescue borderline hits, Suppress weak overlapping hits are not checked, and Composition -based statistics adjustment check, the Maximum number of hits is set to 500, Result mode is set to the Concise.

Prediction of the spatial structure of N proteins

Phyre2: The N protein amino acid sequence was submitted to the online tools phyre2(35) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id = index), which uses an existing model to predict protein sequences.

Prediction of N protein-ligand and binding site

COACH: SARS-CoV–2 (N protein in Genbank: MN908947.3) to predict the PDB file upload to the online tool COACH(36) (https://zhanglab.ccmb.med.umich.edu/COACH/) forecast operation. COACH is a meta-server method for predicting protein-ligand binding sites.

COFACTOR: SARS-CoV–2 protein (N protein in Genbank: MN908947.3) to predict the PDB file upload to the online tool COFACTOR(37) (https://zhanglab.ccmb.med.umich.edu/COFACTOR/) forecast operation. COFACTOR is a structure-based, sequencing and protein-protein interaction (PPI) approach for functional annotation of proteins in molecular biology.

Comparison of N protein structure similarity

For quantitative comparison of SARS - COV (Genbank: AY274119.3) and SARS-CoV–2 (N in Genbank: MN908947.3) protein structure similarity, we adopt TM-score(38) (https://zhanglab.ccmb.med.umich.edu/TM-score/) to measure the similarity of two protein structure. The TM fraction is a measure of the similarity between two protein structures. A score below 0.17 corresponds to a randomly selected independent protein, while a structure above 0.5 is assumed to be generally the same in SCOP/CATH. Input the protein PDB file of SARS-CoV in Input Structure 1, Input the protein PDB file of SARS-CoV–2 in Input Structure 2, parameters These are complex structures are not selected, and then run the comparison calculation.

Calculation of the tendency of N protein residues binding proteins

In order to judge the COACH and the COFACTOR to predict whether the residue of binding sites has more substantial combination orientation, thus to predict the ligand binding, we choose the SCRIBER(39) (http://biomine.cs.vcu.edu/servers/SCRIBER/) to generate each residue number score, the score is used to quantify the residue combined with protein assumed tendency, tend to value, the greater the show that, the higher the possibility of the combination.

The interaction tendency and co-expression of the auxiliary protein RNA and N protein in human tissues were calculated

To study the role of 5 kinds of auxiliary Protein RNA in the N Protein, through the online server express (http://service.tartaglialab.com/page/catrapid_express_group) are calculated by the nucleotide sequence. The N Protein interaction tendentiousness and objectivity, the Protein sequence input N Protein amino acid sequence of the RNA sequence points five input 5 kinds of auxiliary Protein sequences of nucleotide.The results were shown by heat map. The darker the red, the higher the binding tendency.

Results

Analysis of the variation for N protein of SARS-CoV–2

It was found that 24.91% samples (212 out of 851) have changed tyrosine to histidine (PVYLL→PVHLL). These 212 mutated sequences came from 15 countries, among which there were 88 from the USA, 82 from China, 10 from South Korea, 6 from Chile, 5 from Australia, 4 from the Netherlands, 4 from Canada, 3 from Japan. In the UK, Spain and France, there were 2 cases respectively, and in German, India and Belgium, there was only 1 case in each. Arginine and glycine of 124 samples were mutated to lysine and arginine respectively (SSRGTS→SSKRTS)(Fig.1) which accounted for 14.57 %. Among the 124 mutated sequences which came from 11 countries, where there were 66 from the Netherlands, 21 from Switzerland, 16 from the UK, 8 from Brazil, 4 from Finland, 2 from Belgium. As for Portugal, Nigeria, Peru, Mexico and Chile, there was 1 case in each country (Fig 1).

The similarity of SARS-CoV–2 and SARS-CoV in N protein

10 ORFs (open reading frame) were found in the genome sequence obtained by SARS-CoV–2. Compared with SARS-CoV, 4 ORFs are missing, and missed fragments were located before E protein, and before and after N protein, respectively. The ORF regions and the conserved domains of SARS-CoV–2 and SARS-CoV were paralleled compared. ORF 7B shares a 90.5% similarity with the NP in SARS-CoV.

When analyzing the conserved binding domain of SARS-CoV–2, it was found that a total of six conserved binding domains were predicted in the nucleotide sequence near the N protein, of which only Corona_M was specific hits and the others were superfamilies. Moreover, these superfamily members include N proteins (Fig.2 & Supplement Table 1&2). It was generally believed that protein superfamilies are the same modular structures that perform similar functions. So theoretically, the N protein of SARS-CoV–2 also has some of the properties of coronavirus N protein, but it may only be similar, and quite different in effect. Further experimental evidence in the S protein also demonstrated a similar effect among members of the superfamily. The S protein receptor of SARS-CoV–2 is ACE2, which was the same as that of another coronavirus, SARS-CoV, but with different properties. For example, the S protein of SARS-CoV–2 showed stronger binding force (4).

To further research the similarity of the N protein of SARS-CoV–2 and SARS-CoV on ORF, the spatial structure of the two was compared. Homologous modelling of the N protein of SARS-CoV–2 was conducted with the help of Phyre2. The N protein of SARS-CoV was selected from PDB dataset (ID:1ssk)), and quantitatively calculated the similarity between the two proteins with TM-score. In Fig.3a&3b&3c, SARS-CoV and SARS-CoV–2 showed approximately symmetrical structures, and TM-score was 0.1479 (TM-score <0.17 is generally considered to mean that there is no similarity between the two proteins), suggesting that the molecules with an immune response to the N protein of SARS-CoV–2 may be completely different from SARS-CoV.

