Structural genomics and interactomics of 2019 Wuhan novel coronavirus, SARS-CoV-2, indicate evolutionary conserved functional regions of viral proteins

Authors: Hongzhu Cui[1,$], Ziyang Gao[1], Ming Liu[1], Senbao Lu[1], Winnie Mkandawire[1], Oleksandr Narykov[2], Suhas Srinivasan[3], Mo Sun[1], and Dmitry Korkin[1,2,3,*]

Affiliations

[1]Bioinformatics and Computational Biology Program, Worcester Polytechnic Institute, Worcester, MA 01609

[2]Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609

[3]Data Science Program, Worcester Polytechnic Institute, Worcester, MA 01609

*Corresponding author. E-mail: korkin@korkinlab.org

[$]The authors would like to note that the first eight authors are listed alphabetically.

Abstract

During its first month, the recently emerged 2019 Wuhan novel coronavirus (SARS-CoV-2) has already infected many thousands of people in mainland China and worldwide and took hundreds of lives. However, the swiftly spreading virus also caused an unprecedentedly rapid response from the research community facing the unknown health challenge of potentially enormous proportions. Unfortunately, the experimental research to understand the molecular mechanisms behind the viral infection and to design a vaccine or antivirals is costly and takes months to develop. To expedite the advancement of our knowledge we leverage the data about the related coronaviruses that is readily available in public databases, and integrate these data into a single computational pipeline. As a result, we provide a comprehensive structural genomics and interactomics road-maps of SARS-CoV-2 and use these information to infer the possible functional differences and similarities with the related SARS coronavirus. All data are made publicly available to the research community at http://korkinlab.org/wuhan .

Importance

Having infected over 60,000 people and killed more than 1,300, the new Wuhan coronavirus, SARS-CoV-2, has a major health and economic impact worldwide. With a genome of the new virus sequenced, scientists learned that this virus is closely related to other coronaviruses, such as SARS and MERS. This discovery, however, leads to other questions. When looking at the proteins of SARS-CoV-2, virus' main functional building blocks, how different those proteins are from those ones of SARS? Does the new virus target the same proteins in human cells as the known coronaviruses? Most importantly, can the currently developed SAR or MERS drug candidates or other promising compounds be used to treat the new infection? This study lays out the first 3D roadmap of SARS-CoV-2 viral proteins and their functional complexes, enabling scientists to answer the above questions. The roadmap can be used to streamline the search for new antivirals and vaccines.

Introduction

Within month and a half of its initial discovery in Wuhan, China, the novel deadly Coronavirus (SARS-CoV-2) has infected more than 37,000 people, with the death toll already surpassing that of the 2003 SARS epidemic [1-3]. In spite of the instantaneous reaction by the scientific community and extensive world wide efforts to address this health crisis, vaccines may be months and even years away [4, 5]. For instance, a Phase I trial of a vaccine for a closely related to SARS-CoV-2, severe acute respiratory syndrome (SARS) virus, was announced in December 2004, two years after the disease outbreak [6]. Additionally, a vaccine for Middle East respiratory syndrome (MERS), another coronavirus that emerged in 2012, was patented in 2019, the same year Phase I trials were introduced [7, 8]. Nevertheless, in the past two decades, a massive amount of work has been done to understand the molecular basis of the coronavirus evolution and infection, develop effective treatment in forms of both vaccines and antiviral drugs, and propose efficient measures for viral detection and prevention [9-18]. Structures of many

individual proteins of SARS, MERS, and related coronaviruses, as well as their biological interactions with other viral and host proteins have been explored along with the experimental testing of the anti-viral properties of small-molecule ihibitors [16, 19-31].

However, experimental investigation of the same scale for SARS-CoV-2 may take the research community years to obtain. Can it be facilitated? The answer lies in the use of the modern bioinformatics methods that can drastically streamline knowledge discovery by quickly providing the important insights about possible molecular mechanisms behind infection, pinpointing likely protein targets for the anti-viral treatments, and predicting the efficacy of the existing antivirals developed for other coronaviruses. By leveraging previously known information on genome sequences as well as protein structure and function, bioinformaticians have been successfully helping virologists by structurally characterizing proteins of novel viruses, determining the evolutionary trajectories, identifying interactions with host proteins, and providing other important biological insights. In particular, a plethora of results has been achieved through comparative, or homology, modeling principles [32, 33]. In addition to the global structural genomics, an initiative that is focusing on determining the 3D structures of individual proteins on a genome scale [34], as well as to the specific efforts aimed at rapid structural characterization of proteins in emerging viruses [35-38], multiple works have used comparative modeling to predict the structures of protein-protein interaction complexes [39-41], facilitate structure-based drug discovery [33, 42, 43], infer protein functions [44], determine the macromolecular interaction network [45-47], and provide molecular insights into viral evolution [48-50].

Here, using an integrated bioinformatics approach, we provide a comprehensive structural genomics and interactomics analysis of the Wuhan SARS-CoV-2. The structural information on the individual SARS-CoV-2 proteins and their interactions with each other and human proteins allows us to accurately determine the putative functional sites. These functional sites, combined with an evolutionary sequence analysis of SARS-CoV-2 as well as the closely related human SARS and bat coronavirus proteomes,

provide us with a structure-based perspective of the evolutionary diversity of SARS-CoV-2, allowing the estimation of how similar the function of SARS-CoV-2 virus is when compared with SARS. Consequently, we can forecast how likely the antibodies and candidate ligands that are efficient in inhibiting the SARS functions will be efficient in doing the same for SARS-CoV-2.

Results

The recently sequenced genomes of SARS-CoV-2 strains combined with the comparative analysis of the SARS genome organization and transcription allowed us to construct a tentative list of gene products [51]. It has been suggested that SARS-CoV-2 has 16 predicted non-structural proteins (referred to as wNsp1-wNsp16 here) constituting a polyprotein (wORF1ab), followed by (at least) 13 downstream open reading frames (ORFs): Surface, ORF3a, ORF3b, Envelope, Membrane, ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid, ORF9a, ORF9b, and ORF10 that we refer in this work as wS, wORF3a, wORF3b, wE, wM, wORF6, wORF7a, wORF7b, wORF8, wN, wORF9a, wORF9b, and wORF10, respectively. The three viral species whose proteins shared the highest similarity were consistently the same: human SARS (SARS-CoV), bat coronavirus (BtCoV), as well as another bat betacoronavirus (BtRf-BetaCoV).

Comparative analysis of SARS-CoV-2 proteins with the evolutionary related coronavirus proteins reveals unevenly distributed large genomic insertions

Searching against UniProt database [52] resulted in matches for the polyprotein (wORF1ab), all four structural proteins (wS, wE, wM, and wN), and six ORFs (wORF3a, wORF6, wORF7a, wORF7b, wORF8, and wORF10) also referred to as accessory proteins. The closest protein matches from UniProt shared sequence identity with the related SARS-CoV-2 proteins as high as 91% (with wORF1ab and wN) and as low as 57% (with wORF8) (Supl. Table S1). The majority of differences were single-residue substitutions spread across the protein sequence (see multiple sequence alignment files for all proteins in Suppl. Materials).

Perhaps the most profound differences lie in the sequences of the multi-domain protein wNsp3 and surface protein wS: our analysis revealed that, compared to related coronavirus proteins, the two proteins had large sequence inserts (multiple sequence alignments for both proteins can be found in Suppl. Materials). In particular, wNsp3 had a novel large (25–41 res., depending on the alignment method) insert between its two putative functional domains, which are homologous to N-terminal domain and adenosine diphosphate ribose 1″ phosphatase (ADRP) of SARS [53, 54] (Suppl. Materials). (Suppl. Materials). Interestingly, the closest matching peptide, apart from the virus itself, was found in C-Jun-amino-terminal kinase-interacting protein 4 (Seq. identity is 46%) of *Labrus bergylta*, a species of marine ray finned fish. Being significantly more diverse than the other three structural proteins, wS was found to have 4 inserts (4–6 res.) that seemed unique to SARS-CoV-2 and two additional inserts shared with human SARS proteins (Suppl. Materials).

