PLOS ONE

RESEARCH ARTICLE

The analysis on the human protein domain targets and host-like interacting motifs for the MERS-CoV and SARS-CoV/CoV-2 infers the molecular mimicry of coronavirus

Yamelie A. Martínez[1,2], Xianwu Guo[3], Diana P. Portales-Pérez[2], Gildardo Rivera[4], Julio E. Castañeda-Delgado[1,5], Carlos A. García-Pérez[6], José A. Enciso-Moreno[1], Edgar E. Lara-Ramírez[1]*

1 Unidad de Investigación Biomédica de Zacatecas, Instituto Mexicano Del Seguro Social, Zacatecas, México, **2** Laboratorio de Inmunología y Biología Celular y Molecular, Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México, **3** Laboratorio de Biotecnología Genómica, Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Reynosa, México, **4** Laboratorio de Biotecnología Farmacéutica, Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Reynosa, México, **5** Cátedras-CONACYT, Unidad de Investigación Biomédica de Zacatecas, Instituto Mexicano Del Seguro Social, Zacatecas, México, **6** Information and Communication Technology Department (ICT), Complex Systems, Helmholtz Zentrum München, Neuherberg, Germany

* doc_lara_ram@hotmail.com

Citation: Martínez YA, Guo X, Portales-Pérez DP, Rivera G, Castañeda-Delgado JE, García-Pérez CA, et al. (2021) The analysis on the human protein domain targets and host-like interacting motifs for the MERS-CoV and SARS-CoV/CoV-2 infers the molecular mimicry of coronavirus. PLoS ONE 16(2): e0246901. https://doi.org/10.1371/journal.pone.0246901

Editor: Eduardo Andrés-León, Institute of Parasitology and Biomedicine, SPAIN

Received: November 30, 2020

Accepted: January 28, 2021

Published: February 17, 2021

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The authors received no specific funding for this work. Yamelie A. Martínez holds a doctoral CONACYT scholarship (No. CVU 934987).

Competing interests: The authors have declared that no competing interests exist.

Abstract

The MERS-CoV, SARS-CoV, and SARS-CoV-2 are highly pathogenic viruses that can cause severe pneumonic diseases in humans. Unfortunately, there is a non-available effective treatment to combat these viruses. Domain-motif interactions (DMIs) are an essential means by which viruses mimic and hijack the biological processes of host cells. To disentangle how viruses achieve this process can help to develop new rational therapies. Data mining was performed to obtain DMIs stored as regular expressions (regexp) in 3DID and ELM databases. The mined regexp information was mapped on the coronaviruses' proteomes. Most motifs on viral protein that could interact with human proteins are shared across the coronavirus species, indicating that molecular mimicry is a common strategy for coronavirus infection. Enrichment ontology analysis for protein domains showed a shared biological process and molecular function terms related to carbon source utilization and potassium channel regulation. Some of the mapped motifs were nested on B, and T cell epitopes, suggesting that it could be as an alternative way for reverse vaccinology. The information obtained in this study could be used for further theoretic and experimental explorations on coronavirus infection mechanism and development of medicines for treatment.

Introduction

Coronaviruses (CoV) are enveloped single-stranded, positive-sense RNA viruses, responsible very often for mild upper respiratory infections in humans. Nevertheless, remarkably

pathogenic CoVs to humans have been reported. The first one appeared in 2003 in Guangdong, China, leading to an epidemic of severe acute respiratory syndrome (SARS) and this virus was named SARS-CoV [1]. In 2012, another CoV arose in Middle Eastern countries, causing pneumonic syndrome, called MERS-CoV [2]. At the end of 2019, a new CoV emerged in Wuhan, China, causing severe pneumonia [3] and was named SARS-CoV-2 due to its genomic similarity with the past SARS-CoV [4]. This is the first CoV that caused a pandemic disease termed COVID-19. These three CoVs are zoonotic, and its primary origin was traced to bats and other animals [4, 5]. We are still suffering from SARS-CoV-2. This is a serious public health concern, especially for the aged people with increased risk for complications such diabetes mellitus (DM), hypertension, and severe obesity, which cause the high morbidity-mortality rates of COVID-19 [6]. Humans infected by SARS-CoV-2 could be also asymptomatic, but they may transmit the virus [6]. Although numerous efforts are currently underway to develop drugs and vaccines to combat those viruses, there is no effective treatment available yet.

The study on molecular interactions of host-pathogen helps to find new targets for drug discovery or antigens for vaccine development. Host-pathogen relation is mainly explored through protein-protein interaction (PPI) studies. These studies can be experimentally and computationally aided [7]. The computational studies could be preliminary but quick to guide the rational selection of data for experimental confirmations. Experimental approaches have been carried out for SARS-CoV, MERS-CoV [8, 9], and recently for SARS-CoV-2 [10]. A detailed literature mining that surveys experimental and predicted PPIs for several coronaviruses, including the viruses studied herein, was recently published [11]. Also, several computation-aided researches focused on predicting PPI of host and SARS-CoV-2 [7, 12, 13]. Such predictions provided valuable information to help the rational design of treatments against these viral infections.

