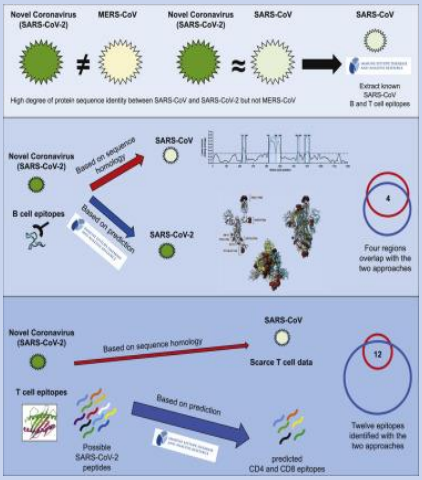
A SEQUENCE HOMOLOGY AND BIOINFORMATIC APPROACH CAN PREDICT CANDIDATE TARGETS FOR IMMUNE RESPONSES TO SARS-COV-2

Abstract :

The Immune Epitope Database and Analysis Resource (IEDB) is used to catalog available data related to other coronaviruses- SARS-CoV-, as there is limited information about (SARS-CoV-2) and the high sequence similarity between them

by identification of specific regions will facilitate effective vaccine design against this virus of high priority.



Keywords :

SARS-CoV; COVID-19; SARS-CoV-2, coronavirus; T cell epitope; B cell epitope; infectious disease; sequence conservation.

Introduction :

we used the IEDB and ViPR resources to compile known epitope sites from other coronaviruses, map corresponding regions in the SARS-CoV-2 sequences, and predict likely epitopes. We also used validated bioinformatic tools to predict B and T cell epitopes that are likely to be recognized in humans and to assess the conservation of these epitopes across different coronavirus species.

Limited information is currently available on which parts of the SARS-CoV-2 sequence are recognized by human immune responses but there is a significant body of information about epitopes for coronaviruses in general, and in particular for BETACORONAVIRUSES like SARS□CoV and MERS-CoV, which cause respiratory disease in humans.

1- Ten experimentally defined regions within SARS-CoV have high homology with SARS-CoV-2.

2- Parallel bioinformatics predicted potential B and T cell epitopes for SARS-CoV-2.

3- Independent approaches identified the same immunodominant regions.

4- The conserved immune regions have implications for vaccine design against multiple CoVs.

Related Work :

The Immune Epitope Database and Analysis Resource (IEDB) is a database of epitope-related material curated from clinical literature in the sense of infectious disease, allergy, and autoimmunity (Vita et al., 2019).

The Immune Epitope Database (IEDB, iedb.org) collects experimental evidence contained in scientific literature statistics, text, and charts, making it freely accessible and readily searchable to the public. The IEDB covers immune epitope evidence from all organisms examined and contains antibody, T cell, and MHC binding contexts consistent with bacterial, allergic, autoimmune, as well as transplant-related diseases After being publicly available for more than a decade, the IEDB's recent emphasis has been on enhanced query and reporting capabilities to satisfy our users' needs to view and summarise data that continues to increase in quantity and complexity. We have an update on our existing activities and strategic goals in this section.

In addition, the IEDB offers bioinformatic methods and algorithms for analysing epitope data and predicting possible epitopes from novel sequences. The Virus Pathogen Resource (ViPR) is a complementary repository of knowledge about human pathogenic viruses that combines genome, gene, and protein sequence information with information about immune epitopes, protein structures, and host

The Virus Pathogen Database and Analysis Resource (ViPR, www.ViPRbrc.org) is an interactive archive of data and analysis resources for various virus families funded by the NIAID Bioinformatics Resource Centers (BRC) network. ViPR provides details on human pathogenic viruses from the Arenaviridae, Bunyaviridae, Caliciviridae, Coronaviridae, Flaviviridae, Filoviridae, Hepeviridae, Herpesviridae, Paramyxoviridae, Picornaviridae, Poxviridae, Reoviridae, Rhabdoviridae, and Togaviridae virus families are currently supported, with hopes to add more virus families in the future. ViPR collects a variety of data, including sequence records. Annotations to genes and proteins, 3D protein structures, immune epitope locations, clinical and surveillance metadata, and novel data obtained from comparative genomics research are all available. There are also methods for metadata-driven statistical sequence analysis, multiple sequence alignment, phylogenetic tree building, BLAST comparison, and determining sequence variance. Workflows for data filtering and interpretation may be mixed, and the results stored in personal 'Workbenches' for future use. ViPR instruments and data are provided free of charge to the virology scientific community in order to aid in the creation of diagnostics, prophylactics, and therapeutics for priority pathogens and other viruses.