Three recent strains of bat SARS-like coronavirus from 2013, 2015, and 2017 share extremely high proteome similarity with SARS-CoV-2

The unusually low conservation of ORF6, ORF8, and surface proteins between SARS-CoV-2 and human SARS, bat coronavirus, as well as another bat betacoronavirus, BtRf-BetaCoV, prompted us to perform an expanded search for ORF8 homologs using NCBI BLAST's blatp tool against a large non-redundant protein sequence repository (nr) [55]. Our search resulted in three new homologs of ORF8 from three different isolates of bat SARS-like coronavirus: bat-SL-CoVZC45 (GenBank ID: MG772933, collected in 2017), bat-SL-CoVZXC21 (GenBank ID: MG772934, collected in 2015), and RaTG13 (GenBank ID: MN996532, collected in 2013) that shared a striking similarity with wORF8, unseen in other strains before: the sequence identities between each of these three homologs and wORF8 ranged between the 94 and 95%. Further analysis showed that the proteomes of these isolates shared even higher sequence identity with the other proteins of SARS-CoV-2: from 88.4% to 100% for 2015 and 2017 isolates, and even higher, 97.4%–100% for the 2013 isolate (Supl. Table S1).

In spite of the significant similarity of the three isolates to SARS-CoV-2, important differences were observed. First, similar to other viruses, the 2017 and 2015 isolates did not have the four sequence inserts that were found in wS. Second, neither of the two isolates had the large insert between the two domains of wNsp3 from wORF1ab described above. On the contrary, the 2013 isolate had both, the four sequence inserts in its surface protein, matching those in wS, and the large insert in Nsp3, although the sequence of the large insert is different from that one in wNsp3. The main difference between SARS-CoV-2 and the 2013 isolate is the lack of Orf10 in the latter: While the genomic sequences of both the 2015 and 2017 isolates can be translated into a protein product that shares 97.4% sequence identity (Supl. Table S1), because of a single nucleotide deletion in that region one cannot translate a full-length ORF10 in RaTG13 because the resulting frameshift causes a premature stop-codon when translating the sequence.

Structural genomics and interactomics analysis of SARS-CoV-2

Next, as a result of a comprehensive comparative modeling effort, we were able to structurally characterize 17 individual proteins, including 13 non-structural proteins of wORF1ab (wNsp1, wNsp3, wNsp4, wNsp5, wNsp7, wNsp8, wNsp9, wNsp10, wNsp12, wNsp14, wNsp15, wNsp16), three structural proteins (wE, wN, and wS), as well as one ORF (wORF7a). For two proteins, wNsp3 and wN, multiple individual domains were modeled (Fig.1, Fig.2). The templates for the majority of the models were homologous protein structures from other coronaviruses, with a high target-to-template sequence similarity (seq. ids: 75–96%, except for ORF8). For two proteins, wN and wNsp3, multiple domains were modeled. N-terminal and C-terminal domains of wN correspond to N-terminal RNA-binding domain and C-terminal dimerization domain of SARS, respectively. The modeled domains 1–6 of wNsp1 correspond to (1) N-terminal domain; (2) adenosine diphosphate ribose 1″ phosphatase domain; (3) SUD domain (SARS-specific unique domain) containing two macrodomains; (4) SUD domain C; (5) the papain-like protease PLPro domain; and (6) Y domain of SARS (Fig. 1). A previously identified transmembrane

domain of SARS Nsp3 was mapped to the sequence of wNsp3, but could not be modeled due to the lack of a template structure.

The structural analysis of the modeled proteins combined with the sequence conservation analysis revealed several findings. First, we found that the mutated residues tend to locate on the protein's surface, supporting previous observations in other families of RNA viruses that the core residues of viral proteins are more conserved than the surface residues [48, 56, 57]. Furthermore, in a substantial number of proteins, distributions of mutated positions exhibited spatial patterns, with groups of mutations found to form clusters on the protein surfaces (Fig.2). These obtained models were then used as reference structures to map and analyze the protein-binding and ligand-binding sites.

Next, using comparative modeling we structurally characterized protein interaction complexes, for both intra-viral (homo- and hetero-oligomers) and host-viral interactions where host proteins were exclusively human. In total, we obtained structural models for 16 homo-oligomeric complexes, three hetero-oligomeric complexes, and eight human-virus interaction complexes (Fig. 3). The intra-viral hetero-oligomeric complexes included exclusively the interactions between the non-structural proteins (wNsp7, wNsp8, wNsp10, wNsp12, wNsp14, and wNsp16). The modeled host-viral interaction complexes included three types of interactions: non-structural protein wNsp3 (papain-like protease, PLpro, domain) interacting with human ubiquitin-aldehyde, surface protein wS (in its trimeric form) interacting with human receptor angiotensin-converting enzyme 2 (ACE2) in different conformations, as well as the same protein wS interacting with several neutralizing antibodies. Based on the obtained models, the protein interaction binding sites were extracted and analyzed with respect to their evolutionary conservation.

Evolutionary conservation and divergence of functional regions of SARS-CoV-2

The analysis of the evolutionary conservation of protein binding sites revealed several patterns. First, we found that all protein binding sites of non-structural proteins involving in the intra-viral heteromeric complexes, wNSP7-wNsp8-wNsp12, wNsp10-wNsp14, and wNsp10-wNsp16, are either fully conserved or allow at most one mutation on the periphery of the binding region, in spite of the fact that each proteins had multiple mutations on their surfaces (Fig. 4A, see Suppl. Materials for the alignments of individual proteins annotated with the protein binding sites). Furthermore, we observed the same behavior when analyzing the interaction between the papain-like protease PLpro domain of wNsp3 and human ubiquitin-aldehyde (Fig. 4B, see Suppl. Materials): the only two mutated residues were located on a border of the binding region and thus were unlikely to disrupt the protein-protein interaction.

The surface of wS presents a striking contrast to the majority of SARS-CoV-2 proteins due to its heavily mutated surface (Fig. 3; see Suppl. Materials for the alignment of wS with related proteins). First, the four novel sequence inserts and two inserts shared with the closest strains were expected to affect the protein's function. Interestingly, three of the novel inserts are located in the first NTD domain, while the fourth one is located immediately before the S2 cleavage site and inside the homo-trimerization interaction interface. While the RBD domain of wS was not affected by those inserts it was the most heavily mutated region of wS with likely disruptive functional effects on the interactions with the human ACE2 receptor and monoclonal antibodies mAb 396 [58] and mAb 80R [59] (Fig. 4B). The NTD domain has been considered a target of another antibody Ab G2 previously shown to work in MERS [60]. However, based on the structural superposition of the NTD domain, it is likely that the expected interaction will be disrupted by the novel sequence inserts to wS.

Finally, the analysis of seven ligand binding sites (LBS) for multiple candidate inhibitors previously identified for four proteins in SARS and MERS showed that many of the LBS were intact in the corresponding SARS-CoV-2 proteins (Fig. D, see Suppl. Materials for the alignments of individual

proteins annotated with the ligand binding sites) such as wN, wNsp3, wNsp5, and wNsp16. For wNsp14, the LBS for several inhibitors was mutated while a co-localized binding site for another ligand was intact.

Joint intra-viral and human-virus protein-protein interaction network for SARS CoV indicates potential system-wide roles of SARS-CoV-2 proteins

After constructing the individual networks of intra-viral interactions and virus-host interactions, they were merged to form a unified network of SARS-CoV interactions, with the hypothesis that most interactions would be conserved in SARS-CoV-2 (Fig. 6). The network analysis of all three networks showed that unifying the intra-viral and virus-host networks reduces the number of components (islands) in the SARS-CoV-Host interactome (Table 2). This suggested that the virus-host interactome map missed some viral interactions that increased the number of components. The clustering coefficient and average node degree were both higher in the unified interactome compared to the virus-host interactome.