However, the analysis of domain-motif interaction (DMI) has paid less attention to those CoVs. Domains in proteins are the functional units involved in the signaling networks within a cell [14]. Its length is up to 200 amino acids, and its folding patterns are independent of the rest of the whole protein [15]. In contrast, motifs are short plastic linear sequences with a length of 3 to 15 amino acids. DMIs are the preferential molecular mechanism by which viruses interact with host cells [16]. Motifs are employed by the viruses to mimic and hijack the host cell's essential process for its survival [17]. Currently, two studies have approached the role of motifs present on essential host proteins for SARS-CoV-2 infection. The research of Mészáros et al. [18] consisted in the prediction of motifs retrieved from Eukaryotic Linear Motif (ELM) resource that were mapped onto the angiotensin-converting enzyme 2 (ACE2) and integrins of the human host. They found conserved motifs on the cytoplasmatic tails of ACE2 and integrin β3 that interacts with several critical regulatory protein domains. This motif information was tested later on experimental binding affinity measurements [19] and found that NHERF3 PDZ1, SHANK1 and SNX27 PDZ domains bind to synthetic peptides of the ACE2, and to the synthetic ATG8 domains, MAP1LC3s and GABARAPs, of integrin β3. Those studies exemplify the utility of motif predictions to guide experimental proposals.

Here contrariwise to the previous researches, we focused on the motifs mapped on the MERS-CoV, SARS-CoV, and SARS-CoV-2 proteomes linked to human protein domains. The frequently matched motifs were compared among the coronaviruses. The motif functionality was inferred through enrichment ontology analysis of its partner domains. The based-motif information obtained could be used as the starting point to develop new therapies to combat these viruses in the future.

Table 1. The total number of non-redundant viral protein sequences for analysis.

Protein	MERS-CoV	SARS-CoV	SARS-CoV-2
ORF1ab	162	4	4003
ORF1a	140		
S	98	5	1135
ORF3a	25	3	421
NS4a	22		
NS4b	36		
NS5	21		
E	6	1	45
M	18	6	125
ORF6		1	73
ORF7a		1	149
ORF8	18	1	146
N	44	1	539
ORF10		1	35
TOTAL	590	24	6671

https://doi.org/10.1371/journal.pone.0246901.t001

Materials and methods

Protein sequence retrieval

The SARS-CoV (taxid:694009) and SARS-CoV-2 (taxid:2697049) sequences were retrieved from the NCBI virus repository (accessed on 01 September 2020) [20] using available predefined filters, such as human for host, length of proteins, and the completeness option for sequences. These sequences were firstly filtered based on its report date; then, sequences before 2019 were put on the SARS-CoV dataset. The redundant amino acid sequences were removed with the perl program "fasta_uniqueseq.pl" obtained from FASTA Tool list web page (http://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html.ncbi/fasta/list.html). The sequences for MERS-CoV were retrieved from the virus variation database [21], using the options as human host, sequence completeness, and collapse for removing redundant sequences. The final number of each viral protein in the datasets ordered by its arrangement on the genome are shown in Table 1. The SARS-CoV protein sequences were grouped together with the SARS-CoV-2 dataset for the analysis due to its small number after eliminating the redundant sequences.

Domain-motif data mining process

Our data mining process is based on our previous reported methodology [22], adapted to the data retrieved for the MERS-CoV and SARS-CoV/CoV-2 viruses. It includes three main steps. 1) Literature search. First, we obtained the human genes associated to the SARS-CoV/CoV-2 and MERS-CoV related diseases with pubtator [23]. This tool allows searching in a straightforward manner the reporting genes related to the infections by these viral pathogens in the PubMed literature. These gene names were compared and unified with the information from a recent research published by Perrin-Cocon et al., [11] to form a list of unique gene names. This list was submitted into the UniProt database [24] to obtain the human UniProt IDs that match our query for the next process. 2) Pfam database [25] mining for human protein domains: From the Pfam we downloaded the latest version of the files "Pfam-A.regions.tsv" and "Pfam-A.clans.tsv". The obtained UniProt IDs that match on the Pfam-A.regions.tsv file were extracted to mine the Pfam-A.clans file. Thereby, it was obtained the Pfam accession, clan ID, Pfam ID, and Pfam description columns that contain information associated with our

UniProt ID list. 3) The domain-motif information was mined from the databases of three-Dimensional Interacting Domains (3DID) [26] and ELM [27]. The motif information for 3DID was retrieved from the 3DID-DMI flat 2019 version. From this file, the Pfam IDs, domain-motif name, and the regular expressions (regexp) were extracted and stored in local files which was used as the target file to draw out the information associated with the Pfam IDs previously obtained. In the ELM database, the information came from the files "elm_interaction_domains.tsv" and "elm_classes.tsv". The first file was the target file to match the Pfam accessions IDs and was then used to take out the domain-motif name, Pfam accession, and the associated regexp from the elm_classes.tsv file. Each regexp was used to match motif amino acid sequences in the protein datasets with the patmatch software [28]. We used linux terminal for each query with the bash command "for ID in 'cat file_of_IDs.txt'; do grep $ID target_file.txt; done > extracted_info_file.txt". The obtained files were also checked manually for concordance with the query IDs.

Identification of potential functional host-like viral motifs

The potential functional motif identification was based on the percentage of regexp that matches a specific amino acid sequence. To this end, we followed 70% cut-off match as in the previous study [29]. For example, a total of 4003 ORF1ab non-redundant sequences were retrieved for SARS-CoV-2; consequently, a regexp present in more than 70% of ORF1ab proteins signifies that a specific motif matched more than 2802 sequences. Those frequent motifs were also queried on shuffled sequences versions of each protein dataset that was produced with the "shuffleseq" function from the EMBOSS suite programs [30]. If those inferred motifs were found scarcely on the randomized sequences, it reinforces as functional motifs.