In combination with all our analyses at proteins, this study aims at the similarity of N protein between SARS-CoV–2 and SARS-CoV, and the specific situation of their N protein fitting. The above analysis indicated that the N protein of SARS-CoV–2 would have a completely new effect on the host.

The mirror image of the N protein

By comparing the N protein prediction model of SARS-CoV–2 with the actual protein model of SARS-COV, it was found that the two have similar mirror image structure(Fig.3a), which may be the reason why SARS-CoV–2 is more infectious. To further confirm this mirroring conclusion, we used TM-score calculation to compare the actual N protein structure of SARS-CoV–2 (PDB: 6m3m) with that of SRAS-COV–2 (PDB:1ssk), the TM-score was calculated as 0.1350, with a low overall similarity, which was consistent with the conclusion of the similarity of the prediction model. But the exciting thing is, it was also a mirror structure (Fig.3d) compared with SARS-CoV. Therefore, the similarity of N protein in SARS-CoV and SARS-CoV–2 was probably very high, except that it was presented as a mirror image.

The binding site and ligand prediction of N protein

In N protein analysis, a total of 8 potential ligands were predicted by the COACH method, which was ranked by c-score as FES, HEC, HEM, U5P, C8E, CA, HIS and BLQ. A total of 5 potential ligands were predicted by the tm-site method, which was ranked by c-score as HEM, U5P, O4B, HIS and C8E. Four potential ligands were obtained by the s-site way, which was ranked by c-score as HEC, GTX, UUU and BLO. A total of three potential ligands were received by the COFACTOR method, which was listed by c-score as ZLD and FES, and the corresponding binding residues were shown in Supplement Table 3. Then, we calculated the binding tendency of each residue in N protein by SCRIBER (Supplement Table 4). Based on all the ligand-binding sites mentioned above, only 306 and 307 of ZLD displayed the binding tendency in SCRIBER's calculation, indicating that ZLD(N-{[(5S) –4-morpholin - 4-ylphenyl) –2-oxo–1, 3-oxazolidin–5-yl] Methylyl} acetamide) (Fig.4) is likely to be a target molecule for N protein binding. This requires further experimental verification.

Comparison of CTL epitopes of N proteins

It was found that the similar residues of SARS-CoV–2 and SARS-CoV were only N49–51, N83–87, N89–94 (Fig.3b&3c). A study using PBMCs from SARS patients two years after infection showed that the main dominant antigen location of N protein was in the c-terminal region (amino acid 331–362)(15). Using the same method, another group found two potential CTL (cytotoxic T lymphocyte) epitopes at sites N211–235 and N330–354(16), and the corresponding dominant response was found in patients who recovered 6 years after infection (N216–230)(17). We found that the spatial structure of amino acid sequences such as N331–362 and N211–235 of SARS-CoV–2 was utterly different from that of SARS-CoV; therefore the CTL epitopes of N protein antigens were completely different from those of SARS, which may be one of the reasons for the different T cell immune responses of the two viruses.

The binding tendency of auxiliary protein RNA to N protein

It has been reported that the N protein of SARS-CoV tends RNA binding. To study the effect of auxiliary protein RNA on the N protein of SARS-CoV–2, we calculated the binding tendency of RNA and N protein of five extra proteins (Fig.5). For ORF3a and N protein, there are regions with binding trend greater than 3, which was mainly located at the c-terminal of N protein and 400–600 range of ORF3a nucleotides. For ORF6 and N proteins, the binding tendency is also higher than 3, and in particular, almost the whole nucleotide sequence of ORF6 shows a high binding tendency with the c-terminal of N proteins. For ORF7a and N protein, there are regions with binding tendency greater than 3, which was mainly located at the c-terminal of N protein and the 0–100 range of ORF7a nucleotides. For ORF8 and N protein, there are regions with binding tendency greater than 2, which are mainly located at the c-terminal of N protein and the 0–100 region of ORF8. For ORF10 and N protein, there is a region with a binding tendency greater than 3, which is mainly located in the c-terminal of N protein and the 0–40 region of ORF10. This suggests that the helper protein RNA may be able to bind to the N protein to complete at least one biochemical reaction.

Discussion

The COVID–19 outbreak brought the coronavirus back into human society, and our study focused on the previously undiscussed N protein. Previous studies have shown that in SARS, the N protein not only protects the virus's genes but also inhibits IFN secretion(12). Moreover, since N protein does undergo glycosylation, it is easy to be expressed in cells. At the meantime, N protein has high immunogenicity, which makes it is an ideal antigen for vaccine development. The N protein of SARS-CoV can induce specific T-cell response(18, 19), and this phenomenon has also been observed in other coronaviruses(20). However, none of these has been reported for SARS-CoV–2. Besides, the N protein of SARS-CoV has some other functions, such as self-dimerization(21), RNA binding capabilities(22), which can even induce mammalian cell apoptosis under stress, and actin recombination(23). Moreover, N protein is not always used as the nucleocapsid of the virus, and its distribution is found in the cytoplasm and nucleus of cells infected with SARS-CoV(24).

Three primary mutations were observed, and these mutations were distributed worldwide, however, how these mutations affect the N protein was still not clear. Other random mutations were also observed. Some of them presented a regional correlation and are particular cases. Thus, these cases were excepted from the variation analysis.