The network topology of the unified interactome indicated the presence of several viral hubs for the structural, non-structural, and accessory proteins (Fig. 6). The viral proteins form all the hubs and there are a few of specific interest. ORF9b is one of the largest hubs in the network thought to play a secondary role only in intra-viral interactions and not necessary for replication [61], but a recent study has shown that ORF9b hinders immunity by targeting the mitochondria and limits host cell responses [62]. NSP8 is the other major hub and interacts with other replicase proteins, including two other hubs, i.e. NSP7 and NSP 12, which together play a crucial role in replication [63].

Another crucial non-structural protein is NSP1, also present in the SARS-CoV-2, which is known to be a primary disruptor of innate immunity. The host interaction partners of NSP1 modulate the Calcineurin/NFAT pathway that plays an important role in immune cell activation [64]. Overexpression of NSP1 is associated with immunopathogenecity and long-term cytokine dysregulation as observed in

severe SARS cases. Additionally, inhibition of cyclophilins/immunopilins (host interaction partners) by cyclosporine A (CspA) blocks the replication of CoVs of all genera.

There is an important interaction between the SARS spike protein (S) and the host angiotensin-converting enzyme 2 (ACE2), as it is associated with cross-species and human-to-human transmissions. The same interaction is also inferred from structural modeling between SARS-CoV-2 wS and ACE2, but might be disrupted due to a substantial number of mutations in the receptor binding site of wS though another recent study proposed that the interaction will be preserved [65]. Similar to SARS-CoV [66], wS protein in SARS-CoV-2 is expected to interact with type II transmembrane protease (TMPRSS2) [67] and is likely to be involved in inhibition of antibody-mediated neutralization. Thus, wS remains an important target for vaccines and drugs previously evaluated in SARS and MERS while a neutralizing antibody targeting the wS protein could provide passive immunity [68]. In addition, there are 7 interactions from SARS-CoV determined by structural characterization of the protein complexes that are predicted to be either conserved or potentially disrupted in SARS-CoV-2 (green edges in Fig. 1). An important target for vaccines and drugs is the surface (S) protein which has been evaluated in SARS-CoV and MERS-CoV with the idea that a neutralizing antibody targeting wS protein could provide passive immunity for SARS-CoV-2 [68]. We also structurally modeled interactions between the SARS-CoV-2 wS protein and three human monoclonal antibodies that were previously studied in SARS-CoV for immunotherapy and mapped the information about the evolutionary conserved and diverse surface residues. We find that two interactions, with mAb 396 [58] and mAb 80R [59], are likely to be disrupted due to the heavily mutated binding sites while another, Ab G2 [60], is likely disrupted due to a novel insert into the sequence of wS. These findings may provide guidelines in the search for potential antibody candidates for treatment.

Discussion

This work provides an initial large-scale structural genomics and interactomics effort towards understanding the structure, function, and evolution of the SARS-CoV-2 virus. The goal of this

computational work is two-fold. First, by making the structural road map and the related findings fully available to the research community, we aim to facilitate the process of structure-guided research where accurate structural models of proteins and their interaction complexes already exist. Second, by providing a comparative analysis between the new virus and its closest relatives from the perspective of protein- and ligand-binding, we hope to help experimental scientists in their deciphering of the molecular mechanisms implicated in infection by the new coronavirus as well as in vaccine development and antiviral drug discovery.

Through integrating the information on structure, function, and evolution in a comparative study of SARS-CoV-2 and the closely related coronaviruses, one can make several preliminary conclusions. First, the extended peptide sequence newly introduced to wNsp3 between two structurally and possibly functionally independent domains of this protein, might act as a long inter-domain linker, thus extending the conformational flexibility of this multi-domain protein. Second, the presence of the four novel inserts and one highly variable region of the surface protein wS and the analysis of this large-scale sequence change with respect to intra-viral and viral-host interactions leads us to conclude that these inserts might have structural impact on the homo-trimeric form of the protein as well as impact the functions carried out by NTD domain. Third, the structurally modellable repertoire of SARS-CoV-2 proteome also pinpoints to the interesting targets for structural biology and hybrid methods. For instance, the whole structures of the multi-domain proteins wN and wNsp3 could be resolved by integrating individual models with the lower-resolution but whole-protein covering techniques such as cryogenic electron microscopy (CryoEM). The structure of wNsp3 is especially interesting because of the presence of the novel peptide introduced between the two structural domains of the protein.

The evolutionary analysis of the protein- and ligand-binding sites mapped on the surfaces of SARS-CoV-2 may provide new insights into the virus functioning and its future treatment. The 100% or near 100%

evolutionary conservation of the protein binding sites on the surfaces of non-structural proteins wNsp7, wNsp8, wNsp10, wNsp12, wNsp14, and wNsp16 that correspond to the intra-viral interactions for three complexes is consistent with our previous observations that the intra-viral interactions are significantly more conserved than viral-host interactions [48, 49, 69]. However, the near-perfect conservation of the human ubiquitin-aldehyde protein binding site on the surface of wNsp3 is rather intriguing, suggesting the critical role of this interaction in the functioning of SARS-like coronaviruses. Lastly, the conservation of the ligand-binding sites for many putative inhibitors previously developed for SARS and MERS suggests the development of antiviral drugs as a promising direction for addressing this new global health threat.

Materials & Methods

The goal of this work is to identify the evolutionary differences between SARS-CoV-2 and the closest coronavirus species, human SARS coronavirus and bat SARS-like coronavirus, and to predict their possible functional implications. To do so, we structurally characterize individual proteins as well as intra-viral and human-virus protein complexes, extract the information on their interaction interfaces and ligand binding, and superpose the evolutionary difference and conservation information with the binding information. Specifically, our integrative computational pipeline will include the following five steps. First, for each of the candidate SARS-CoV-2 proteins a set of sequentially similar coronavirus proteins is determined and aligned. Second, structural models of SARS-CoV-2 proteins are obtained using template-based, or homology, modeling. Third, the protein-protein interaction complex structures of SARS-CoV-2 proteins interacting with each other and/or with the human proteins are determined using a multi-chain comparative modeling protocol. Fourth, the protein-binding sites will be extracted from the obtained models of protein-protein interaction complexes, and protein-ligand binding sites will be extracted from the evolutionary close coronavirus protein-ligand templates and mapped to the relevant structural models of SARS-CoV-2 through structural alignment. Fifth, the information on the evolutionary differences and conservations observed between the protein sequences of SARS-CoV-2 and the related coronaviruses and

extracted from the above protein sequence alignments will mapped onto the structural models of the SARS-CoV-2 proteins to determine if the protein- and ligand-binding sites are functionally conserved. Lastly, a joint human-virus and virus-virus interactome is predicted through homology and analyzed. All the obtained models and sequence alignments have been made publicly available to the research community.

Protein sequence data collection and analysis

Available sequences for protein candidates wS, wORF3a, wE, wM, wORF6, wORF7a, wORF7b, wORF8, wN, and wORF10 are extracted from NCBI Virus repository [70] (collected on January 29, 2019) and then used in the sequence analysis and structural modeling (Suppl. Materials). The Uniprot BLAST-based search is performed for each of the proteins using default parameters. From the results of each search, the final selection is done based on the pairwise sequence identity (>60%) as well as the relationship (*Coronaviridae* family). Each of SARS-CoV-2 proteins is then aligned with the corresponding coronavirus proteins using the multiple sequence alignment method Clustal Omega [71].

Structural characterization of protein and protein complexes

The structure of each protein is determined using a single-template comparative modeling protocols with the MODELLER software package [72]. First, the template for each protein sequence is identified using a PSI-BLAST search in Protein Data Bank (PDB) [73]. In general, a structural template with the highest sequence identity is selected out of those that cover at least 50 residues of the target sequence with at least 30% sequence identity. The polyprotein wORF1ab was first split into 16 putative proteins based on its alignment with the human SARS polyprotein, with each protein independently searched against PDB. In total, structural templates for 17 proteins were chosen (Table 1, Figs. 1-3). In some cases, several independent templates, each covering an individual protein domain, of a target SARS-CoV-2 protein are selected. The obtained template is used in the comparative modeling protocol, generating five models.

Each model is assessed using the DOPE statistical potential [74]; the best-scoring model is selected as a final prediction.