Protein domain enrichment analysis

The protein domain enrichment analysis was carried out with the dgOR package [31] for R statistical language. For this analysis, the Pfam accession numbers were used as input data and the first ten significant ($p < 0.05$) ontologies based on the hypergeometric test related to gene ontology biological process (GOBP) and Gene ontology molecular function (GOMF) were analyzed.

Identification of motifs as immune epitopes

The immune epitope database (IEDB) [32] was manually queried for motif sequences with ≥ 5 amino acids, setting the blast parameter of identity more than 70%, and selecting the options "human host", "all assay types", and the disease option "COVID-19 and Severe acute respiratory syndrome" as filters. This query analysis was omitted for the MERS-CoV because there is not available information for this pathogen on the IEDB.

Statistics

The statistics rests on descriptive statistics of the frequent motifs. The obtained information was analyzed by its conjunction and disjunction relationships based on the matching patterns. This analysis was carried out with the help of the web tool for the calculation and drawing of custom Venn diagrams (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Table 2. The total number of human gene names obtained from the PubMed literature and compared with Perrin-Cocon et al. [11].

	Present study	Present study ∩ Perrin-Cocon et al.,	Perrin-Cocon et al.,
MERS-CoV	55	10	7
SARS-CoV/CoV2	383	114	352

*∩ means the intersection in the conjunction-disjunction analysis.

https://doi.org/10.1371/journal.pone.0246901.t002

Results

Literature mining

After removing duplicate gene names among the reviewed publications (data in S1 File), 497 human genes for SARS-CoV/CoV-2 and 65 for MERS-CoV infection were found involved in pathogenesis (Table 2, data in S2 File). The comparison of our mined information with Perrin-Cocon et al [11] showed overlapped gene information (n = 124), and the newly acquired (n = 438), especially for the MERS-CoV viruses. After eliminating the duplicated the rest are theunique gene names (data in S2 File), which were used to search its corresponding UniProt IDs to mine the Pfam, 3DID, and ELM databases for the subsequent regexp match analysis.

Identification of functional viral protein motifs

The functional regions of proteins are either structured or disordered. However, the proteins of coronaviruses were found mainly ordered according to IUPRED (S1 Fig) [33]. For example, most amino acids of the largest protein ORF1ab and the spike (S) protein were found below the 0.5 score. However, few regions of viral protein were disordered, such as the nucleocapsid (N) protein. In this study, the whole regexp lists obtained from the 3DID and ELM databases (data in S3 File) were mapped on the whole viral protein sequences. The frequent (>70%) regexps that matched amino acid motifs are shown in Table 3 and the data in S4 File.

Table 3. Total number of motifs frequently matched by regexp.

	3DID			ELM		
Protein	**M-CoV**	**M-CoV ∩ S-CoV/CoV-2**	**S-CoV/CoV-2**	**M-CoV**	**M-CoV ∩ S-CoV/CoV-2**	**S-CoV/CoV-2**
ORF1ab	65	148	31	8	78	5
S	47	50	38	11	44	12
ORF3a	4	6	24	1	13	28
NS4a	14			25		
NS4b	46			35		
NS5	20			28		
E	19	0	5	6	5	4
M	9	5	15	11	18	9
ORF6			4			18
ORF7a			20			27
ORF8	23	1	14	8	7	15
N	23	27	31	9	27	14
ORF10			2			12
TOTAL	270	237	184	142	192	144

*∩ means the intersection in the conjunction-disjunction analysis.

https://doi.org/10.1371/journal.pone.0246901.t003

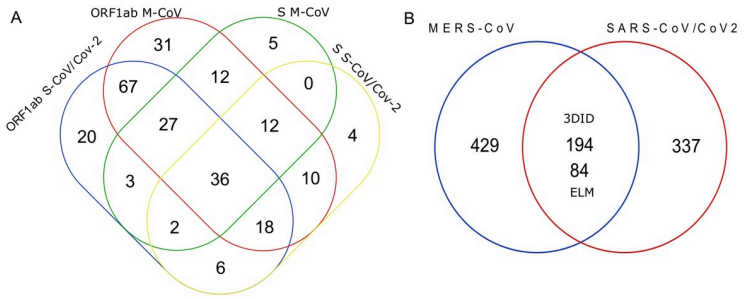

Fig 1. Venn diagrams show the redundant or non-redundant regexp motifs among the proteins or viral proteomes. (A) Venn diagram to show the redundant regexp numbers mapped on the ORF1ab and Spike proteins. (B) Venn diagram of total non-redundant regexp mapped in MERS-CoV and SARS-CoV-2 obtained from the two databases.

https://doi.org/10.1371/journal.pone.0246901.g001

The ORF1ab, S, and N sequences were matched by the regexp more than the other proteins from databases. A high number of motifs were shared among three CoVs in the ORF1ab (n = 148 and 78), followed by the S (n = 50 and 44) and the N (n = 27 and 27). The regexp motifs were redundant among the proteins or viral proteomes (data in S4 File); for example, the ORF1ab and S shared the same motifs (Fig 1A); and a high number of motifs shared between the MERS-CoV and SARS-CoV/CoV-2 after removing the redundant (Fig 1B, data in S5 File). Most of these motifs were scarcely on the shuffled sequences; thus, all were considered in the subsequent analysis.