By analyzing the ORF quantity of SARS-CoV–2 and SARS-CoV, we found that SARS-CoV–2 was 4 ORF less than SARS-CoV, but the total nucleotide sequence length was within 0.5%, which was a weird thing. Compared with SARS-CoV, hCoV had fewer auxiliary proteins but was more infectious. In comparison, the ORF positions of SARS-CoV–2 and SARS-CoV were ORF3B and ORF7B. ORF8B, ORF13 and ORF14 are directly replaced by another new open reading frame, ORF10. In previous reports, ORF3B protein contains two predicted nuclear localization signals and is located in the nucleoli of transfected cells in the absence of any other SARS-CoV protein(25). Also, overexpression of ORF3B may cause cell cycle arrest and apoptosis in the G0/G1 phase(26). This may mean that SARS-CoV–2 lacks this open reading frame, leading to the longer survival time of its host cells. As a result, the proteins in the host cells are used more thoroughly, resulting in more SARS-CoV–2 production, but the cell lyse more slowly, which may be one of the reasons for the more extended incubation period of COVID–19. The lack of ORF7B was reported in SARS, although the corresponding antibody was detected(27). Unfortunately, the expression of ORF7b in host cells has not been confirmed, and a study claimed that after subculture of SARS-CoV, the virus could still infect and replicate after ORF7b loss(28). In addition, reverse genetic studies also showed that ORF7b deletion had no significant effect on virus replication in both cell models and mouse models[26], So maybe there is no ORF7b in SARS-CoV–2, and it has no effect on its replication. In the translation and transcription process of SARS-CoV, ORF8b fuses with ORF8a to form a protein. In SARS-CoV–2, the length of ORF8 is 123, while the total length of ORF8a and ORF8b is 121. After blast analysis of ORF8 of SARS-COV–2 and ORF8 of SARS-CoV, the similarity was only 30.16%. In the late period of SARS, the ORF8 region was absent to different degrees(29). In the latest report on SARS-CoV–2, the ORF8 region is also missing in some region(30). This site suggests that ORF8 may only act as a cofactor in coronavirus and has no adverse effect on the survival of the virus. When analyzing the RNA sequence of ORF8, we found that its nucleotide may bind to the N protein, and more experiments are needed to verify whether it plays a role in weakening viral virulence.

Then we analyse the five kinds of auxiliary SARS-CoV–2 protein RNA will combine with N protein, surprisingly, five types of protein N of RNA and protein have some near the 3 ‘end of the combination of tendency, it may show that SARS-CoV–2 of 5 kinds of proteins in the auxiliary or RNA sequences, can with the N protein, the formation of its possible catalytic N protein.

In the further study of the spatial structure, we found that the SARS - CoV and SARS-CoV–2 TM - score only 0.14, in the sense of TM score, which below 0.17 correspond to randomly chosen unrelated proteins. But for SARS-CoV–2, observe the space structure, can be found that the structure of both the approximate image(Fig.3d), that may mean may be due to the mirror image of the structure of the led to such a low score. And there may not be much difference in function. Studies of chiral molecules suggest

that two mutually symmetric molecules may have a bipolar effect on the same function. Could this mirror structure protect the virus's genes better? As a result, it can reproduce better in cells than SARS-CoV.

On molecular biology, N protein is considered to be a more conserved fragment than S protein and M protein(31). We first compared the differences in N protein binding domains (Fig.2). In SARS-CoV–2, the SARS-CoV binding domains associated with the N protein are all members of the superfamily, so the binding sites of the two proteins in the human body may be completely various, and therefore the cytokines are different. Studies have shown that the N protein of SARS-CoV is not only an essential B-cell immunogen, but also can trigger a broad cellular immune response, and the immune mice have a strong delayed hypersensitivity reaction to the N protein(31). According to existing studies, most newly infected people have mild symptoms, but when they become severe or critically ill, there is a sudden acceleration, which is an significant cause of death in both harsh and critically ill patients(9). This may be closely related to the delayed hypersensitivity of N protein, and we suspect that the symmetrical mirror structure of N protein may make this process faster. Studies on the SARS-CoV's N protein have shown that causes a significant increase in the secretion of IgG–2a, IFN-γ and IL–2 in mice(31). Therefore, when people suffer from COVID–19, the immune system secretes different antibodies to different antigens. Among them, the cytokines produced by the N protein may be relatively slow, but also more destructive. Therefore, the cytokine storm is induced at a breakneck speed and later time, and the patient is thus transferred to the severe disease.

We integrated two methods for predicting protein binding residues, COACH and SCRIBER, and found that only two residues, 306 and 307, had a binding tendency of one (one was considered to have a greater binding tendency, while zero had a lower binding tendency). The corresponding predicted ligand of 306 and 307 binding sites in COACH was ZLD(Fig.5)(Supplement Table 3), a class of drugs used to treat bacterial infections, has been reported to bind to residues of A385 and F386 of the main transporter protein Arcb in escherichia coil(32). Although there is a residual difference between the two groups, the spatial structure around them is similar to that around Arcb's 385 and 386. Its BS score is calculated quantitatively to reach 0.81. It is generally considered that a BS score more massive than 1.0 is considered as a significant match, and 0.81 is considered as a large matching value. Since ZLD has been marketed as a drug, its effect on COVID–19 can be tested in cell and mouse models to determine further whether it can treat COVID–19.

Declarations

Author contribution: Zehua Zeng, Zhi Luo and Hongwu Du designed and performed research; Yubang Shan, Chuan Chen, Luna Wang collected data; Jingjie Yang guided the polish; Zhi Luo and Zehua Zeng analyzed data; and Zehua Zeng, Zhi Luo, and Hongwu Du wrote the paper.

The authors declare no competing interest.