To model a protein-protein interaction complex, a multi-chain modeling protocol is used [39]. Specifically, we align the corresponding pairs of homologous proteins and combine them into a single alignment where the individual chains are separated by "/" symbol. The alignment is used as an input together with the multi-chain structural template of a homologous complex. In case of viral-human interaction (these are the only virus-host structural templates found), the human protein remains the same in the alignment. In total, 18 structural templates of protein complexes involving homologs of 14 SARS-CoV-2 proteins were retrieved (Fig. 3). Similar to the single-protein protocol, five candidate models are generated for each complex conformation, and each model is assessed using the DOPE statistical potential following selection of the best-scoring model as a final prediction.

Mapping of functional regions and evolutionary conservation

We next extract protein- and ligand-binding sites and map them onto the models of SARS-CoV-2 proteins. For protein binding sites, the obtained modeled structures of protein complexes that involve SARS-CoV-2 proteins are considered. For each SARS-CoV-2 protein in a protein-protein interaction complex, we identify all binding residues that constitute its protein-binding site. Given an interaction between two proteins, a residue on one protein is defined as a protein binding site residue if there is at least one pair of atoms, one from this residue and another from a residue in the second protein, with Van Der Waals (VDW) surfaces not farther than 1.0 Å. Using this definition, the binding sites are identified with UCSF Chimera [75]. The ligand binding sites are identified and mapped using a different protocol—we rely on the default definition of protein-ligand binding site residues from PDB 3D Ligand View, since this standard is widely accepted. Once all ligand binding site residues are identified for a related coronavirus, the residues are mapped onto the surface of the related SARS-CoV-2 protein guided by a

structural alignment between the two proteins, which puts residues from both proteins into a one-to-one correspondence.

Inferring intra-viral and virus-host protein-protein interaction networks

Next, we leverage the obtained protein homology information to predict and map all possible intra-viral and virus-host protein interactions at a systems level. Since the SARS-CoV-2 genome exhibits substantial similarity to the 2002 SARS-CoV genome [76] and proteome [77], we hypothesize that many of the interactions observed in the SARS proteome will be preserved in the SARS-CoV-2 proteome as well, unless the corresponding binding sites are affected. The information in the interactome will help us understand the global mechanistic processes of the viral molecular machinery during viral infection, survival within the host, and replication. With this knowledge, we can discern the protein interactions that are crucial for transmission and replication—such interactions are potential candidates for inhibitory drugs [78]. Furthermore, using the systems approach, we can identify hubs and bottlenecks new to SARS-CoV-2 that could again be targeted by the antiviral drugs.

For this purpose, we create a comprehensive integrated SARS-CoV interactome that consists of both intra-viral and virus-host interactions. The SARS intra-viral interactome was created using published data, while the SARS-CoV ORFeome was cloned and a genome-wide analysis of viral protein interactions through yeast-two-hybrid (Y2H) matrix screens was performed [61]. The Y2H matrix screen was summarized by combining interactions in one direction, both directions and self-interaction. We also included intra-viral interactions gathered from the literature review [61]. The aggregated intra-viral interaction network consists of 31 proteins and 86 unique interactions.

We then construct the SARS-CoV-Host interactome through published interaction data, where first the SARS-CoV ORFeome was cloned for Y2H screens and used as a bait [64]. Specifically, a cDNA library encoding 5,000 different human genes were used that acted as prey molecules. We also include virus-host interactions mined from literature survey of 5,000 abstracts [64]. The curated virus-host interaction network consists of 118 proteins, including 93 host proteins, and 114 unique virus-host interactions. Next, to create an integrated network, the two individual networks were imported in Cytoscape [79] and merged to form a unified interactome representing both intra-viral and virus-host interactions. Finally, we include our predictions from structural modeling of SARS-CoV-2 intra-viral and virus-host interactions, surveyed recent literature on predicted SARS-CoV-2 interactions and annotated them accordingly in the unified interactome. The unified interactome consists of 125 proteins (94 host proteins) and 200 unique interactions. In addition, network analysis is performed to compute a set of topological statistics (degree distribution, clustering coefficient, and other important characteristics) to characterize parameters of the three networks, before which the networks are pruned for duplicate edges. Finally, the 200 putative interactions were annotated based on the extent to which the protein binding sites of the SARS-CoV-2 proteins were altered, compared to their SARS homologs. Based on the evolutionary conservation analysis of the putative protein binding sites extracted from the modeled complexes, some interactions were annotated as potentially disrupted.

Availability: All data generated in this work is freely available to the research community via Korkin Lab web-server (korkinlab.org/wuhan).

Funding

This work was supported by National Science Foundation (1458267) and National Institute of Health (LM012772–01A1) to D.K.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

Figures

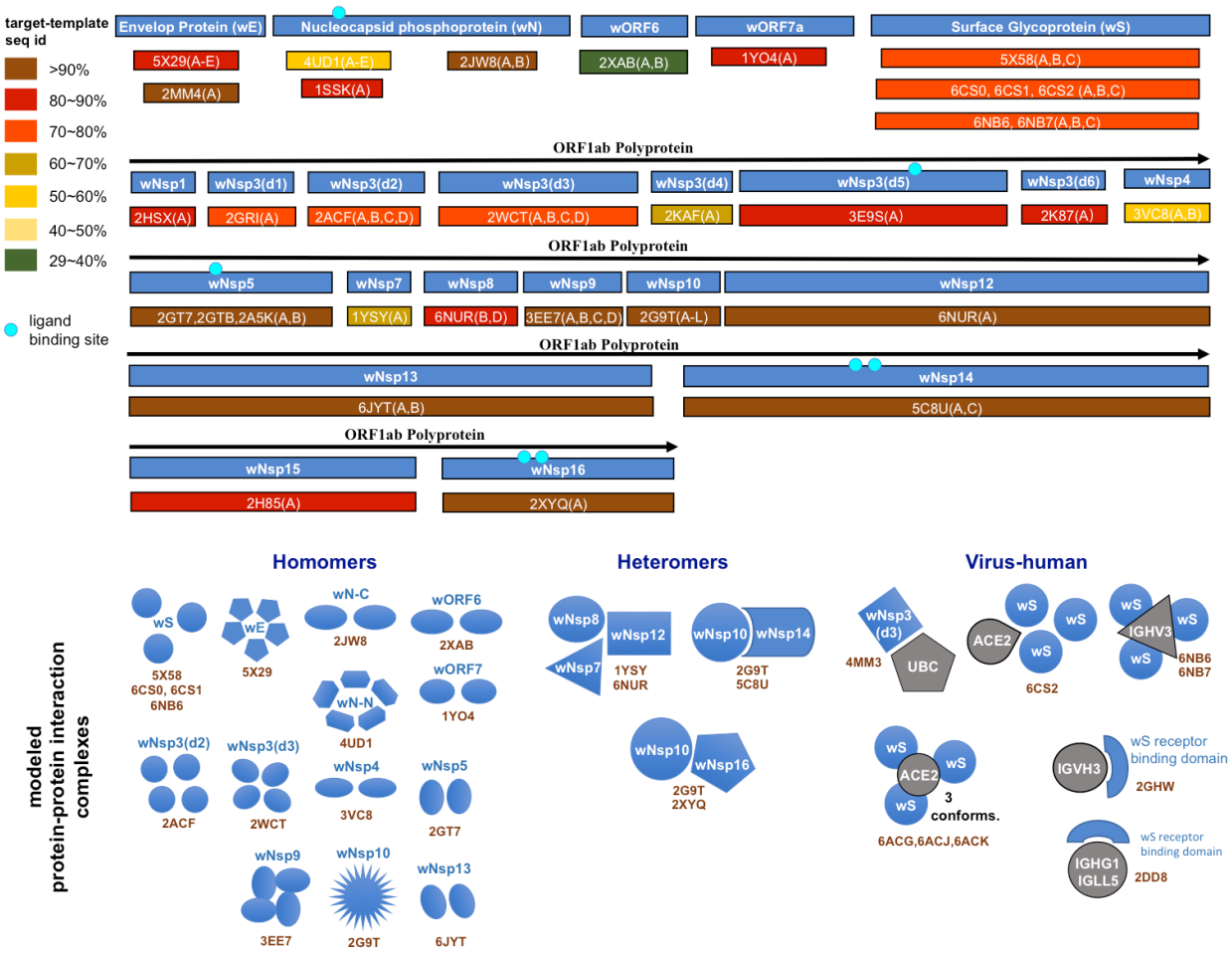

Figure 1. Structural genomics and interactomics road map. Shown are the individual proteins and protein complexes targeted for structural characterization together with PDB ID of their templates from Protein Data Bank (PDB).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