While no epitope data for SARS-CoV-2 are currently available, there is a substantial body of knowledge about coronavirus epitopes in general, and especially for Betacoronaviruses such as SARS-CoV and MERS-CoV, which cause respiratory disease in humans (de Wit et al., 2016, Song et al., 2019).

Coronaviruses (CoVs) were previously thought to be relatively harmless respiratory pathogens in humans. However, as a result of zoonotic CoVs breaching the species boundary, two cases of severe respiratory tract infection caused by the serious acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV) caused elevated pathogenicity

and morality rates in human population. This drew international attention to CoVs and emphasised the importance of monitoring infectious diseases at international boundaries. In this study, we concentrate on our current understanding of the epidemiology, pathogenesis, prevention, and treatment of SARS-CoV and MERS□CoV, as well as the critical structure and role of the spike proteins (S proteins) on the surface of each of the viruses. We compare existing pathogenesis-replicating animal models and summarise the possible function of host receptors in leading to complex host affinity in different organisms. We summarise the study that remains to be done to thoroughly understand the pathogenic mechanism of these viruses, to ducible animal models, and, finally, to establish countermeasures to defeat not only SARS and MERS-CoV,but also these emerging coronaviral diseases.

Method:

IEDB Analysis of Coronavirus T and B Epitopes

Coronavirus T and B cell epitopes were discovered by searching the IEDB at the end of January 2020. Positive assays in T cell, B cell, and/or ligand contexts were chosen from a large number of coronaviruses (taxonomy ID no.11118) queries. The total number of donors tested, as well as the corresponding total number of donors with positive responses in B or T cell assays, and as a function of host, were tabulated, as were the characteristics of each particular epitope (i.e., organisms, protein of provenance, positive assay type(s), MHC restriction). Finally, individual response frequency scores (RF) for T or B cell assays were determined for both hosts or for specific contexts (e.g., T cell assays in humans). RF = [(r – sqrt(r)]/t, where r is the total number of donors who responded and t is the total number of donors who were tested.

Comparison of Coronavirus Sequences to SARS-CoV-2

SARS-CoV and MERS-CoV full-length protein sequences were obtained from ViPR (https://www.viprbrc.org/brc/home.spg?decorator=corona) on 31 January 2020. Sequences from ‘‘unknown," rat, and monkey hosts were omitted from study to keep experimental strains out. In ViPR, the remaining sequences were matched using the MUSCLE algorithm. Sequences causing poor alignments in a preliminary analysis were removed before computing the final alignment. The consensus protein sequences of each virus group were determined from the final alignments using the Sequence Variation Analysis tool in ViPR. For epitope sequence analysis, protein sequences from natural virus isolates with sequences similar to the SARS-CoV and MERS-CoV consensus were chosen.

Determination of SARS-CoV-2 Sequence Conservation

Each Wuhan-Hu-1 (GeneBank: MN908947) protein sequence was compared to SARSCoV and MERS-CoV consensus protein sequences, as well as protein sequences from bat relatives (bat-SL-CoVZXC21) using the BLAST algorithm (ViPR;https://www.viprbrc.org/brc/blast.spg?method=ShowCleanInputPage&decorator=corona) to compute the pairwise identity between Wuhan-Hu-1 proteins and their comparison target.

QUANTIFICATION AND STATISTICAL ANALYSIS

The current theoretical research, which was focused on evidence from the existing literature and publicly accessible datasets, did not use any mathematical analyses. Percent identity and response factor scores were calculated as mentioned above in the Method Details section.