References

1.Xu X, Chen P, Wang J, Feng J, Zhong W, Zhou H, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. science china life sciences.

2.Wan S, Yi Q, Fan S, Lv J, Zhang X, Guo L, et al. Characteristics of lymphocyte subsets and cytokines in peripheral blood of 123 hospitalized patients with 2019 novel coronavirus pneumonia (NCP). medRxiv. 2020 2020–01–01:2020–2.

3.Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID–19: consider cytokine storm syndromes and immunosuppression. The Lancet. 2020.

4.Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. NATURE. 2020 2020–01–01;579(7798):270–3.

5.Holappa M, Vapaatalo H, Vaajanen A. Many Faces of Renin-angiotensin System - Focus on Eye. The open ophthalmology journal. 2017 2017–01–01;11(1):122–42.

6.Sun Y, Liu L, Pan X, Jing M. Mechanism of the action between the SARS-CoV S240 protein and the ACE2 receptor in eyes. INT J OPHTHALMOL-CHI. 2006;6(4):783–6.

7.Hamming I, Timens W, Bulthuis MLC, Lely AT, Goor HV. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J PATHOL. 2004;203(2):631–7.

8.Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of 2019 novel coronavirus infection in China. NEJM. 2020 2020–01–01:2020–2.

9.Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet. 2020;395(10223):507–13.

10.Mahy BWJ. Virology: Molecular biology of the coronaviruses. NATURE. 1983 1983–01–01;305(5934):474–5.

11.Tan Y, Lim SG, Hong W. Characterization of viral proteins encoded by the SARS-coronavirus genome. ANTIVIR RES. 2005 2005–01–01;65(2):69–78.

12.Chen J, Ly H. Immunosuppression by viral N proteins. Oncotarget. 2017 2017–08–01;8(31):50331–2.

13.Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. NAT PROTOC. 2010 2010–01–01;5(4):725–38.

14.Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein structure and sequence re-analysis of 2019-nCoV genome does not indicate snakes as its intermediate host or the unique similarity between its

spike protein insertions and HIV–1. 2020 2020–01–01.

15.Peng H, Yang L, Wang L, Li J, Huang J, Lu Z, et al. Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients.;351(2):466–75.

16.Li T, Jing X, Yuxian H, Hongwei F, Laurence B, Zhifeng Q, et al. Long-Term Persistence of Robust Antibody and Cytotoxic T Cell Responses in Recovered Patients Infected with SARS Coronavirus. PLOS ONE.;1(1):e24.

17.Oh HLJ, Chia A, Chang CXL, Leong HN, Bertoletti A. Engineering T Cells Specific for a Dominant Severe Acute Respiratory Syndrome Coronavirus CD8 T Cell Epitope. J VIROL. 2011;85(20):10464–71.

18.Lau SKP, Woo PCY, Wong BHL, Tsoi HW, Woo GKS, Poon RWS, et al. Detection of Severe Acute Respiratory Syndrome (SARS) Coronavirus Nucleocapsid Protein in SARS Patients by Enzyme-Linked Immunosorbent Assay. J CLIN MICROBIOL.;42(7):2884–9.

19.Gao W, Tamin A, Soloff A, D'Aiuto L, Nwanegbo E, Robbins PD, et al. Effects of a SARS-associated coronavirus vaccine in monkeys. The Lancet. 2003 2003–01–01;362(9399):1895–6.

20.Siddell SG. The Coronaviridae; 1995.

21.He R, Dobie F, Ballantine M, Leeson A, Li Y, Bastien N, et al. Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochemical & Biophysical Research Communications.;316(2):483.

22.Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, et al. Structure of the N-Terminal RNA-Binding Domain of the SARS CoV Nucleocapsid Protein. BIOCHEMISTRY-US. 2004 2004–01–01;43(20):6059–63.

23.Surjit M, Liu B, Jameel S, Chow VTK, Lal SK. The SARS coronavirus nucleocapsid protein induces actin reorganization and apoptosis in COS–1 cells in the absence of growth factors. The Biochemical journal. 2004 2004–01–01;383(Pt 1):13–8.

24.Chang MS, Lu Y, Ho S, Wu C, Wei T, Chen C, et al. Antibody detection of SARS-CoV spike and nucleocapsid protein. Biochemical & Biophysical Research Communications.;314(4):936.

25.Yuan X, Yao Z, Shan Y, Chen B, Yang Z, Wu J, et al. Nucleolar localization of non-structural protein 3b, a protein specifically encoded by the severe acute respiratory syndrome coronavirus. VIRUS RES. 2005 2005–01–01;114(1):70–9.

26.Yuan X, Shan Y, Zhao Z, Chen J, Cong Y. G0/G1 arrest and apoptosis induced by SARS-CoV 3b protein in transfected cells. VIROL J. 2005 2005–01–01;2(1):66.

27.Thiel V, Ivanov KA, Putics A, Hertzig T, Ziebuhr J. Mechanisms and enzymes involved in SARS coronavirus genome expression. J GEN VIROL. 2003;84(Pt 9):2305–15.

28.Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, Sparks J, et al. Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. J VIROL.;79(23):14909–22.

29.Consortium TCSM. Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. SCIENCE.;303.

30.Su YC, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, et al. Discovery of a 382-nt deletion during the early evolution of SARS-CoV–2. 2020.

31.Zhao P, Cao J, Zhao LJ, Qin ZL, Qi ZT. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. VIROLOGY. 2005;331(1):128–35.

