Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

ELSEVIER

Web resources facilitate drug discovery in treatment of COVID-19

Long-Can Mei [a,b], Yin Jin [c], Zheng Wang [c], Ge-Fei Hao [a,b,c,*], Guang-Fu Yang [a,b,d]

[a] Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, China
[b] International Joint Research Center for Intelligent Biosensor Technology and Health, Central China Normal University, Wuhan 430079, China
[c] State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Research and Development Center for Fine Chemicals, Guizhou University, Guiyang 550000, China
[d] Collaborative Innovation Center of Chemical Science and Engineering, Tianjin 300072, China

The infectious disease Coronavirus 2019 (COVID-19) continues to cause a global pandemic and, thus, the need for effective therapeutics remains urgent. Global research targeting COVID-19 treatments has produced numerous therapy-related data and established data repositories. However, these data are disseminated throughout the literature and web resources, which could lead to a reduction in the levels of their use. In this review, we introduce resource repositories for the development of COVID-19 therapeutics, from the genome and proteome to antiviral drugs, vaccines, and monoclonal antibodies. We briefly describe the data and usage, and how they advance research for therapies. Finally, we discuss the opportunities and challenges to preventing the pandemic from developing further.

Keywords: Bioinformatics; SARS-CoV-2; Sequence and structure; Drug design; Vaccines; Monoclonal antibodies

Introduction

Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread across the world since it was first reported in late 2019 [1]. This infectious disease has caused millions of deaths, and the death rate continues to grow. The pandemic has also had a significant effect on global health, economies, societies, and education. Following on from SARS-CoV, first reported in 2002, and Middle East respiratory syndrome coronavirus (MERS-CoV) reported in 2012, SARS-CoV-2 is the third coronavirus (CoV) epidemic of the 21st century. However, whereas these CoVs can lead to severe and potential fatal respiratory tract infections, other known CoVs cause mild respiratory symptoms similar to a common cold [2].

Unprecedented efforts have been devoted to developing therapeutic interventions to mitigate the current pandemic. Based on research into SARS-CoV and MERS-CoV, researchers and clinicians could learn lessons to aid the development of therapies against SARS-CoV-2, from repurposing drugs to searching for novel therapies [3]. Promising therapies mainly include antiviral drugs, vaccines and monoclonal antibodies (mAbs).

Antiviral drugs are used for treating viral infections by preventing the virus from entering host cells, replicating, or releasing viral particles to infect other cells. Remdesivir was the first antiviral approved by the US Food and Drug Administration (FDA) for emergency use as a treatment for COVID-19. However, not all clinical trials proved remdesivir to be effective. A vaccine trains the immune system to recognize and attack a virus to protect the host before it is exposed to SARS-CoV-2. At the time of writing (January 2021), 13 vaccines had been approved or authorized for limited use, including mRNA-based vaccines, adenovirus vaccines, recombinant adenovirus vaccines, nonreplicating viral vectors, inactivated vaccines, and peptide vaccines. Moreover, 78 vaccines were being testing in clinical trials.

Antibodies protect the host from viral attack by triggering the immune system. The SARS-CoV-2 spike glycoprotein (S protein)

* *Corresponding author:* Hao, G.-F. (gefei_hao@foxmail.com)