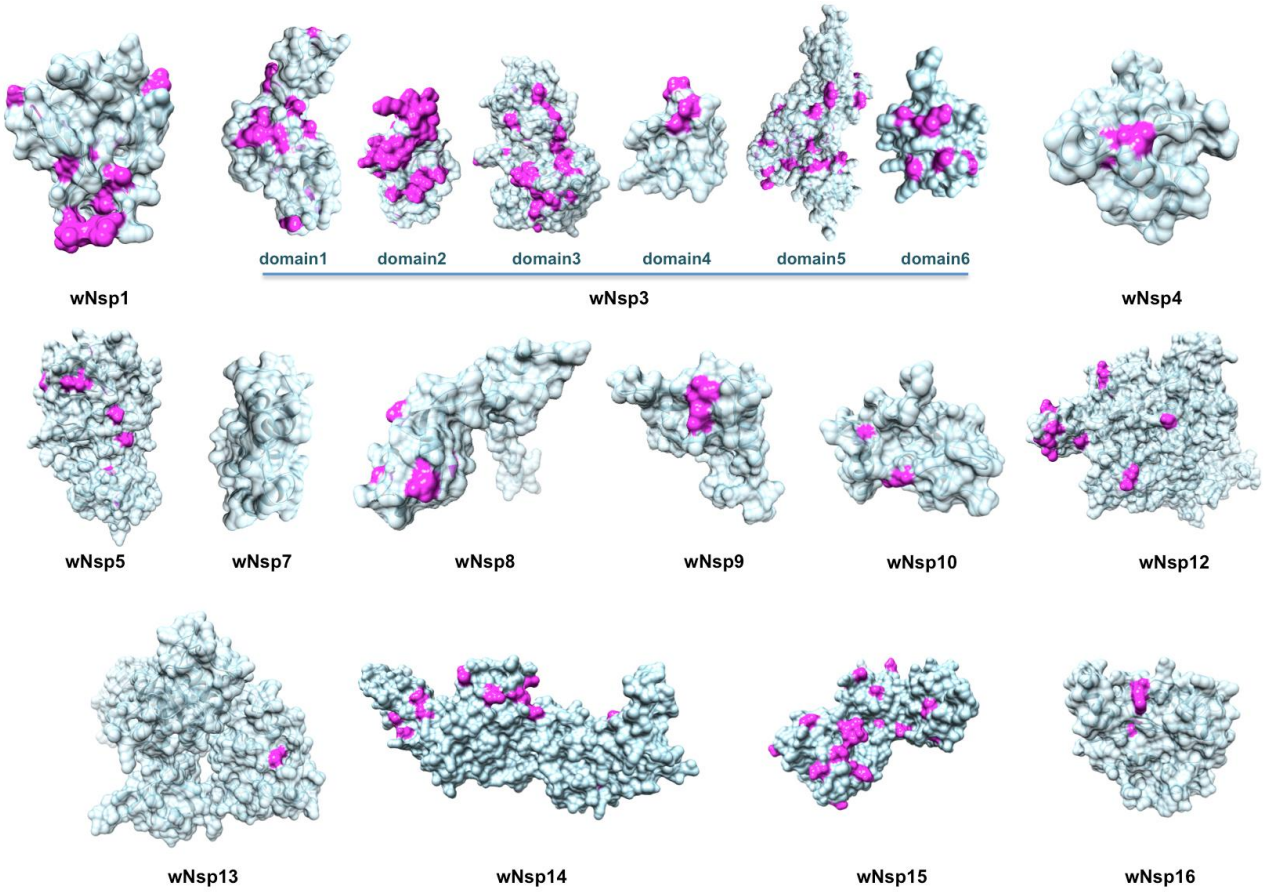

Figure 2. Structurally characterized non-structural proteins of SARS-CoV-2. Highlighted in pink are mutations found when aligning the proteins against their homologs from the closest related coronaviruses: human SARS, bat coronavirus, and another bat betacoronavirus BtRf-BetaCoV. The structurally resolved part of wNsp7 shares 100% sequence identity to its homolog.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

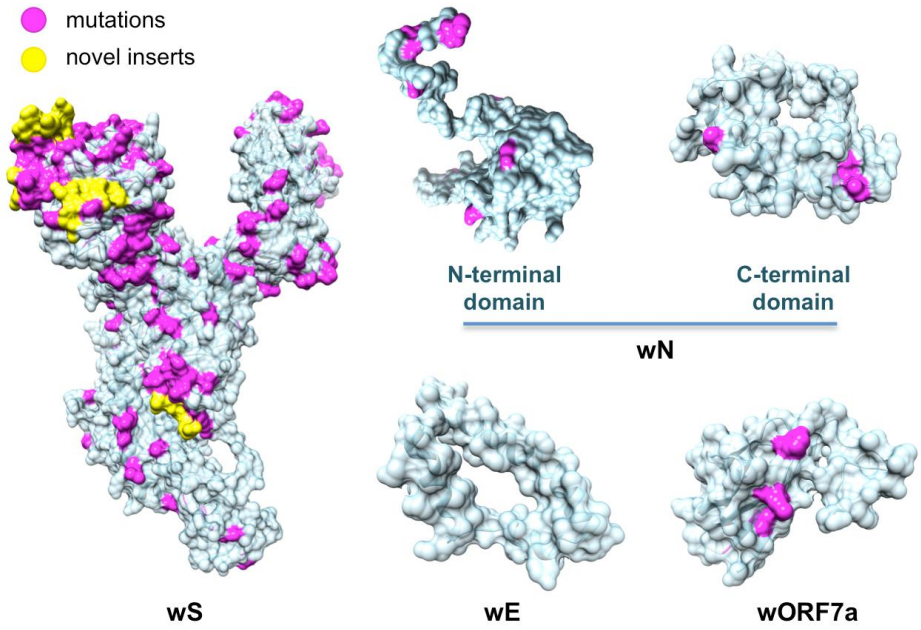

Figure 3. Structurally characterized structural proteins and an ORF of SARS-CoV-2. Highlighted in pink are mutations found when aligning the proteins against their homologs from the closest related coronaviruses: human SARS, bat coronavirus, and another bat betacoronavirus BtRf-BetaCoV. Highlighted in yellow are novel protein inserts found in wS.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