32.Hung L, Kim H, Murakami S, Gupta G, Kim C, Terwilliger TC. Crystal structure of AcrB complexed with linezolid at 3.5 Å resolution. Journal of Structural and Functional Genomics. 2013 2013–01–01;14(2):71–5.

33.Madden T. The BLAST Sequence Analysis Tool. The NCBI Handbook [Internet]. 2nd edition.; 2013.

34.Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. NUCLEIC ACIDS RES. 2019 2019–11–28;48(D1):D265–8.

35.Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. NAT PROTOC.;10(6):845–58.

36.Jianyi, Yang, Ambrish, Roy, Zhang. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment.

37.Zhang C, Freddolino PL, Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. NUCLEIC ACIDS RES. 2017 2017–05–02;45(W1):W291–9.

38.Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins Structure Function and Bioinformatics.;57(4):702–10.

39.Zhang J, Kurgan L. Review and comparative assessment of sequence-based predictors of protein-binding residues. BRIEF BIOINFORM. 2017 2017–03–01;19.

40.Liu W, Xie Y, Ma J, Luo X, Peng N, Zuo Z, et al. IBS: an illustrator for the presentation and visualization of biological sequences. BIOINFORMATICS. 2015(20):20.

Figures

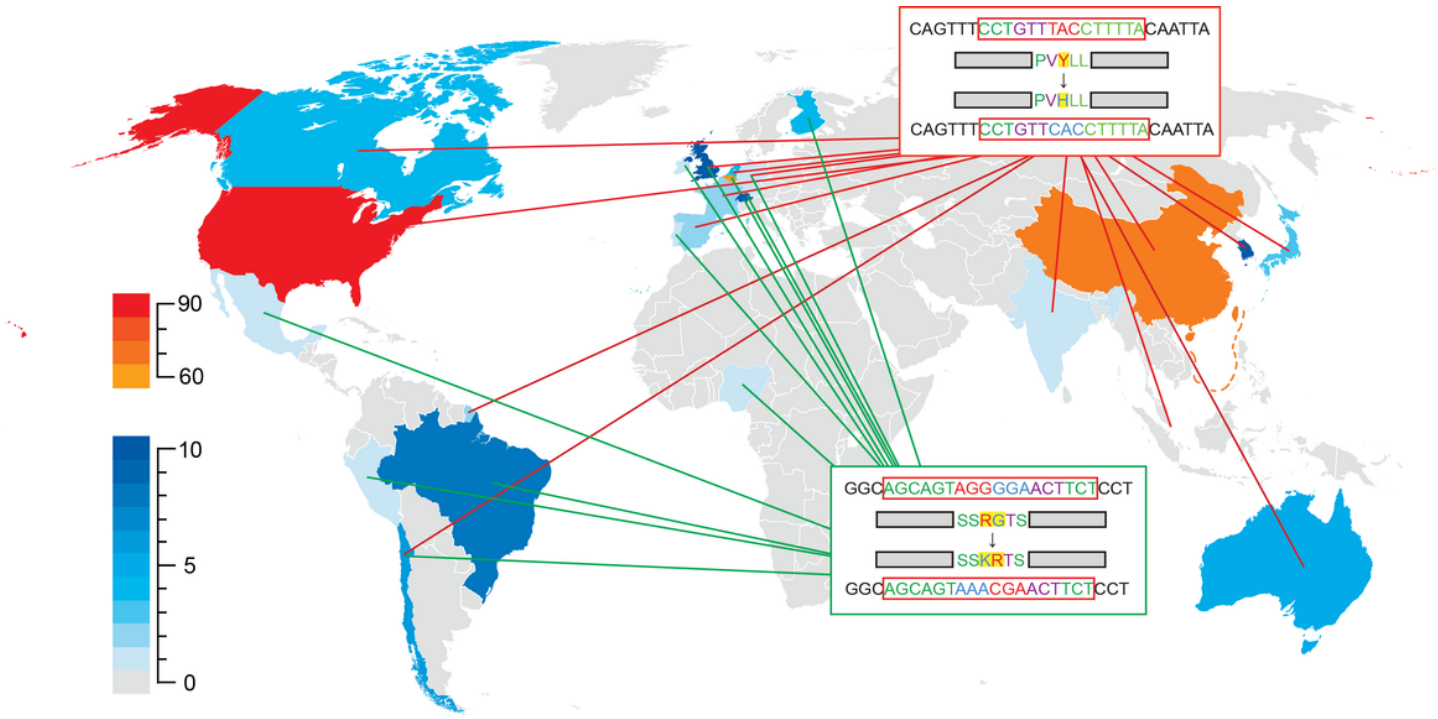

Figure 1

The distribution of mutations in worldwide and a part of the nucleotide sequence (in the red box) and its translated amino acid sequence(grey box represents amino acid sequence which is translated from nucleotide outside the red box in the figure). TAC mutated into CAC, which led to Tyrosine (Y) mutating into histidine(H). AGG mutated into AAA and GGA mutated into CGA, which made arginine(R) mutate into lysine(K) and glycine(G) mutate into arginine(R). the amino acid mutations were highlighted in yellow. The distribution of 2 mutations was distinguished by Red and Green segments. The colour of each country represents the number of cases. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