https://doi.org/10.1016/j.drudis.2021.04.018

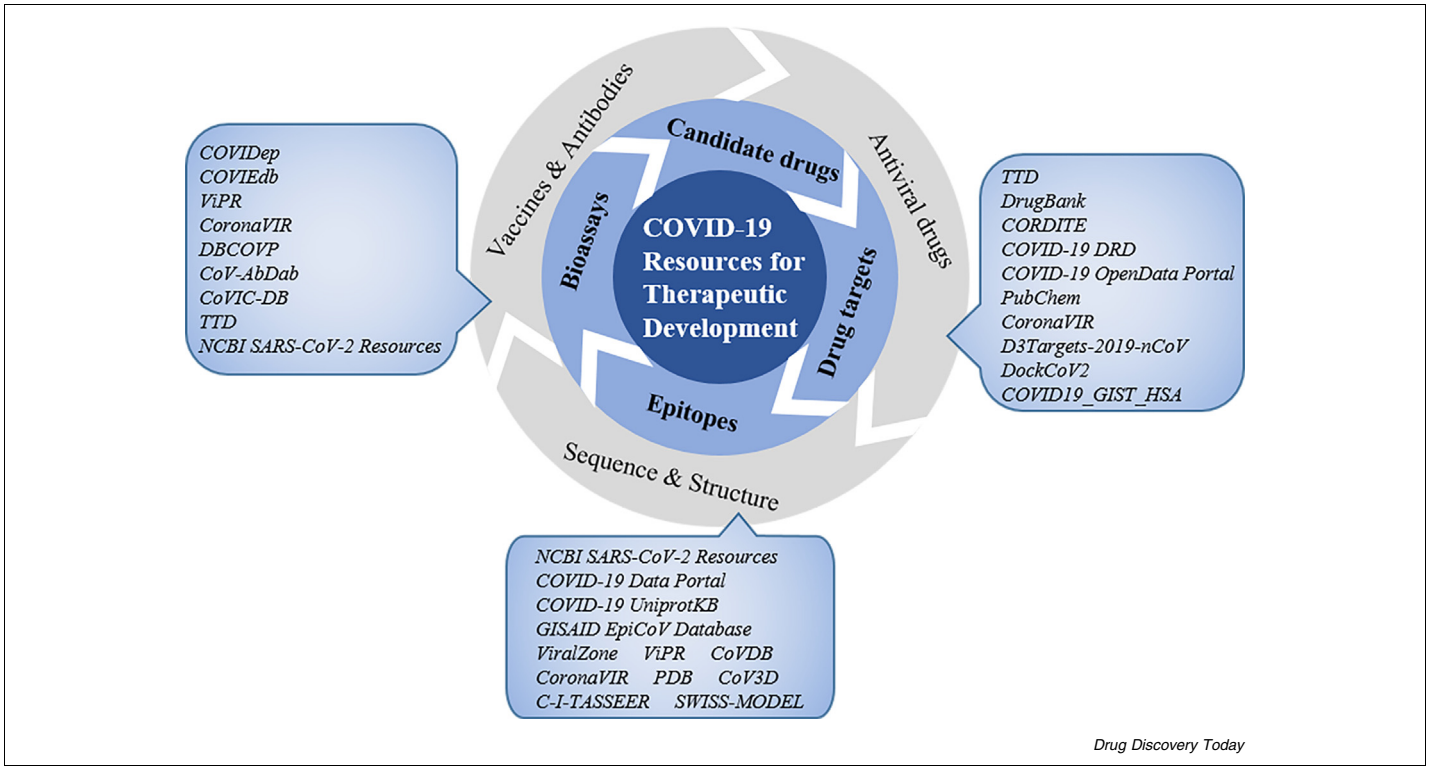

FIGURE 1

Multidisciplinary resources for the development of therapeutic strategies for Coronavirus 2019 (COVID-19). Existing resource repositories contain research data including genome and proteome information, antiviral drugs and drug targets, antigen epitopes, bioassays, and clinical trials. These resources are classified by data type and usage: genome and proteome data; relevant data for the development of anti-COVID-19 drugs, as well as vaccines and monoclonal antibodies (mAbs).

is one of the most significant targets for antibody programs. At the time of writing, two antibody therapeutics had been approved (in Russia and India), and 79 antibodies were in clinical trials. Despite such unprecedented efforts across the globe, there remains much to explore that could be helpful in developing COVID-19 therapies.

Given the focus on CoVs, numerous data have been collected. Many scientific institutes and research groups have established COVID-19-related resource hubs for data sharing. These resources cover the virus genome and proteome sequence, COVID-19-related targets, potential drug leads and bioactivity assays, and epitopes. For example, National Center for Biotechnology Information (NCBI) SARS-CoV-2 Resources contains comprehensive SARS-CoV-2 sequence data for diagnostics and epitope mapping; DrugBank houses data related to anticoronavirus drugs, therapeutic targets, and corresponding pathways. These can help the scientific community to keep up with the rapidly growing number of biological and chemical resources associated with the virus, and make the latest information available to a wide audience. Recent studies have summarized the use of various tools for research into COVID-19 evolution, pathogenesis, and treatment [4–6]. However, these online resources remain underutilized, even though they could serve as the starting point for relevant studies.