Protein domain enrichment analysis for non-redundant motifs

First, it was examined the conjunction-disjunction relationships for the total number of Pfam accessions associated with non-redundant motifs described above. A total of 78 non-redundant domains were shared for MERS-CoV and SARS-CoV/CoV-2 irrespective of the database source, and few were specific to MERS-CoV (n = 8) and SARS-CoV/CoV-2 (n = 9) (Fig 2A, data in S5 File). Protein domain enrichment analysis of the 78 shared domains for GOBP identifies general terms related to metabolic and cellular processes. Five GOBP significant terms were related to energy reserve and glycogen biosynthesis metabolism (Fig 2B, data in S6 File). GOMF analysis also identifies five important terms related to channel regulation in which potassium channel regulator activity was the most significant (Fig 2C, data in S6 File). The study of specific domains for MERS-CoV and SARS-CoV-2 also showed terms associated with the same biological processes and molecular functions of the 78 shared domains. Thus, those domains could be the primary targets for molecular mimicry generated by MERS-CoV and SARS-CoV/CoV-2 to manipulate the host cell machinery.

Analysis of significant domains present on distinct host proteins

The analysis described above allows us to identify specific proteins linked to the domains involved with significant ontology terms. Four domains (Pfam accession ID: PF00656, PF00026, PF00082, PF00089) related to the glycogen biosynthetic process were present in 26 proteins that matched our gene lists. Among them, the PF00089 related to trypsin domain function is the more promiscuous present on most of the proteins (Fig 3A). This domain was associated with the protease TMPRSS2, an endothelial cell surface protein involved in the entry and spread of CoVs and influenza virus [34], so that this protein has been proposed as a potential drug target to combat those viruses. It was also found the domains associated with the potassium channel regulator activity (Fig 3B).

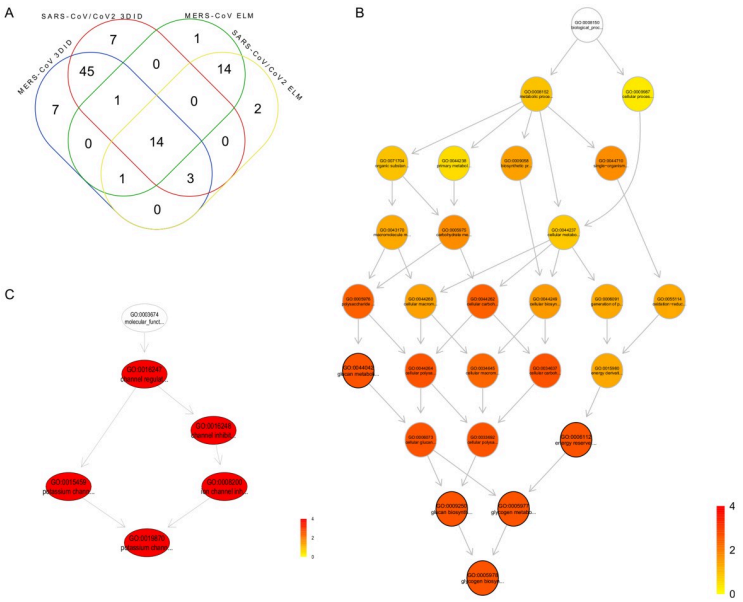

Fig 2. Protein domain enrichment analysis that produced the significant gene ontology terms for non-redundant motifs. (A) Venn diagram for the non-redundant domains. (B) Gene ontology terms for biological processes and (C) molecular functions terms of the non-redundant domains. Nodes are colored according to adjusted p-values.

https://doi.org/10.1371/journal.pone.0246901.g002

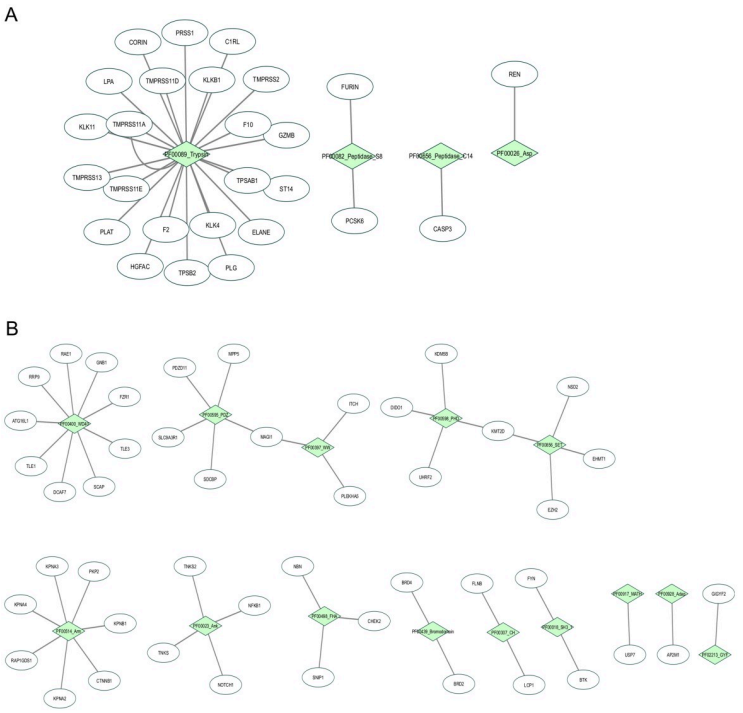

Fig 3. Network representation of significant domains linked to proteins and their gene ontology terms. (A) Biological processes and (B) Molecular functions. The green light diamonds represent the domains, and the ellipses represent the protein names associated with the domains. The images were generated with the cytoscape software [35].