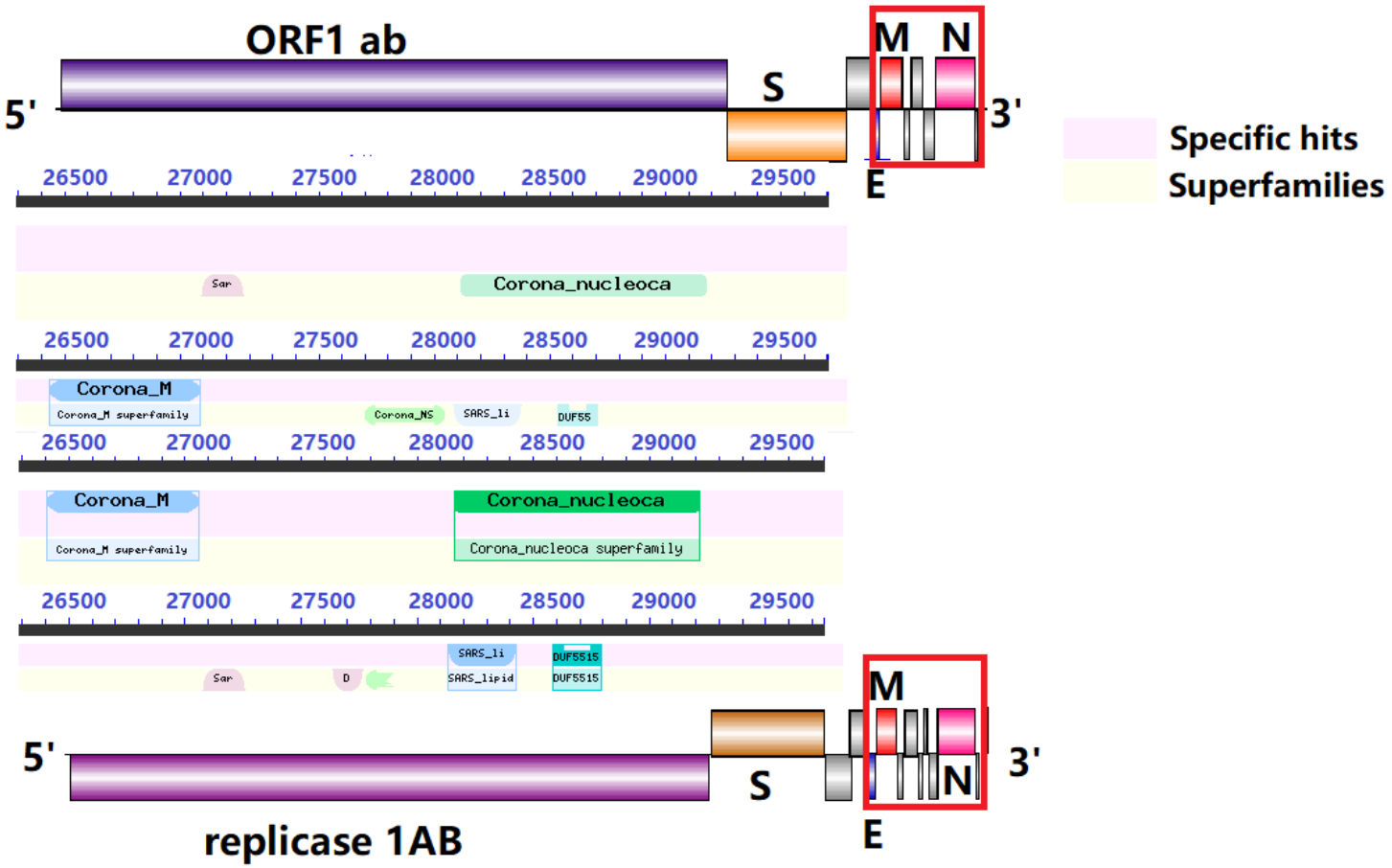

Figure 2

The ORF analysis. The upper half of the figure is divided into SARS-CoV-2, and the lower half is SARS-CoV. The five main open reading frames (ORF) of the virus are shown in colour: purple for RNA replicase, orange for surface glycoprotein, blue for E protein, red for M protein, pink for N protein, and the remaining open reading frames are shown in grey. For the ORF, lilac represents the specific hits, and pale yellow represents the superfamily. This image shows a graphical summary of conserved domains identified in the query sequence. The Show Concise/Full Display button at the top of the page can be used to select the desired level of detail: only top-scoring hits (labelled illustration) or all hits (labelled illustration). Domains are colour-coded according to superfamilies to which they have been assigned. Hits with scores that pass a domain-specific threshold (specific hits) are drawn in bright colours. Others (non-specific hits) and superfamily placeholders are drawn in pastel colours. With IBS drawing(40).