Sharing available data with researchers and clinicians as quickly and widely as possible could help to accelerate therapeutic research, given that rapid data sharing could be the basis for understanding the disease and developing treatments. In this review, we provide a comprehensive up-to-date survey of research resources for SARS-CoV-2 and COVID-19 (Fig. 1). According to data associations, these resources can be categorized into three groups: genome and proteome; antiviral drugs; and immunotherapy. We compare and analyze their scales of data and scopes of usage, and discuss the treatment experience learned from previous studies and the possible risks from drug repurposing and viral genome variability. We anticipate that our review will aid the use of multidisciplinary resources and cooperation between the global research groups, which will further facilitate the development of therapeutic approaches.

SARS-CoV-2 protein data for therapeutic target identification

Research on SARS-CoV-2 protein data, including sequence and structure, could provide insights into the development of structure-guided therapeutic approaches. Investigations based on SARS-CoV-2 sequence data enable the discovery of new therapeutic targets for molecular assays and mutation analyses for predicting effects on existing assays [7,8] .The availability of sequence data also provides fundamental information for epitope mapping and modeling [9,10]. Since the first release of the SARS-CoV-2 genome sequence, next-generation sequencing has provided large numbers of high-quality genome sequences. Additionally, structural biology techniques have been devoted to resolving the SARS-CoV-2 protein structures, which provide a basis for structure-based drug design campaigns [11]. For

example, resolving the structure of the spike protein could help to predict its interactions with human receptor proteins, and facilitate the discovery of promising small molecules to prevent interactions that would then block viral infection and spread [12,13]. SARS-CoV-2 sequence and structure data have been stored and integrated in different public databases, freely available to all users (Table 1).

Resources for gene and protein sequence information

Several databases have been established to store, retrieve, and manage SARS-CoV-2 sequence information. The most widely used databases are from NCBI and the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBL). NCBI SARS-CoV-2 Resources offers compilations of nucleotide and protein sequences collected from numerous scientific publications and different research institutes. At the time of writing, it contained 50 975 SARS-CoV-2 nucleotide sequences, including a RefSeq genome record NC_045512, and 555 716 SARS-CoV-2 protein sequences. It coordinates with individual laboratories and allows individuals to upload their SARS-CoV-2 sequence data. Similarly, the COVID-19 Data Portal launched by EMBL-EBL provides comprehensive sequence information, and exchanges data with NCBI SARS-CoV-2 Resources constantly. At the time of writing, it contained 46 730 CoVs nucleotide sequences, six human host nucleotide sequences, as well as 16 CoV and 47 human receptor protein sequences. The unique data in COVID-19 Data Portal included 87 sets of gene expression data for understanding the biology of the virus. To provide comprehensive protein sequence and function annotation data, COVID-19 UniProtKB focuses on SARS-CoV-2 and human receptor protein information [14]. At the time of writing, it had 79 protein entries, including 47 human proteins, 15 SARS-CoV proteins, and 17 SARS-CoV-2 proteins. Each protein entry is annotated in detail, including function, names and taxonomy, pathology, expression, sequence, and structure.

Other databases related to virology and epidemiology store SARS-CoV-2 sequence data, such as GISAID EpiCoV Database [15], ViPR [16], and ViralZone [17]. The GISAID EpiCoV Database

TABLE 1

SARS-CoV-2 sequence and structure data repositories.[a]