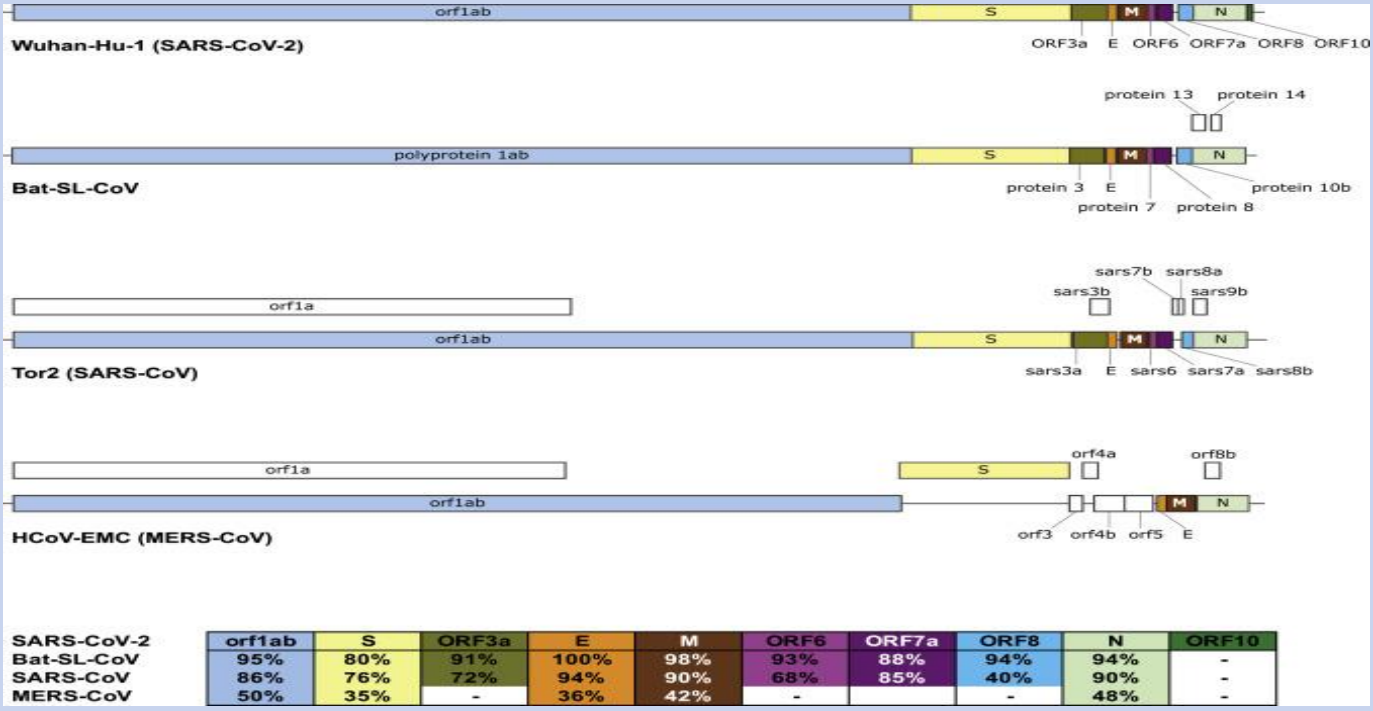
DATA AND CODE AVAILABILITY

As previously mentioned, all data provided and evaluated in this analysis was obtained from the IEDB and PDB. All data produced or analysed during this analysis is presented in the following tables, figures, and Supplementary Materials, which are included in the published paper. The accompanying author may have text files of data downloaded from the IEDB upon request.

Result:

A Wealth of Data Related to Coronaviruses Is Available in the IEDB

long belongs to the Coronaviradae family in the Nidovirales order, which is divided into four genera (Alpha-, Beta-, Gamma-, and Deltacoronaviruses). Several Alpha- and Betacoronaviruses infect humans, causing minor respiratory illnesses and common cold symptoms; some are zoonotic, infecting birds, goats, bats, and other species. In addition to SARS-CoV-2, two other coronaviruses, SARS-CoV and MERS-CoV, have caused major disease outbreaks with elevated (10–30%) lethality rates and widespread societal effects (Figure 1)



	orf1ab	S	ORF3a	E	M	ORF6	ORF7a	ORF8	N	ORF10
SARS-CoV-2										
Bat-SL-CoV	95%	80%	91%	100%	98%	93%	88%	94%	94%	-
SARS-CoV	86%	76%	72%	94%	90%	68%	85%	40%	90%	-
MERS-CoV	50%	35%	-	36%	42%	-		-	48%	-

Figure 1. Comparison of SARS-CoV-2 (Wuhan-Hu-1) Genome Structure with Its Closest Bat Relative (bat-SL-CoVZXC21), Tor2 SARS-CoV, and HCoV-EMC MERS-CoV.

Correspondence between the Epitopes Identified by the Two Different Approaches

The five SARS-CoV membrane protein and nucleoprotein regions that mapped to SARS-CoV-2 and those predicted by BebiPred 2.0 had no overlap. For those two proteins, no Discotope 2.0 prediction was available, as previously mentioned. Two of the possible epitope regions identified on the basis of SARS-CoV data are independently confirmed by the Discotope 2.0 prediction study based on the SARS-CoV-2 spike glycoprotein PDB structure. Finally, the 888–909 region is narrowly missed, because residue 914, which is predicted, is right outside of the epitope.