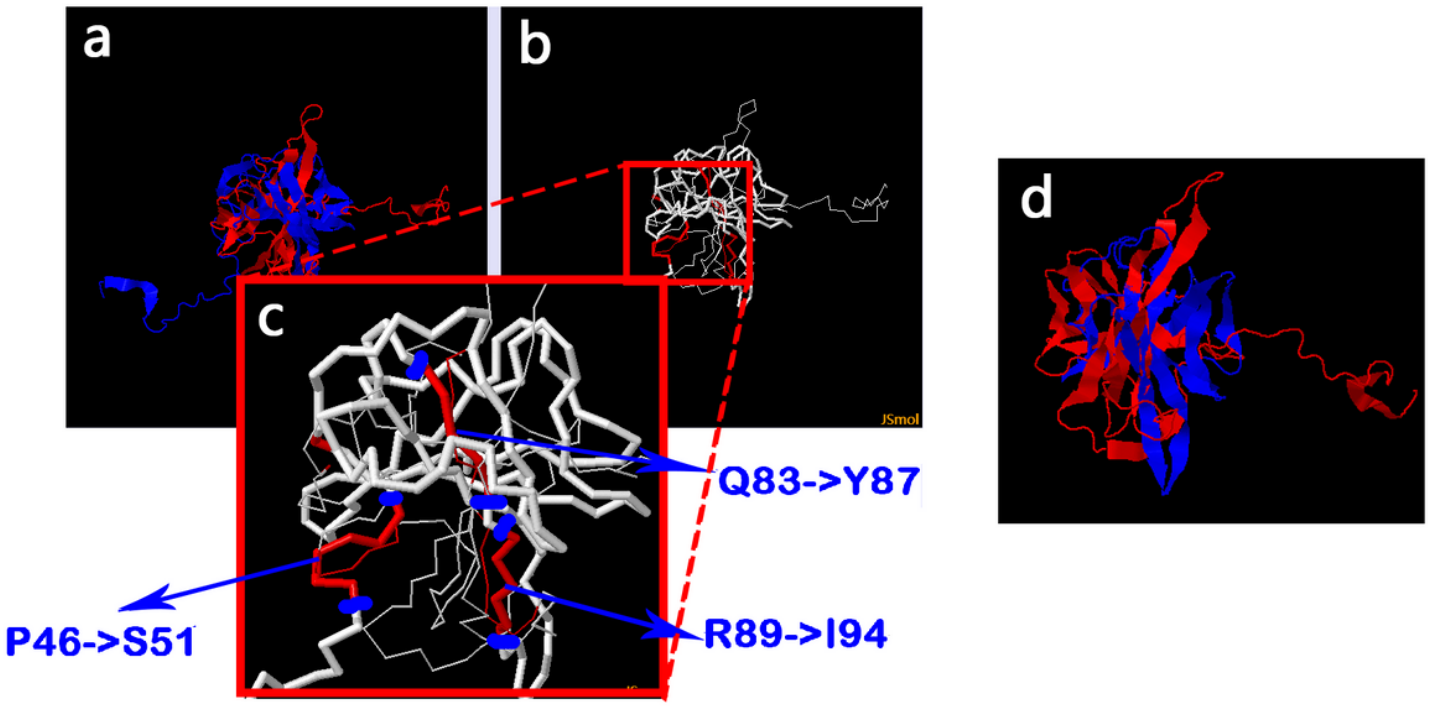

Figure 3

Spatial structure difference. a. The cartoon was protein structure composite diagram (SARS-CoV-2 Structure in blue and SARS-CoV Structure in red). b. The cartoon was skeleton mixed model (the N protein of SARS-CoV-2 in thick and the N protein of SARS-CoV in thin wireframes; Residues with d<5A in red). c. The cartoon was a partial enlarged drawing of the mixed skeleton model. d. Comparison of SARS-CoV(in Red) and SARS-CoV-2(in Blue) structure cartoon model on the right side of figure