Resource/URL	Description	No. of gene sequences		No. of protein sequences		No. of protein structures	
		CoV	Human[b]	CoV	Human[b]	CoV	Human[b]
NCBI SARS-CoV-2 Resources (www.ncbi.nlm.nih.gov/sars-cov-2/)	SARS-CoV-2 genome sequence and annotation, as well as protein sequence and structures from NCBI resource databases	50 975	132	555 716	–	27	–
COVID-19 Data Portal (www.covid19dataportal.org/)	CoV and host genome sequences; SARS-CoV-2 and associated receptor protein sequences and structures	46 730	6	16	47	9	43
COVID-19 UniprotKB (https://covid-19.uniprot.org/uniprotkb?query=*)	SARS-CoV-2 and human receptor protein sequences and functional information	–	–	17	47	9	43
GISAID EpiCoV Database (https://db.cngb.org/gisaid/)	Collection of SARS-CoV-2 genome sequences and related clinical and epidemiological data	396 118	–	–	–	–	–
ViralZone (https://viralzone.expasy.org/9056)	COVID-19-relevant genome expression data, protein sequences, and SARS-CoV-2 interactome	–	–	17	–	–	–
ViPR (www.viprbrc.org/brc/home.spg?decorator=corona_ncov)	SARS-CoV-2-relevant genome, proteome, phylogeny, epidemiology, and clinical data	46 263	–	1 058 542	–	303	–
CoVdb (http://covdb.popgenetics.net/v3/)	CoV genome and proteome database for evolutionary analysis	5709	–	50 000	–	3 000 000[c]	–
CoronaVIR (https://webs.iiitd.edu.in/raghava/coronavir/)	SARS-CoV-2 genomic and proteomic data from public databases and literature	53	–	10	–	12	–
Protein Data Bank (www.rcsb.org/)	Experimentally determined structures of biological macromolecules	–	–	–	–	921	135
CoV3D (https://cov3d.ibbr.umd.edu/cov3d)	Repository of CoV protein structures	–	–	70	–	770	–
Coronavirus3D (https://coronavirus3d.org)	3D visualization and analysis of SARS-CoV-2 protein structures with respect to CoV-2 mutational patterns	482 468	–	30	–	36	–
C-I-TASSER (https://zhanglab.ccmb.med.umich.edu/COVID-19/)	Genome-wide structures of SARS-CoV-2 modeled by C-I-TASSER	–	–	24	–	24	–
SWISS-MODEL (https://swissmodel.expasy.org/repository/species/2697049)	Full SARS-CoV-2 proteome modeled by SWISS-MODEL	–	–	–	–	39	–

[a] Data up to January 2021.
[b] Human sequence and structure data related to SARS-CoV-2.
[c] CoV gene sequences mappings with an E-value < 0.05 and a coverage >50%.

contains 396 118 genome sequences, and related clinical and epidemiological data. Similarly, ViPR contains genome and proteome sequences for phylogeny and epidemiology research. Unlike the GISAID EpiCoV Database and ViPR, ViralZone provides information about SARS-CoV-2 genome expression, protein sequences, and the interactome with human proteins. Nucleotide and protein sequences in ViralZone are retrieved from NCBI and UniProt, respectively. Newly emerging databases also provide sequence data. For example, CoronaVIR is an integrated multi-omics repository for diagnosis and treatment [18], which maintains 53 genome sequences of SARS-CoV-2 strains from NCBI. CoVdb contains comprehensive CoV genomic information across more than 5000 CoV strains [19,20]. These resources contribute to the research on viral sequences.

Comparative analysis of these sequence resources reveals the divergence in data volume and type (Table 1). NCBI SARS-CoV-2 Resources and the COVID-19 Data Portal contain the most comprehensive sequence data relating to SARS-CoV-2, with both containing over 46 000 sequence records. SARS-CoV-2 related human sequence data are absent in most resource repositories, althoughNCBI SARS-CoV-2 Resources provides 132 relevant human gene sequence records. For protein sequence information, COVID-19 UniprotKB provides 64 viral and human protein sequences and annotations. Some of these databases provide tools to analyze specialized content. However, CoronaVIR is not updated continuously, and its data volume is relatively small. The sudden accumulation of SARS-CoV-2 genome sequences comes with challenges for big data analytics. Convenient and versatile bioinformatics tools are needed to solve this problem, which could help control not only the current COVID-19 pandemic, but also future outbreaks of infectious disease. We recommend that researchers use NCBI SARS-CoV-2 Resources, COVID-19 Data Portal, and UniProt for studies of genome variability and vaccine targets mapping, and GISAID EpiCoV Database and ViPR for epidemiology research.

Resources for protein structure information

Several databases provide structural data obtained using experimental and computational techniques (Table 1). Protein Data Bank (PDB) contains experimentally determined structures for biomacromolecules [21], including all the resolved structures of SARS-CoV-2 proteins and human receptor proteins. At the time of writing, it contained 921 SARS-CoV-2 protein structures and 135 relevant human receptor structures. To facilitate the structure-based design of antibodies and vaccines, CoV3D was developed to focus on betacoronavirus protein structures combined with drug targeting and antibody recognition information [22]. Besides experimentally determined structures, computational modeled structures can also support a better understanding of possible drug targets [23]. Zhang's research group modeled structures of all SARS-CoV-2 proteins by using the C-I-TASSER algorithm [24,25], which is an integration of deep convolutional neural-networks and the I-TASSER algorithm. Unlike the sequence-based method of C-I-TASSER, SWISS-MODEL is a homology modeling-based method to predict protein structure, developed by Schewede's group at the Swiss Institute of Bioinformatics (SIB) [26]. They modeled the full SARS-CoV-2 proteome based on the NCBI reference sequence NC_045512 [27]. For the systematic analysis of SARS-CoV-2 genome mutations and their influences, the Coronavirus3D online platform was developed for exploring the distribution of the mutations in the context of the corresponding protein structures [28]. It contains 27 092 mutations based on 482 468 genomes from SARS-CoV-2 protein structures. It allows users to analyze the mutations visually in their 3D context.