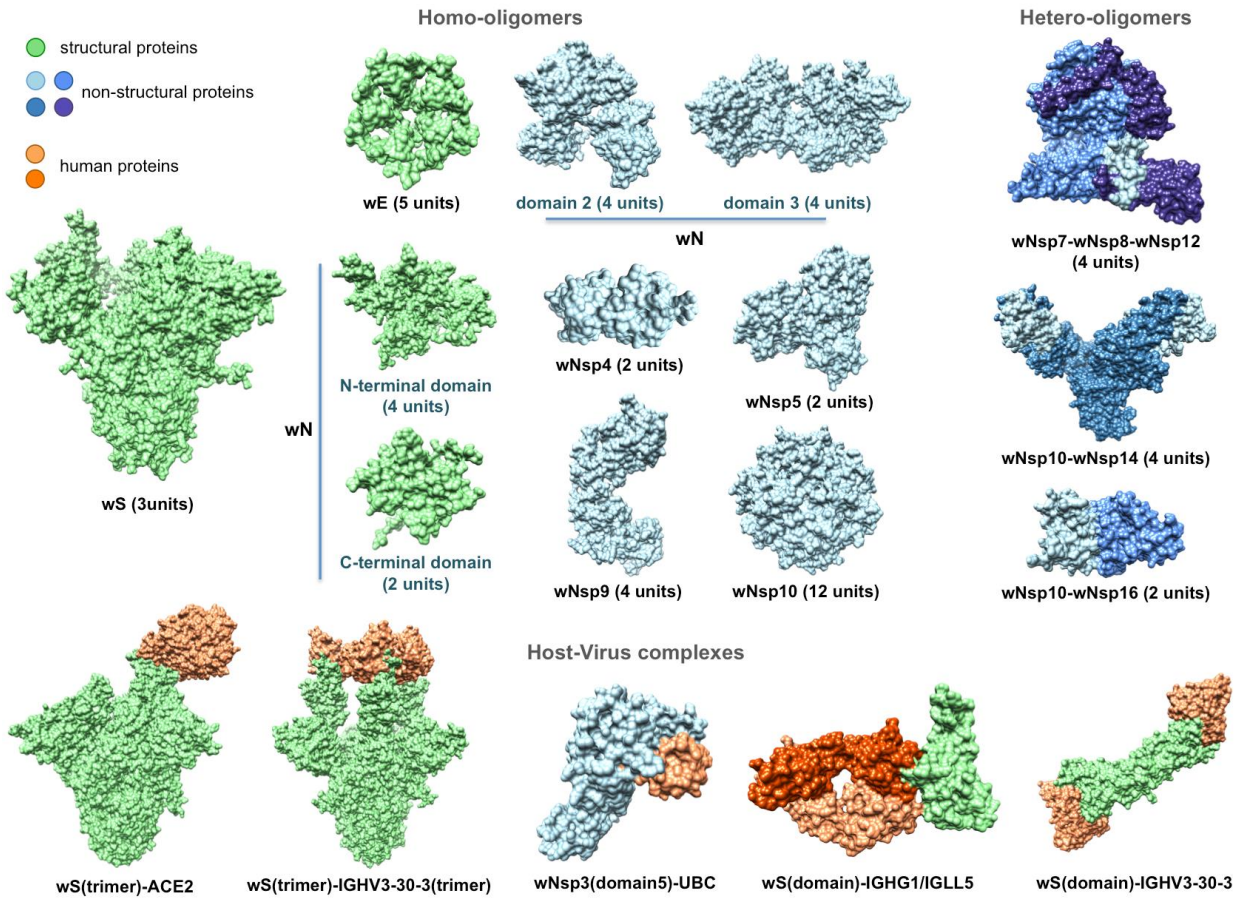

Figure 4. Structurally characterized intra-viral and host-viral protein-protein interaction complexes of SARS-CoV-2. Human proteins (colored in orange) are identified through their gene names. For each intra-viral structure, the number of subunits involved in the interaction is specified.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

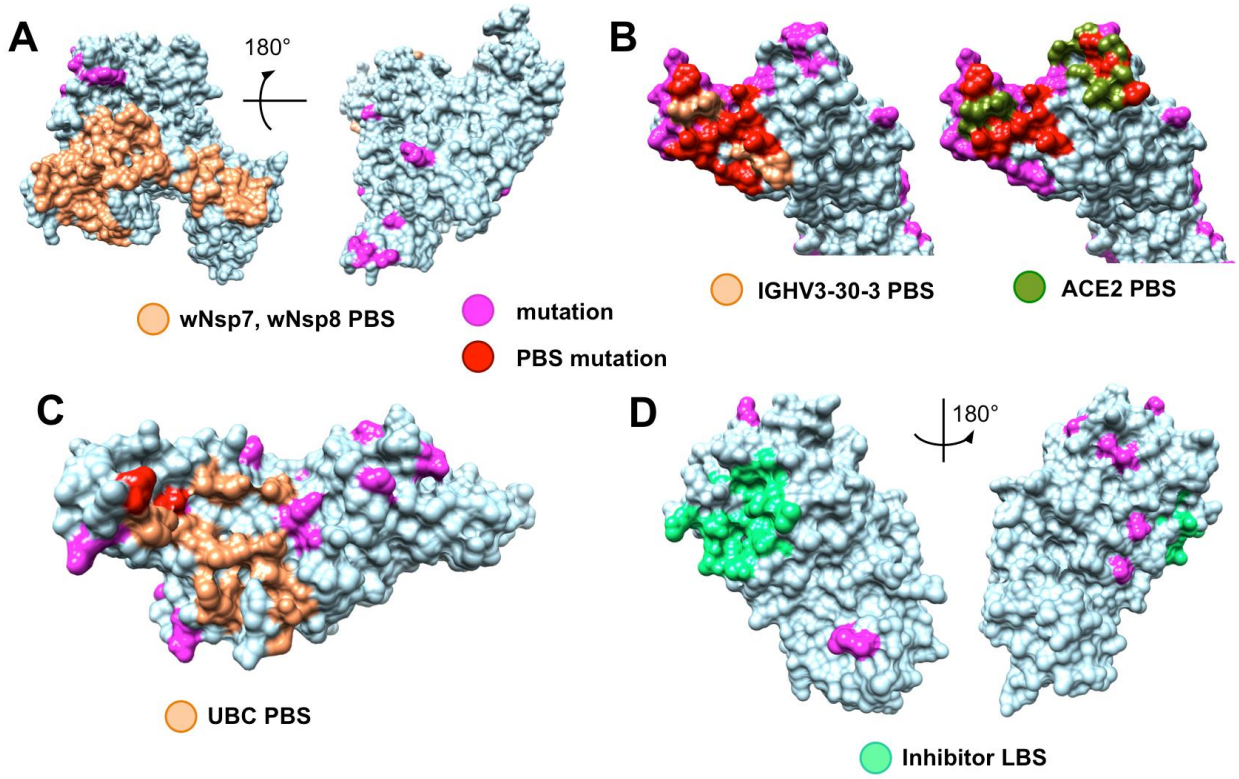

Figure 5. Evolutionary conservation of functional sites in SARS-CoV-2 proteins. A. Fully conserved protein binding sites (PBS, light orange) of wNsp12 in its interaction with wNsp7 and wNsp8 while other parts of the protein surface shows mutations (magenta); B. Both major monoclonal antibody binding site (light orange) and ACE2 receptor binding site (dark green) of wS are heavily mutated (binding site mutations are shown in red) compared to the same binding sites in other coronaviruses; mutations not located on the two binding sites are shown in magenta; C. Nearly intact protein binding site (light orange) of wNsp (papain-like protease PLpro domain) for its putative interaction with human ubiquitin-aldehyde (binding site mutations for the only two residues are shown in red, non-binding site mutations are shown in magenta); D. Fully conserved inhibitor ligand binding site (LBS, green) for wNsp5; non-binding site mutations are shown in magenta.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