https://doi.org/10.1371/journal.pone.0246901.g003

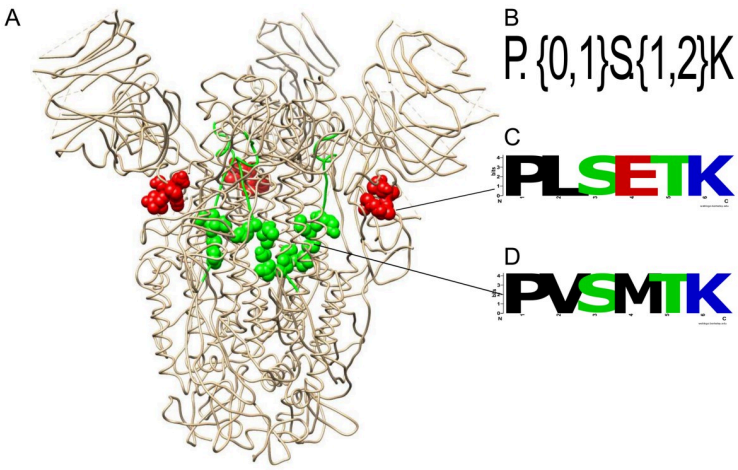

Fig 4. Some motifs matched the epitopes on the spike protein. (A) Spike protein of SARS-CoV-2 (PDBID:6XS6). (B) The regexp. (C) The motif PLSETK. Red balls indicate the PLSETK seqlogo motif mapped at amino acid positions 295 to 300 (D) The motif PVSMTK. Green balls and sticks showed the total length of the epitope ILPVSMTKTSVDCTMY ICGD, including the PVSMTK seqlogo motif mapped (the balls) at amino acids positions 728 to 733.

https://doi.org/10.1371/journal.pone.0246901.g004

Identification of amino acid motif sequences as immune epitopes

The non-redundant motifs $\geq$ 5 amino acids were searched for a match with epitopes reported on the IEDB, which were experimentally confirmed. The amino acid sequences of several motifs matched on epitopes sequences for SARS-CoV/CoV-2 that recognize B and T cells specific to class I or II MHC (data in S7 File). These motifs had the following main characteristics. 1) The epitope linear motifs contain the nested motifs recognized by both B and T cells. For example, the motif matched with the regexp [DE]..[IMV].[ST] was found on the B cell and T cell epitope PK**EITVA**TSRTLSYYK (IEDB ID: 48052) in the M protein [36] of SARS-CoV and SARS-CoV-2 [37]. 2) Motifs matched by the same regexp are prone to occur in different protein structural locations. For example, the regexp motif P.{0,1}S.{1,2}K matches the amino acid sequences PLSETK and PVSMTK locating to varying coordinates on the S protein (Fig 4). 3) Motifs maintain its crucial amino acids, and little variations occur at neighbor sites. For example, the **PVSMTK** motif nested on the B cell linear epitope IL**PVSMTKT**SVDCTMYICGD (IEDB ID:1309493) of SARS-CoV-2 (Fig 4A and 4D) [38] varied a little on the epitope sequence **PVSMAK**TSVDCNMYICGDS (IEDB ID: 49968) of the SARS-CoV, maintaining its main amino acid anchors P,S and K. PVSMAK was found only in one SARS-CoV-2 sequence (NCBI ID: QKV39263) isolated from Washington, Yakima County.

Discussion

In this work, we employed our previous data mining methodology [22] to identify potential functional motifs but applied to MERS-CoV and SARS-CoV/CoV-2 viruses. The main advantage of this method is the search restricted to human protein targets involved in the virus pathogenesis. The initial step allows us to reduce *a priori* the query on the 3DID and ELM databases. As a result, the unsheathed domain-motif information is potentially associated with human genes related to pathogenesis of the MERS-CoV and SARS-CoV/CoV2. Our approach is then similar to the methods used by Hagai, T., et al., Becerra, A. et al and Zhang, A et al [29, 39, 40] in predicting functional motifs. These methods include some distinctive features such as predicting disordered regions on the protein, the high frequency of amino acid motifs in the

protein sequences datasets under study, and the scarcity of amino acid motifs on shuffled sequences. The filters were tailored according to the information obtained in each data mining process. All those filtered steps guided our analysis to a more specificity that linked the predicted functional motifs as part of immune epitopes as previously we did for influenza A viruses [22]. It is distinctive of our prediction approach, because it was used to reduce the high rate of false positives associated with the computational prediction of motifs [41]. Furthermore, our method could be an alternative for computer-aided reverse vaccinology.

One interesting result is that the tendency of matched motifs occurred in the most variable proteins, the ORF1ab, and the S protein of the coronavirus proteomes. The ORF1ab contains the nonstructural proteins responsible for the translation machinery of viruses in the intracellular environment [42] and the S protein is essential for the virus's attachment to the host cell [43]. The tendency of motifs to appear on the proteins involved in virus replication was also observed in influenza viruses [44]. Thus, the high frequency of host-like motifs in those viral proteins suggests that such proteins could be the master kidnappers. Another finding is the high number of shared motifs across the proteome or distinct proteins of a proteome, reflecting the viral motifs to evolve independently in light of acquiring host-like mechanisms for the success in the invasion of host cells.