A summary of these structure resources is provided in Table 1. The PDB database provides the most comprehensive structural data, including 921 CoV proteins and 135 human receptor proteins, whereas CoV3D focuses on the spike glycoprotein, which could be used as an antibody target. Coronavirus3D integrates structural information with genome mutation data, making it possible to explore SARS-CoV-2 genome variability in the context of protein structure. Computational approaches provide alternatives to model unknown proteins, which could accelerate research of the viral proteome. C-I-TASSER is capable of structural modeling using only the amino acid sequence, and SWISS-MODEL is dependent on a similar structure template. In addition to individual viral protein models, it is necessary to model the complexes of virus proteins and human receptors. For example, SWISS-MODEL predicts five complex structures between the viral and host proteins. However, the quality of these modeled structures requires improvement in some cases. The reliability of the structure modeled by SWISS-MODEL and C-I-TASSER has not yet been validated by experiments. Thus, researchers should verify with caution the conclusions derived from these structural data. The SWISS-MODEL might not perform as well as the C-I-TASSER model for proteins lacking homologous templates. Thus, given the data quality, the PDB database is recommended for structural analysis because of its validated structures, whereas Coronovirus3D is recommended for variability exploration.

Resources available for anti-COVID-19 drug discovery

The efficiency of drug discovery depends crucially on the identification of the therapeutic target and therapeutic agents. Comprehensive information about therapeutic targets and early drug candidates is useful to promote further drug discovery. In addition, repurposing of approved drugs could significantly accelerate the development of therapies for COVID-19 [29,30]. During the early days of the outbreak, drug repurposing was used a preference to screen the safety and efficacy of existing drugs rapidly for the treatment of COVID-19. The development of targeted therapeutics can be promoted by data of drug and target structures, *in vitro* and *in vivo* assays, and clinical trials. Several public repositories have been established for sharing relevant pharmaceutical data (Table 2).

Data relating to candidate drugs, drug targets, and drug–target interactions have been collected for public access. The Therapeutic Target Database (TTD) and DrugBank are two notable databases [31,32]. The main purpose of TTD is to provide therapeutic targets for corresponding drugs, and target and drug cross-linked pathways. At the time of writing, it contained 210 anticoronavirus drugs and 45 corresponding therapeutic targets from public reports. Compared with TTD, DrugBank focuses on drug interactions, pharmacology, and clinical trial information. At the time of writing, it contained 81 potential drug targets

TABLE 2

Resource repositories for anti-COVID-19 drug discovery and development.[a]

Resource/URL	Description	No. of drugs/compounds	No. of targets/proteins	DTI[b]	Bioassay data[c]	Clinical data[d]
Therapeutic Target Database (http://db.idrblab.net/ttd/ttd-search/covid-19-profile)	Collection of anticoronavirus drugs and associated targets and pathways	210	45	97	–	–
DrugBank (https://go.drugbank.com/covid-19)	COVID-19-related drugs, drug targets, clinical trials, and publications	43	81	126	–	3307
CORDITE (https://cordite.mathematik.uni-marburg.de/#/)	Drug interactions with SARS-CoV-2 and human receptor proteins from literature and clinical trials	823	29	1172	–	247
COVID-19 Drug Repurposing Database (www.excelra.com/covid-19-drug-repurposing-database/)	Approved small molecules or promising drug candidates with bioactivity data from publications, reports, and databases	106	15	92	–	–
COVID-19 OpenData Portal (https://opendata.ncats.nih.gov/covid19/)	Repository for sharing SARS-CoV-2 screening data and assay protocols	10 000	–	–	9958[e]	–
PubChem (https://pubchem.ncbi.nlm.nih.gov/#query=covid-19)	Information on antiviral compounds, relevant targets, pathways, and bioassays	1201	228	2197	275[f]	–
Chemical Checker (https://sbnb.irbbarcelona.org/covid19/)	Bioactive chemical compounds with chemical and bioactivity features similar to drug candidates	10 307	24	–	–	–
CoronaVIR (https://webs.iiitd.edu.in/raghava/coronavir/)	COVID-19-related potential drugs and drug targets manually curated from literature	204	9	–	–	–
DockCoV2 (https://covirus.cc/drugs/)	Molecular docking data of approved drugs against SARS-CoV-2 and human receptor proteins	3109	7	21 539	–	–