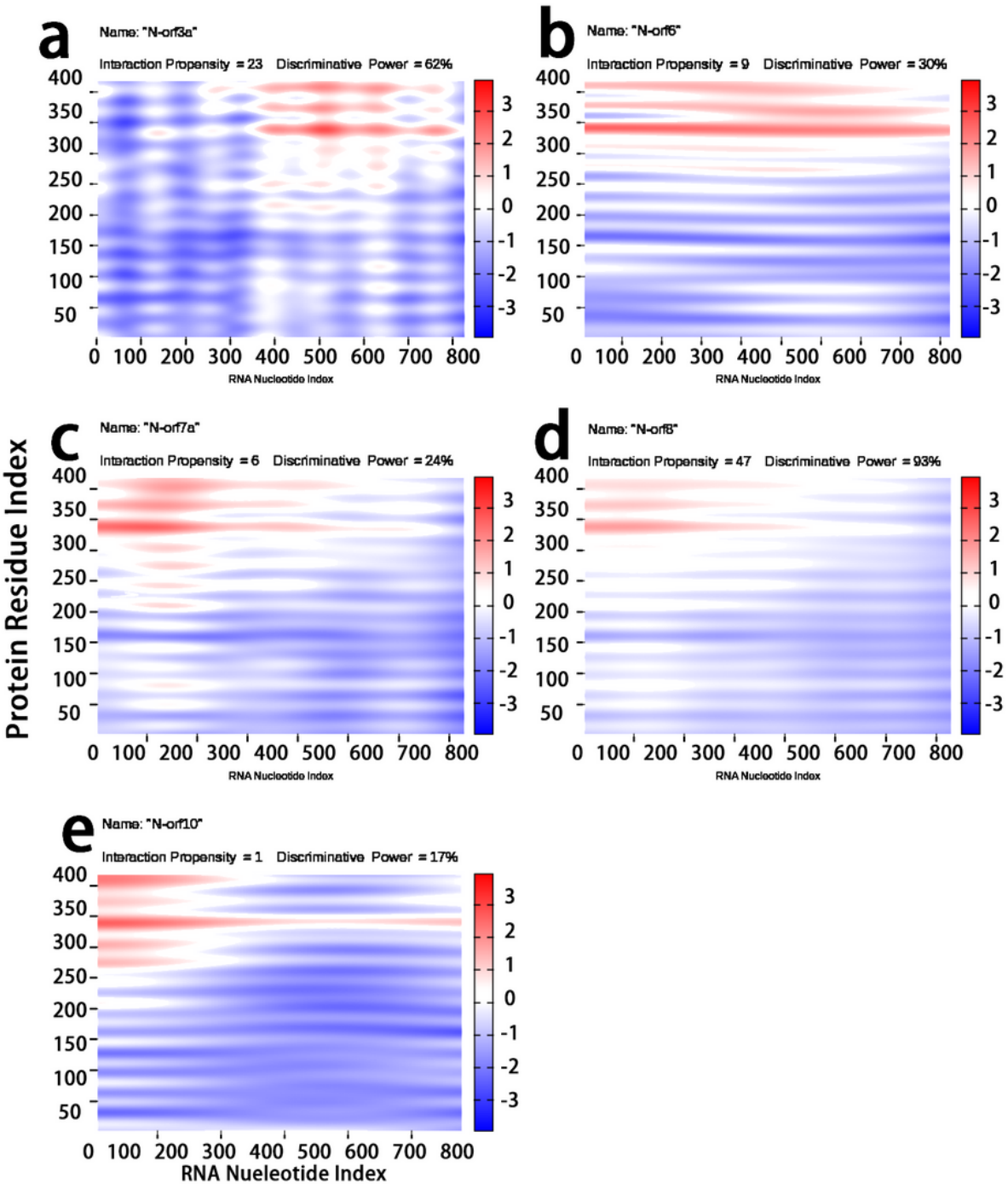

Figure 4

The tendency of proteins bind to RNA. The colours in the heatmap indicate the interaction score (ranging from -3 to +3) of the individual amino acid and nucleotide pairs. The total sum represents the overall interaction score. The abscissa represents the RNA nucleotide sequence, 5' to 3' from left to right. The ordinate represents the sequence of amino acids, from the top-down, from the C’ to the N’.

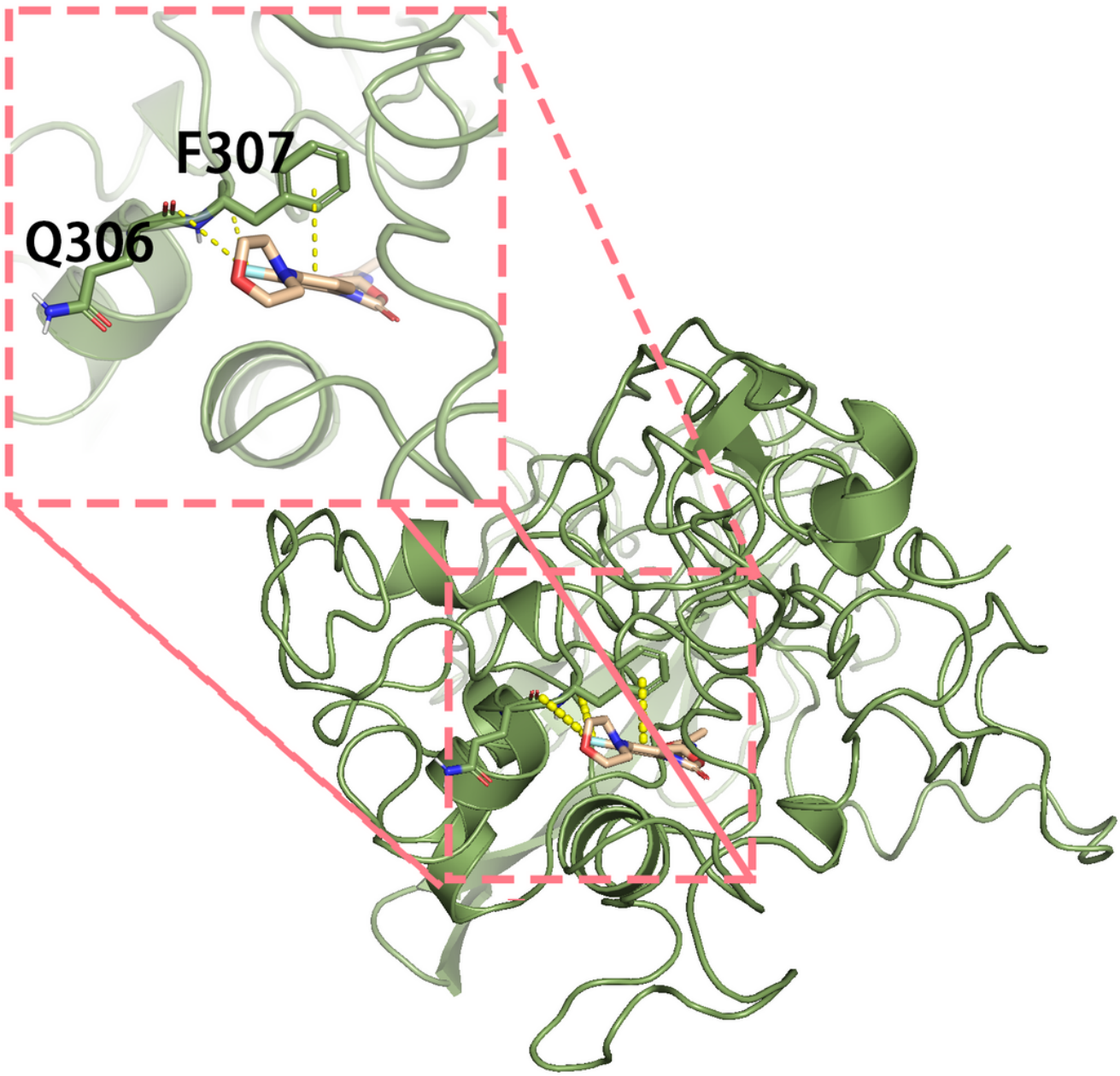

Figure 5

Schematic diagram of ZLD binding to N protein. The green cartoon shows the spatial structure of N protein, the brown one shows the ZLD molecule, the yellow dotted line shows the interaction force, and the black one represents the residue name and the corresponding site.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement.docx