The domain enrichment analysis showed that the general biological processes, and molecular functions could be the consequence of the MERS-CoV and SARS-CoV/CoV-2 mimicry to hijack the host cell. The most significant ontology terms are the energy-saving and glycogen biosynthesis metabolism association. This result agrees with that viruses use the infected cells' carbon sources to achieve viral replication and virion production [45]. It is reasonable that glycogen, a storage form of glucose, is utilized in unexpected, exhausting cell activity [46] as infected. On the other hand, as this biosynthetic pathway is vital for the viruses' survival, targeting essential components such as the glycogen synthase kinase could help treat virus infections. It was reported that the use of two glycogen synthase inhibitors altered the hepatitis C virus assembly and release [47]. Hence, the proteins we found in the present study could be used to explore them as drug targets.

In another context, motifs have been suggested as potential immunogens [41]. It took our attention to search motif that matched with immune epitopes. Indeed we found that some motifs matched to the epitopes on the IEDB. Some of them were nested on the epitopes of earlier SARS-CoV and also present on those new SARS-CoV-2. It reaffirms the evidence of cross-reactive immune responses to coronavirus infections by SARS-CoV and SARS-CoV-2 [48–51]. Additionally, our study identified the epitopes harboring motifs that could interact with human protein domains. It is quite relevant because such domain-motifs shared in the different coronavirus can trigger a common molecular mimicry process that could lead to autoimmune diseases. It was demonstrated that antibodies derived from Flu vaccinated patients react with homologous sequences of the nucleoprotein of influenza A virus and the hypocretin receptor 2 domain of humans, the latter of which was involved in narcolepsy, an autoimmune adverse effect attributed to the Flu-vaccine [52]. Influenza immunization is also attributed to Guillain-Barré syndrome [53], a disease in which its pathogenesis is associated with several bacterial and viral pathogens' molecular mimicry [54–56]. Thus, our results are vital to helping in the currently underway rational vaccine development efforts, mainly because several autoimmune diseases have been associated with COVID-19 [57].

Conclusions

In conclusion, this study showed that our method's adaptability and practicality could guide a rational inference of domain targets and their interacting host-like motifs on the MERS-CoV

and SARS-CoV/CoV-2 proteomes. A high number of motifs were shared in the different CoVs, and it could interact with human proteins, indicating that molecular mimicry is a common strategy for CoVs. The finding of motifs as part of immune epitopes makes our method a suitable alternative for reverse vaccinology. The obtained information could be the starting point for future theoretic and experimental studies to develop new drugs and peptidic vaccines to combat those viruses.

Supporting information

S1 Fig. Order and disorder regions for the MERS-CoV, SARS-CoV, and SARS-CoV-2 proteins arranged by its known genome order.
(PDF)

S1 File. The literature mined information.
(XLSX)

S2 File. The merged gene name lists for the MERS-CoV and SARS-CoV/CoV-2.
(XLSX)

S3 File. The regexp lists obtained from 3DID and ELM.
(XLSX)

S4 File. The redundant regexp matched on the MERS-CoV and SARS-CoV/CoV-2 proteomes.
(XLSX)

S5 File. The non-redundant motifs with its domain accession partner.
(XLSX)

S6 File. GOBP and GOMF for the significant domains.
(XLSX)

S7 File. The motifs nested on linear sequences of epitopes from IEDB.
(XLSX)

Author Contributions

Conceptualization: Edgar E. Lara-Ramírez.

Data curation: Yamelie A. Martínez, Edgar E. Lara-Ramírez.

Formal analysis: Yamelie A. Martínez, Xianwu Guo, Edgar E. Lara-Ramírez.

Methodology: Edgar E. Lara-Ramírez.

Writing – original draft: Xianwu Guo, Edgar E. Lara-Ramírez.

Writing – review & editing: Diana P. Portales-Pérez, Gildardo Rivera, Julio E. Castañeda-Delgado, Carlos A. García-Pérez, José A. Enciso-Moreno.

References

1. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. New England Journal of Medicine. 2003; 348: 1967–1976. https://doi.org/10.1056/NEJMoa030747 PMID: 12690091
2. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. New England Journal of Medicine. 2012; 367: 1814–1820. https://doi.org/10.1056/NEJMoa1211721 PMID: 23075143