[a] Data up to January 2021.
[b] Number of drug–target interaction data.
[c] Number of bioassay data relating to drugs/compounds.
[d] Number of clinical trial records relating to drugs/compounds.
[e] Bioassay data types mainly include proximity, biophysical, biochemical, cell-based, and cell viability data.
[f] Bioassay data types mainly include biochemical, Cell-Titer Glo viability assay, and cell-based data.

and 43 drug candidates. Three newly established databases, CoronaVIR [18], CORDITE [33], and COVID-19 Drug Repurposing Database (DRD), provide drug–target interactions (DTIs) relevant to COVID-19. CoronaVIR collates potential drug and targets from the literature and, at the time of writing, it contained nine potential drug targets and 204 candidate drugs. CORDITE provides interactions of drugs with viral and human proteins, hosting 839 drugs, 29 targets, 1172 sets of drug–target interactions, and 247 clinical trial records. The data in CORDITE are manually curated from the literature. By contrast, COVID-19 DRD focuses on drug repurposing research, containing at the time of writing 106 small-molecule drugs against 15 drug targets, the modes and mechanisms of action, and the relevant literature.

Moreover, information about antiviral agents has been integrated in public databases. COVID-19 OpenData Portal [34] and PubChem [35] have collated antiviral agents and detailed descriptions of their properties and bioassay data. COVID-19 OpenData Portal, managed by the National Center for Advancing Translational Sciences (NCATS), contains screening assay data for more than 10 000 compounds, including nearly 3000 approved drugs. Compared with COVID-19 OpenData Portal, PubChem provides not only the bioassays of the antiviral chemicals, but also medication information, and information of the pharmacology, and biological pathways. To expand the list of drugs to fight SARS-CoV-2, a list of bioactive chemical compounds with potential to be effective against COVID-19 was filtered from 800 000 bioactive compounds according to similar chemical and bioactivity features. This includes 307 drug molecule candidates from the COVID-19 literature and 10 000 bioactive compounds from the larger chemical space of Chemical Checker [36]. Additionally, computational techniques have contributed to the discovery of antiviral agents [37]. DockCoV2 is a database housing the molecular docking results of 3109 FDA-approved and Taiwan National Health Insurance (NHI) drugs with seven SARS-CoV-2 and human receptor proteins [38]. It aims to predict binding affinity between drugs and protein targets.

A statistical analysis reveals the divergence among these resources (Table 2). For therapeutic targets, DrugBank and TTD provide more comprehensive drug targets in the clinical stage than other databases, including 81 and 45 drug targets, respectively. TTD focuses only on the identified therapeutic targets corresponding to a drug, overlooking relevant protein interactions. For anticoronavirus drugs in the clinical stage, TTD contains the most drug data (up to 210 drug records). Chemical checker provides the largest number of bioactive molecules(up to 10 000), which need to be further validated by biological assays. For DTIs, CORDITE contains the largest number of experimentally identified interactions (~1172 DTIs). Few resources provide bioassay data relating to the drugs and compounds, except COVID-19 OpenData Portal and PubChem. COVID-19 OpenData Portal contains the largest number of experimental tested antiviral compounds(more than 10 000), and the unique data in COVID-19 OpenData Portal are the assay protocols. DrugBank provides the widest coverage of data relating to drugs in the clinical stage.