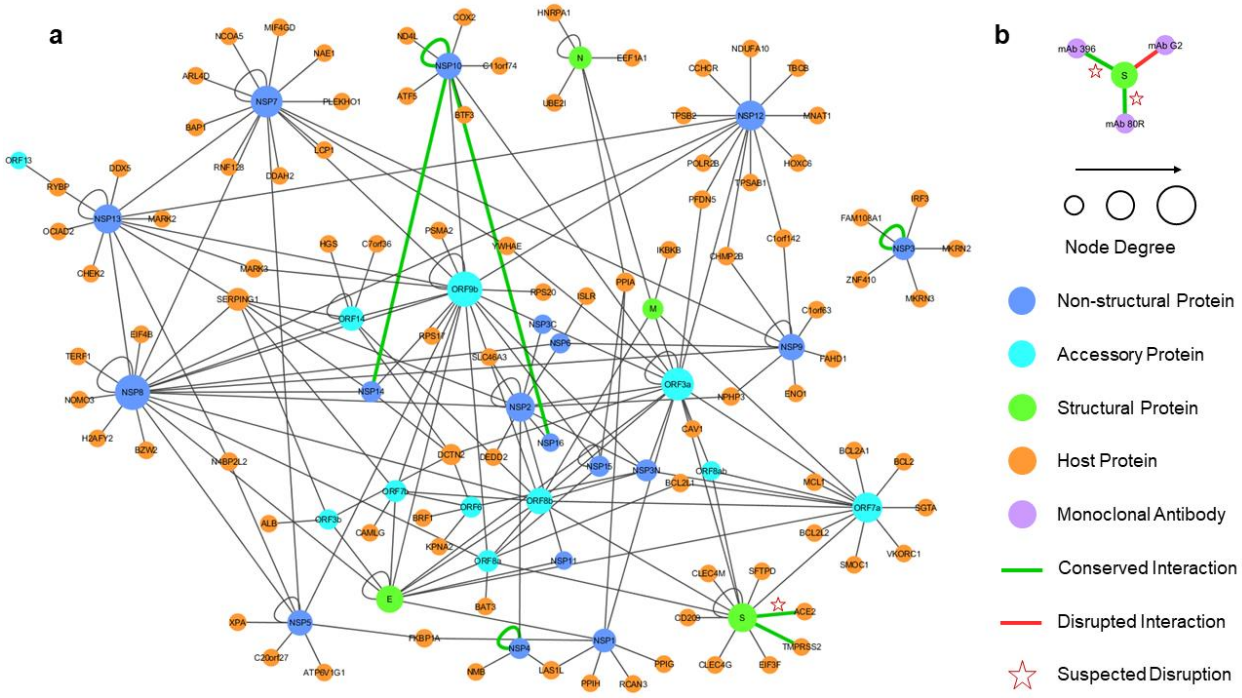

Figure 6. The unified interactome with SARS-CoV-2 interactions inferred through homology. (a) The SARS-CoV intra-viral and virus-host interactions are merged to create a unified interaction network indicating the types of proteins; the size of a node reflects the node degree. The SARS-CoV-2 interactions that are inferred from SARS-CoV are represented by green edges, with stars indicating the suspected disruption of the interaction based on the evolutionary conservation of the binding sites. (b) Structural modeling prediction of interaction between SARS-CoV-2 S protein and three monoclonal antibodies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0372.v1

Peer-reviewed version available at *Viruses* **2020**, *12*, 360; doi:10.3390/v12040360

Tables

Table 1. The list of SARS-CoV-2 proteins analyzed and structurally characterized in this work.

protein	accession	length	template PDB id	Trgt-tmplt Seq ID	organism
wS, surface glycoprotein	YP_009724390	1273	6ACK	75%	SARS
wE, envelope protein	YP_009724392	75	5X29	89%	SARS
wORF7a	YP_009724395	121	1YO4	90%	SARS
wN, nucleocapsid phosphoprotein	YP_009724397	419	2JW8	96%	SARS
			1SSK	83%	SARS
			4UD1	51%	MERS
wNsp1	YP_009725297	115	2HSX	86%	SARS
wNsp3-domain1	YP_009725299	107	2GRI	79%	SARS
wNsp3-domain2	YP_009725299	175	2ACF	72%	SARS
wNsp3-domain3	YP_009725299	264	2WCT	76%	SARS
wNsp3-domain4	YP_009725299	67	2KAF	70%	SARS
wNsp3-domain5	YP_009725299	315	3E9S	82%	SARS
wNsp3-domain6	YP_009725299	113	2K87	82%	SARS
wNsp4	YP_009725300	91	3VC8	60%	MHV
wNsp5	YP_009725301	306	2GT7	96%	SARS
wNsp7	YP_009725302	83	1YSY	67%	SARS
wNsp8	YP_009725304	114	6NUR	85%	SARS
wNsp9	YP_009725305	110	3EE7	99%	SARS
wNsp10	YP_009725306	121	2G9T	98%	SARS
wNsp12	YP_009725307	770	6NUR	97%	SARS
wNsp13	YP_009725308	596	6JYT	100%	SARS
wNsp14	YP_009725309	526	5C8U	95%	SARS
wNsp15	YP_009725310	346	2H85	86%	SARS
wNsp16	YP_009725311	288	2XYQ	94%	SARS

Table 2. Network Parameters of the SARS-CoV intra-viral, virus-host and unified networks. The table shows topological statistics for the three networks. Among the many computed statistics, the shown parameters include the number of nodes and edges in the networks, the average degree, number of components (independent networks), diameter (maximum shortest path), and clustering coefficient.

Network Parameters	SARS-CoV Intra-viral Interactome	SARS-CoV-Host Interactome	Unified Interactome
No. of nodes	31	118	125
No. of edges	86	114	206
No. of components	1	8	2
Diameter	4	14	7
Average degree	4.710	1.95	3.04
Clustering coefficient	0.448	0.0	0.068

References

1. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.* The Lancet, 2020.
2. Hui, D.S., et al., *The continuing SARS-CoV-2 epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China.* International Journal of Infectious Diseases, 2020. **91**: p. 264-266.
3. Zhu, N., et al., *A novel coronavirus from patients with pneumonia in China, 2019.* New England Journal of Medicine, 2020.
4. Nabel, G.J., *Designing tomorrow's vaccines.* New England Journal of Medicine, 2013. **368**(6): p. 551-560.
5. Plotkin, S.A. and S.L. Plotkin, *The development of vaccines: how the past led to the future.* Nature Reviews Microbiology, 2011. **9**(12): p. 889-893.
6. Martin, J.E., et al., *A SARS DNA vaccine induces neutralizing antibody and cellular immune responses in healthy adults in a Phase I clinical trial.* Vaccine, 2008. **26**(50): p. 6338-6343.
7. Zumla, A., et al., *Coronaviruses—drug discovery and therapeutic options.* Nature reviews Drug discovery, 2016. **15**(5): p. 327.
8. Zumla, A., et al., *Vaccine against Middle East respiratory syndrome coronavirus.* The Lancet Infectious Diseases, 2019. **19**(10): p. 1054-1055.
9. Cotten, M., et al., *Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study.* The Lancet, 2013. **382**(9909): p. 1993-2002.
10. de Wit, E., et al., *SARS and MERS: recent insights into emerging coronaviruses.* Nature Reviews Microbiology, 2016. **14**(8): p. 523.
11. Dyall, J., et al., *Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection.* Antimicrobial agents and chemotherapy, 2014. **58**(8): p. 4885-4893.
12. Li, F., *Structure, function, and evolution of coronavirus spike proteins.* Annual review of virology, 2016. **3**: p. 237-261.
13. Lu, G., et al., *Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26.* Nature, 2013. **500**(7461): p. 227-231.
14. Luk, H.K., et al., *Molecular epidemiology, evolution and phylogeny of SARS coronavirus.* Infection, Genetics and Evolution, 2019.
15. Millet, J.K. and G.R. Whittaker, *Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells.* Virology, 2018. **517**: p. 3-8.
16. Song, W., et al., *Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2.* PLoS pathogens, 2018. **14**(8): p. e1007236.
17. Walls, A.C., et al., *Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer.* Nature, 2016. **531**(7592): p. 114-117.
18. Yuan, Y., et al., *Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains.* Nature communications, 2017. **8**: p. 15092.
19. Cho, C.-C., et al., *Macro Domain from Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Is an Efficient ADP-ribose Binding Module CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES.* Journal of Biological Chemistry, 2016. **291**(10): p. 4894-4902.
20. Gui, M., et al., *Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding.* Cell research, 2017. **27**(1): p. 119-129.