3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine. 2020 [cited 29 Oct 2020]. https://doi.org/10.1056/NEJMoa2001017 PMID: 31978945
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020; 5: 536–544. https://doi.org/10.1038/s41564-020-0695-z PMID: 32123347
5. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 2020 [cited 15 Jun 2020]. https://doi.org/10.1093/nsr/nwaa036
6. Apicella M, Campopiano MC, Mantuano M, Mazoni L, Coppelli A, Prato SD. COVID-19 in people with diabetes: understanding the reasons for worse outcomes. The Lancet Diabetes & Endocrinology. 2020; 8: 782–792. https://doi.org/10.1016/S2213-8587(20)30238-2 PMID: 32687793
7. Khorsand B, Savadi A, Naghibzadeh M. SARS-CoV-2-human protein-protein interaction network. Informatics in Medicine Unlocked. 2020; 20: 100413. https://doi.org/10.1016/j.imu.2020.100413 PMID: 32838020
8. He R, Leeson A, Ballantine M, Andonov A, Baker L, Dobie F, et al. Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. 2004; 105: 121–125. https://doi.org/10.1016/j.virusres.2004.05.002 PMID: 15351485
9. Vidalain P-O, Jacob Y, Hagemeijer MC, Jones LM, Neveu G, Roussarie J-P, et al. A Field-Proven Yeast Two-Hybrid Protocol Used to Identify Coronavirus–Host Protein–Protein Interactions. Coronaviruses. 2014; 1282: 213–229. https://doi.org/10.1007/978-1-4939-2438-7_18 PMID: 25720483
10. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020; 583: 459–468. https://doi.org/10.1038/s41586-020-2286-9 PMID: 32353859
11. Perrin-Cocon L, Diaz O, Jacquemin C, Barthel V, Ogire E, Ramière C, et al. The current landscape of coronavirus-host protein-protein interactions. J Transl Med. 2020; 18: 319. https://doi.org/10.1186/s12967-020-02480-z PMID: 32811513
12. Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nature Communications. 2020; 11: 3518. https://doi.org/10.1038/s41467-020-17189-2 PMID: 32665542
13. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery. 2020; 6: 1–18. https://doi.org/10.1038/s41421-020-0153-3 PMID: 32194980
14. Basu MK, Poliakov E, Rogozin IB. Domain mobility in proteins: functional and evolutionary implications. Brief Bioinform. 2009; 10: 205–216. https://doi.org/10.1093/bib/bbn057 PMID: 19151098
15. Lin MM, Zewail AH. Hydrophobic forces and the length limit of foldable protein domains. PNAS. 2012; 109: 9851–9856. https://doi.org/10.1073/pnas.1207382109 PMID: 22665780
16. Garamszegi S, Franzosa EA, Xia Y. Signatures of Pleiotropy, Economy and Convergent Evolution in a Domain-Resolved Map of Human–Virus Protein–Protein Interaction Networks. PLOS Pathog. 2013; 9: e1003778. https://doi.org/10.1371/journal.ppat.1003778 PMID: 24339775
17. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci. 2011; 36: 159–169. https://doi.org/10.1016/j.tibs.2010.10.002 PMID: 21146412
18. Mészáros B, Sámano-Sánchez H, Alvarado-Valverde J, Čalyševa J, Martínez-Pérez E, Alves R, et al. Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. Sci Signal. 2021; 14. https://doi.org/10.1126/scisignal.abd0334 PMID: 33436497
19. Kliche J, Kuss H, Ali M, Ivarsson Y. Cytoplasmic short linear motifs in ACE2 and integrin β3 link SARS-CoV-2 host cell receptors to mediators of endocytosis and autophagy. Sci Signal. 2021; 14. https://doi.org/10.1126/scisignal.abf1117 PMID: 33436498
20. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. Nucleic Acids Res. 2015; 43: D571–577. https://doi.org/10.1093/nar/gku1207 PMID: 25428358
21. Brister JR, Bao Y, Zhdanov SA, Ostapchuck Y, Chetvernin V, Kiryutin B, et al. Virus Variation Resource—recent updates and future directions. Nucl Acids Res. 2013; gkt1268. https://doi.org/10.1093/nar/gkt1268 PMID: 24304891
22. García-Pérez CA, Guo X, Navarro JG, Aguilar DAG, Lara-Ramírez EE. Proteome-wide analysis of human motif-domain interactions mapped on influenza a virus. BMC Bioinformatics. 2018; 19: 238. https://doi.org/10.1186/s12859-018-2237-8 PMID: 29940841
23. Wei C-H, Kao H-Y, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res. 2013; 41: W518–W522. https://doi.org/10.1093/nar/gkt441 PMID: 23703206
24. Consortium TU. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019; 47: D506–D515. https://doi.org/10.1093/nar/gky1049 PMID: 30395287

25. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019; 47: D427–D432. https://doi.org/10.1093/nar/gky995 PMID: 30357350

26. Mosca R, Céol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 2014; 42: D374–379. https://doi.org/10.1093/nar/gkt887 PMID: 24081580

27. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, et al. ELM—the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020; 48: D296–D306. https://doi.org/10.1093/nar/gkz1030 PMID: 31680160

28. Yan T, Yoo D, Berardini TZ, Mueller LA, Weems DC, Weng S, et al. PatMatch: a program for finding patterns in peptide and nucleotide sequences. Nucleic Acids Res. 2005; 33: W262–W266. https://doi.org/10.1093/nar/gki368 PMID: 15980466

29. Hagai T, Azia A, Babu MM, Andino R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. Cell Rep. 2014; 7: 1729–1739. https://doi.org/10.1016/j.celrep.2014.04.052 PMID: 24882001

30. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 2000; 16: 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2 PMID: 10827456

31. Fang H. dcGOR: An R Package for Analysing Ontologies and Protein Domain Annotations. PLoS Comput Biol. 2014; 10. https://doi.org/10.1371/journal.pcbi.1003929 PMID: 25356683

32. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, et al. The Immune Epitope Database 2.0. Nucleic Acids Res. 2010; 38: D854–D862. https://doi.org/10.1093/nar/gkp1004 PMID: 19906713

33. Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018; 46: W329–W337. https://doi.org/10.1093/nar/gky384 PMID: 29860432

34. Shen LW, Mao HJ, Wu YL, Tanaka Y, Zhang W. TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. Biochimie. 2017; 142: 1–10. https://doi.org/10.1016/j.biochi.2017.07.016 PMID: 28778717

35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003; 13: 2498–2504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