TTD, DrugBank, and PubChem are recommended for their large-scale coverage of clinical drug and target data, whereas COVID-19 OpenData Portal and PubChem are recommended for bioassay data. For pharmaceutical chemists in the field of computer-aid drug design, DockCoV2 is recommended for efficient discovery of lead compounds. The DTI data in DockCoV2 are from the computational prediction of molecular docking. It explores the possible drug–target binding modes, which provide clues for the validation process. Molecular docking has made it possible to highlight small molecules that would be worth testing as COVID-19 therapy, most of which are supported by their physicochemical and structural properties; therefore their probability of success as antiviral treatments is relatively high. However, most of these results have not yet been evaluated in biological assays. Assessing the quality of predictions is challenging, because many studies do not perform experimental evaluation. A general recommendation for evaluating the quality of predictions is to determine the similarity between the predicted compounds and the drugs undergoing clinical trials. There are numerous candidate drug compounds that might be useful for treating COVID-19, some of which are undergoing clinical trials. Structure–activity analysis helps us to understand structural requirements for higher activity. For inhibitors targeting angiotensin-converting enzyme-2 (ACE-2), the hydrophilic functional groups are common in the structures of the bioactive compounds. For inhibitors targeting the S-protein, the small-molecule structures commonly comprise an amide bond, aromatic rings, and a piperidine ring.

Available resources for developing vaccines and monoclonal antibodies

Immunotherapy is an effective method to cure CoV infection. Many immunotherapy trials are underway against COVID-19, as previously developed for SARS-CoV and MERS-CoV [39,40]. However, determination of epitopes is crucial in epitope-based vaccine design, and could stimulate an enhanced immune response against the viral infection. This process is slow because of our limited understanding of SARS-CoV-2 and its infection mechanism. An essential problem for antibody tests is the identification of antigens or proteins from the viral coat that trigger the immune response, leading to the production of antibodies.

To provide potential vaccine epitopes for vaccine design, several resource repositories cover immune epitopes of CoVs, such as COVIDep [41], COVIEdb, ViPR [16], DBCOVP [42], and CoronaVIR [18] (Table 3). COVIDep was the first platform developed for epitope recommendations for SARS-CoV-2 [41]. It contains 58 identified B cell and 285 identified T cell epitopes that have been determined experimentally for SARS-CoV and have high genetic similarities with SARS-CoV-2. Thus, those epitopes could contribute to SARS-CoV-2 vaccine design. Numerous efforts have been devoted to the development of computational methods for the prediction of epitopes. Some databases target predicted epitopes based on different computational philosophies. For example, CoronaVIR provides B cell epitopes predicted computationally by LBtope [43], T cell epitopes by CTLpred [44], major histocompatibility complex binders by ProPred 1 and ProPred [45]. Compared with CoronaVIR, COVIEdb contains not only the predicted cell epitopes for four CoVs, but also the experimentally validated epitopes from the literature, in which the B cell epitopes were predicted by seven computational tools and the T cell epitopes were predicted by another seven tools [46]. Different from COVIEdb, ViPR contains the epitopes predicted by NetCTL server [47], and the experimentally determined epitopes from the Immune Epitope Database (IEDB) [48]. DBCOVP focuses on virulent glycoproteins from Betacoronavirus genera. It houses the predicted B/T cell epitopes combined with detailed annotation including the conservancy analysis, allergenicity, antigenicity, toxicity, and structural properties. These resources provide abundant immune epitope data for epitope-based vaccine design.

Moreover, databases related to mAbs have been established to integrate and characterize anticoronavirus antibodies for

TABLE 3

Resources for the development of vaccines and mAbs for COVID-19.[a]

Resource/URL	Description	T cell epitopes	B cell epitopes	Vaccines	mAbs[b]	Clinical data[c]
COVIDep (https://covidep.ust.hk/)	Potential vaccine targets for SARS-CoV-2	285	58	–	–	–
COVIEdb (http://biopharm.zju.edu.cn/coviedb/)	Predicted vaccine epitopes for CoVs	14 489	11	–	–	–
ViPR (www.viprbrc.org/brc/home.spg?decorator=corona_ncov)	Predicted and experimentally determined epitopes	1586	476	–	–	–
CoronaVIR (https://webs.iiitd.edu.in/raghava/coronavir/)	Collection of experimentally validated epitopes from IEDB	25	59	17	–	–
DBCOVP (http://covp.immt.res.in/Default.aspx)	Functional and immunological properties of CoV virulent glycoproteins	21	4	–	–	–
CoV-AbDab (http://opig.stats.ox.ac.uk/webapps/covabdab/)	Sequences and structures of anticoronavirus antibodies and nanobodies	–	–	–	1402	–
CoVIC-DB (https://covic.lji.org/databases/)	Potency, efficacy, and structural analyses of antibodies against SARS-CoV-2	–	–	–	52	–
Therapeutic Target Database (http://db.idrblab.net/ttd/ttd-search/covid-19-profile)	Information about vaccines and mAbs for CoV	–	–	29	22	–
NCBI SARS-CoV-2 Resources (www.ncbi.nlm.nih.gov/sars-cov-2/)	Clinical studies related to vaccines and antibodies	–	–	201	58	259