21. Jacobs, J., et al., *Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease.* Journal of medicinal chemistry, 2013. **56**(2): p. 534-546.
22. Jia, Z., et al., *Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis.* Nucleic acids research, 2019. **47**(12): p. 6538-6550.
23. Kankanamalage, A.C.G., et al., *Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element.* European journal of medicinal chemistry, 2018. **150**: p. 334-346.
24. Kirchdoerfer, R.N., et al., *Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis.* Scientific reports, 2018. **8**(1): p. 1-11.
25. Kirchdoerfer, R.N. and A.B. Ward, *Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors.* Nature communications, 2019. **10**(1): p. 1-9.
26. Li, Y., et al., *A humanized neutralizing antibody against MERS-CoV targeting the receptor-binding domain of the spike protein.* Cell research, 2015. **25**(11): p. 1237-1249.
27. Ma, Y., et al., *Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex.* Proceedings of the National Academy of Sciences, 2015. **112**(30): p. 9436-9441.
28. Ratia, K., et al., *Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease.* PLoS pathogens, 2014. **10**(5).
29. Shimamoto, Y., et al., *Fused-ring structure of decahydroisoquinolin as a novel scaffold for SARS 3CL protease inhibitors.* Bioorganic & Medicinal Chemistry, 2015. **23**(4): p. 876-890.
30. Su, D., et al., *Dodecamer structure of severe acute respiratory syndrome coronavirus non-structural protein nsp10.* Journal of virology, 2006. **80**(16): p. 7902-7908.
31. Wang, N., et al., *Structural Definition of a Neutralization-sensitive Epitope on the MERS-CoV S1-NTD.* Cell reports, 2019. **28**(13): p. 3395-3405. e6.
32. Martí-Renom, M.A., et al., *Comparative protein structure modeling of genes and genomes.* Annual review of biophysics and biomolecular structure, 2000. **29**(1): p. 291-325.
33. Cavasotto, C.N. and S.S. Phatak, *Homology modeling in drug discovery: current trends and applications.* Drug discovery today, 2009. **14**(13-14): p. 676-683.
34. Burley, S.K., et al., *Structural genomics: beyond the human genome project.* Nature genetics, 1999. **23**(2): p. 151-157.
35. Yan, L., et al., *Assessment of putative protein targets derived from the SARS genome.* FEBS letters, 2003. **554**(3): p. 257-263.
36. Wichapong, K., et al., *Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction.* Journal of Molecular Recognition: An Interdisciplinary Journal, 2010. **23**(3): p. 283-300.
37. Ekins, S., et al., *Illustrating and homology modeling the proteins of the Zika virus.* F1000Research, 2016. **5**.
38. Prabakaran, P., X. Xiao, and D.S. Dimitrov, *A model of the ACE2 structure and function as a SARS-CoV receptor.* Biochemical and biophysical research communications, 2004. **314**(1): p. 235-241.
39. Davis, F.P., et al., *Host–pathogen protein interactions predicted by comparative modeling.* Protein Science, 2007. **16**(12): p. 2585-2596.
40. Russell, R.B., et al., *A structural perspective on protein–protein interactions.* Current opinion in structural biology, 2004. **14**(3): p. 313-324.
41. Zhang, Q.C., et al., *Structure-based prediction of protein–protein interactions on a genome-wide scale.* Nature, 2012. **490**(7421): p. 556-560.

42. Cavasotto, C.N., et al., *Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening.* Journal of medicinal chemistry, 2008. **51**(3): p. 581-588.
43. Wang, S.-Q., Q.-S. Du, and K.-C. Chou, *Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases.* Biochemical and biophysical research communications, 2007. **354**(3): p. 634-640.
44. Loewenstein, Y., et al., *Protein function annotation by homology-based inference.* Genome biology, 2009. **10**(2): p. 207.
45. Durmuş, S. and K.Ö. Ülgen, *Comparative interactomics for virus–human protein–protein interactions: DNA viruses versus RNA viruses.* FEBS open bio, 2017. **7**(1): p. 96-107.
46. Zhang, A., L. He, and Y. Wang, *Prediction of GCRV virus-host protein interactome based on structural motif-domain interactions.* BMC bioinformatics, 2017. **18**(1): p. 145.
47. Vidalain, P.-O. and F. Tangy, *Virus-host protein interactions in RNA viruses.* Microbes and infection, 2010. **12**(14-15): p. 1134-1143.
48. Warren, S., et al., *Extreme evolutionary conservation of functionally important regions in H1N1 influenza proteome.* PloS one, 2013. **8**(11).
49. Wang, L., et al., *From mosquitos to humans: genetic evolution of Zika virus.* Cell host & microbe, 2016. **19**(5): p. 561-565.
50. Patel, H. and A. Kukol, *Prediction of ligands to universally conserved binding sites of the influenza a virus nuclear export protein.* Virology, 2019. **537**: p. 97-103.
51. Sawicki, S.G., D.L. Sawicki, and S.G. Siddell, *A contemporary view of coronavirus transcription.* Journal of virology, 2007. **81**(1): p. 20-29.
52. *UniProt: the universal protein knowledgebase.* Nucleic acids research, 2017. **45**(D1): p. D158-D169.
53. Saikatendu, K.S., et al., *Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1 ”-phosphate dephosphorylation by a conserved domain of nsP3.* Structure, 2005. **13**(11): p. 1665-1675.
54. Serrano, P., et al., *Nuclear magnetic resonance structure of the N-terminal domain of non-structural protein 3 from the severe acute respiratory syndrome coronavirus.* Journal of virology, 2007. **81**(21): p. 12049-12060.
55. Johnson, M., et al., *NCBI BLAST: a better web interface.* Nucleic acids research, 2008. **36**(suppl_2): p. W5-W9.
56. Cheng, S. and C.L. Brooks III, *Viral capsid proteins are segregated in structural fold space.* PLoS computational biology, 2013. **9**(2).
57. Patel, H. and A. Kukol, *Evolutionary conservation of influenza A PB2 sequences reveals potential target sites for small molecule inhibitors.* Virology, 2017. **509**: p. 112-120.
58. Zhu, Z., et al., *Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies.* Proceedings of the National Academy of Sciences, 2007. **104**(29): p. 12123-12128.
59. Sui, J., et al., *Effects of human anti-spike protein receptor binding domain antibodies on severe acute respiratory syndrome coronavirus neutralization escape and fitness.* Journal of virology, 2014. **88**(23): p. 13769-13780.
60. Coughlin, M.M., J. Babcook, and B.S. Prabhakar, *Human monoclonal antibodies to SARS-coronavirus inhibit infection by different mechanisms.* Virology, 2009. **394**(1): p. 39-46.
61. von Brunn, A., et al., *Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome.* PloS one, 2007. **2**(5).

62. Shi, C.-S., et al., *SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome.* The Journal of Immunology, 2014. **193**(6): p. 3080-3089.
63. Subissi, L., et al., *One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities.* Proceedings of the National Academy of Sciences, 2014. **111**(37): p. E3900-E3909.
64. Pfefferle, S., et al., *The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors.* PLoS pathogens, 2011. **7**(10).
65. Wan, Y., et al., *Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS.* Journal of Virology, 2020.
66. Glowacka, I., et al., *Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response.* Journal of virology, 2011. **85**(9): p. 4122-4134.
67. Hoffmann, M., et al., *The novel coronavirus 2019 (SARS-CoV-2) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells.* bioRxiv, 2020.
68. Kruse, R.L., *Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China.* F1000Research, 2020. **9**(72): p. 72.
69. Voitenko, O.S., et al., *Patterns of amino acid conservation in human and animal immunodeficiency viruses.* Bioinformatics, 2016. **32**(17): p. i685-i692.
70. Brister, J.R., et al., *NCBI viral genomes resource.* Nucleic acids research, 2015. **43**(D1): p. D571-D577.
71. Sievers, F. and D.G. Higgins, *Clustal Omega for making accurate alignments of many protein sequences.* Protein Science, 2018. **27**(1): p. 135-145.
72. Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Current protocols in bioinformatics, 2006. **15**(1): p. 5.6. 1-5.6. 30.
73. Berman, H.M., et al., *The protein data bank*, in *Protein Structure*. 2003, CRC Press. p. 394-410.
74. Shen, M.y. and A. Sali, *Statistical potential for assessment and prediction of protein structures.* Protein science, 2006. **15**(11): p. 2507-2524.
75. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
76. Chan, J.F.-W., et al., *Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan.* Emerging Microbes & Infections, 2020. **9**(1): p. 221-236.
77. Dong, N., et al., *Genomic and protein structure modelling analysis depicts the origin and infectivity of SARS-CoV-2, a new coronavirus which caused a pneumonia outbreak in Wuhan, China.* bioRxiv, 2020.
78. De Chassey, B., et al., *Virus-host interactomics: new insights and opportunities for antiviral drug discovery.* Genome medicine, 2014. **6**(11): p. 115.
79. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.