36. He Y, Zhou Y, Siddiqui P, Niu J, Jiang S. Identification of immunodominant epitopes on the membrane protein of the severe acute respiratory syndrome-associated coronavirus. J Clin Microbiol. 2005; 43: 3718–3726. https://doi.org/10.1128/JCM.43.8.3718-3726.2005 PMID: 16081901

37. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat Immunol. 2020; 21: 1336–1345. https://doi.org/10.1038/s41590-020-0782-6 PMID: 32887977

38. Yi Z, Ling Y, Zhang X, Chen J, Hu K, Wang Y, et al. Functional mapping of B-cell linear epitopes of SARS-CoV-2 in COVID-19 convalescent population. Emerg Microbes Infect. 2020; 9: 1988–1996. https://doi.org/10.1080/22221751.2020.1815591 PMID: 32844713

39. Becerra A, Bucheli VA, Moreno PA. Prediction of virus-host protein-protein interactions mediated by short linear motifs. BMC Bioinformatics. 2017; 18: 163. https://doi.org/10.1186/s12859-017-1570-7 PMID: 28279163

40. Zhang A, He L, Wang Y. Prediction of GCRV virus-host protein interactome based on structural motif-domain interactions. BMC Bioinformatics. 2017; 18: 145. https://doi.org/10.1186/s12859-017-1500-8 PMID: 28253857

41. Hraber P, O'Maille PE, Silberfarb A, Davis-Anderson K, Generous N, McMahon BH, et al. Resources to Discover and Use Short Linear Motifs in Viral Proteins. Trends in Biotechnology. 2020; 38: 113–127. https://doi.org/10.1016/j.tibtech.2019.07.004 PMID: 31427097

42. Yoshimoto FK. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. Protein J. 2020; 39: 198–216. https://doi.org/10.1007/s10930-020-09901-4 PMID: 32447571

43. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020; 181: 281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058 PMID: 32155444

44. Yang C-W. A Comparative Study of Short Linear Motif Compositions of the Influenza A Virus Ribonucleoproteins. PLoS One. 2012; 7. https://doi.org/10.1371/journal.pone.0038637 PMID: 22715401

45. Sanchez EL, Lagunoff M. Viral activation of cellular metabolism. Virology. 2015; 479–480: 609–618. https://doi.org/10.1016/j.virol.2015.02.038 PMID: 25812764

46. Berg JM, Tymoczko JL, Stryer L. Glycogen Metabolism. Biochemistry 5th edition. 2002 [cited 29 Oct 2020]. https://www.ncbi.nlm.nih.gov/books/NBK21190/

47. Sarhan MA, Abdel-Hakeem MS, Mason AL, Tyrrell DL, Houghton M. Glycogen synthase kinase 3β inhibitors prevent hepatitis C virus release/assembly through perturbation of lipid metabolism. Scientific Reports. 2017; 7: 2495. https://doi.org/10.1038/s41598-017-02648-6 PMID: 28566716

48. Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, et al. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. Science. 2020; 370: 89–94. https://doi.org/10.1126/science.abd3871 PMID: 32753554

49. Tai W, Zhang X, He Y, Jiang S, Du L. Identification of SARS-CoV RBD-targeting monoclonal antibodies with cross-reactive or neutralizing activity against SARS-CoV-2. Antiviral Res. 2020; 179: 104820. https://doi.org/10.1016/j.antiviral.2020.104820 PMID: 32405117

50. Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nature Reviews Immunology. 2020; 20: 457–458. https://doi.org/10.1038/s41577-020-0389-z PMID: 32636479

51. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell. 2020; 181: 1489–1501.e15. https://doi.org/10.1016/j.cell.2020.05.015 PMID: 32473127

52. Ahmed SS, Volkmuth W, Duca J, Corti L, Pallaoro M, Pezzicoli A, et al. Antibodies to influenza nucleoprotein cross-react with human hypocretin receptor 2. Science Translational Medicine. 2015; 7: 294ra105–294ra105. https://doi.org/10.1126/scitranslmed.aab2354 PMID: 26136476

53. Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, Keenlyside RA, Ziegler DW, Retailliau HF, et al. Guillain-Barre syndrome following vaccination in the National Influenza Immunization Program, United States, 1976–1977. Am J Epidemiol. 1979; 110: 105–123. https://doi.org/10.1093/oxfordjournals.aje.a112795 PMID: 463869

54. Rees JH, Soudain SE, Gregson NA, Hughes RAC. Campylobacter jejuni Infection and Guillain–Barré Syndrome. New England Journal of Medicine. 1995; 333: 1374–1379. https://doi.org/10.1056/NEJM199511233332102 PMID: 7477117

55. Steininger C, Popow-Kraupp T, Seiser A, Gueler N, Stanek G, Puchhammer E. Presence of Cytomegalovirus in Cerebrospinal Fluid of Patients with Guillain-Barré Syndrome. J Infect Dis. 2004; 189: 984–989. https://doi.org/10.1086/382192 PMID: 14999600

56. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. J Autoimmun. 2018; 95: 100–123. https://doi.org/10.1016/j.jaut.2018.10.012 PMID: 30509385

57. Ehrenfeld M, Tincani A, Andreoli L, Cattalini M, Greenbaum A, Kanduc D, et al. Covid-19 and autoimmunity. Autoimmun Rev. 2020; 19: 102597. https://doi.org/10.1016/j.autrev.2020.102597 PMID: 32535093