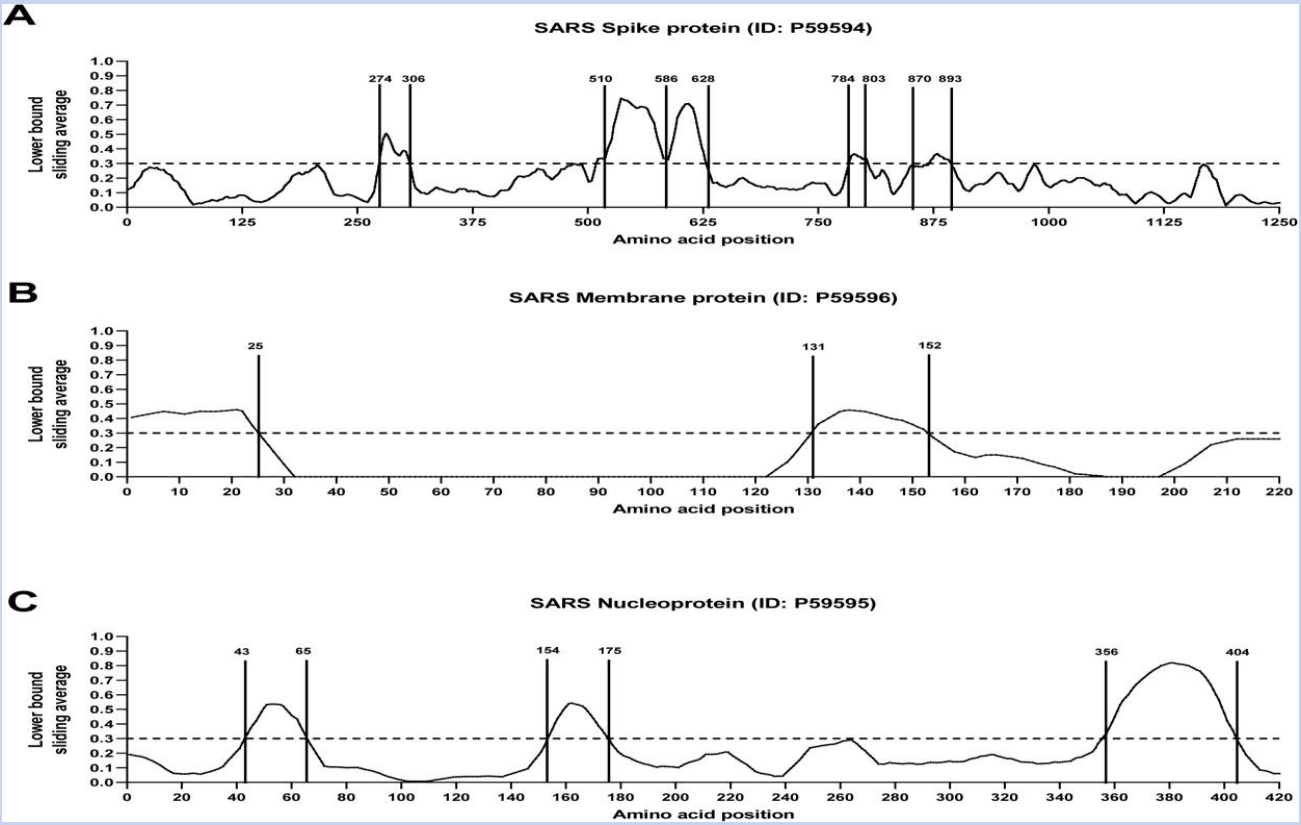


Figure 2. B Cell Immunodominant Regions Based on SARS-Specific Epitope Mapping

When we compared the SARS-CoV T cell epitopes that map ped to SARS-CoV-2 with the predicted CD4 and CD8 T cell epitopes, We discovered that the two methods independently classified 12 of 17 SARS-CoV-2 T cell epitopes with strong sequence identity (R90%) to the SARS-CoV. Another 7 out of 16 epitopes have a moderate sequence identity (70–89%), Both approaches have classified 6 of 12 epitopes with low sequence identity (70 percent). Provided that the experimental results are obtained from a distorted series of HLA restrictions (primarily HLA A*02:01), the absence of absolute correspondence is not unexpected , and that our HLA class I prediction approach focused on a smaller range of alleles chosen to represent the most common worldwide variants; however, class II predictions are predicted to cover 50% of class II responses.