[a] Data up to January 2021.
[b] Number of mAb resources for CoVs.
[c] Number of clinical trial records for vaccines and mAbs for CoVs.

yielding new insights for the development of antibody-based therapeutics. CoV-AbDab was the first established antibody database for betacoronaviruses, and contains manually curated entries of 1402 antibodies from the literature and patents [49]. Of those entries, 1131 are for SARS-CoV-2, 483 for SARS-CoV-2, and 147 for MERS-CoV. To characterize the biological activity, another mAb database, CoVIC-DB Database, provides the tested antibodies combined with the binding and inhibitor assay data. Moreover, some databases provide information on the proved or clinical vaccines and antibodies. For example, the TTD database contains records for 29 anticoronavirus vaccines and 22 mAbs in clinical and preclinical trials. NCBI SARS-CoV-2 Resources provides access to the relevant clinical studies from ClinicalTrials.gov database, including 201 vaccines and 58 mAbs. These resources could be used collectively to discover potential vaccine epitopes and to facilitate the design of mAbs.

We compared the data type and volume of the above resources (Table 3). COVIEdb contains the largest number of predicted vaccine epitopes (~14 500 predicted epitopes), whereas COVIDep has collated the most epitope data demonstrated by bioactivity trails (~343 identified epitopes). These databases provide computationally predicted epitopes not validated by experiments. Thus, the epitope data require further rigorous tests to prove their effectiveness. For the studies of antibodies against betacoronaviruses, CoV-AbDab is the most comprehensive database, housing 1402 experimentally determined antibody sequence and structure data. Few resources have collated clinical trials of vaccines and mAbs. As mentioned earlier, NCBI SARS-CoV-2 Resources provides external access to clinical trials recorded in ClinicalTrials.gov database. Together, these resources are not only valuable for research on vaccine- and mAb-based therapeutics, but also helpful for the development of computational tools to predict vaccine and antibody targets. According to their maturity and annotation quality, COVIDep and COVIEdb are highly recommended for vaccine target discovery, and CoV-AbDab for mAb research. Given its role in viral entry, the SARS-CoV-2 S protein has become an attractive antigen candidate. However, missense mutations in S protein alter its properties, which could also influence the epitopes. If these mutations reduce vaccine efficacy, other viral proteins will be required for antigens, such as the N and M proteins. Farrera-Soler *et al.* developed a peptide array for the epitope mapping of S protein and identified three immunodominant linear epitopes [50]. With the SARS-CoV-2 proteome, Yarmarkovich *et al.* reported 65 33-mer peptides conserved across 15 related CoVs, which are predicted to induce long-term immunity [51].

Opportunities and challenges

The COVID-19 pandemic is the third CoV outbreak over the past 20 years. Much experience has been gained from the fight against CoV. First, the crucial structures of viral proteins have been resolved, which promotes understanding of virus infection and replication and facilitates structure-based drug design [52–55]. Second, drug repurposing could substantially shorten the timeline and reduce the risks of failures in drug development. Third, the high similarity between the SARS-CoV-2 and SARS-CoV genomes makes it possible that recommended SARS-CoV-2 epitopes are similar to those epitopes for SARS-CoV [56,57]. In fact, experimental results have demonstrated that SARS-CoV immunological data could be used to inform vaccine epitopes for SARS-CoV-2 [58,59].

Although significant experience and great advances have been gained as a result of CoV epidemics, we need to realize the challenges still in our way. First, a high mutation rate is a noteworthy feature of RNA viruses [60]. This drives viral genome variability, leading to drug resistance [61,62]. Thus, we should always pay attention to the emergence of new SARS-CoV-2 variants [63]. It is necessary to study and characterize viral genome variability for recognizing the correlation with possible drug-resistance phenotypes [64]. Second, repurposing existing drugs still presents challenges [30,65]. A drug commonly affects a single pathway or multiple pathways to regulate several physiological processes in humans, which could cause unexpected adverse effects. Thus, researchers aiming to repurpose existing drugs must carefully evaluate the potential benefit. Since the outbreak of COVID-19, many studies have been conducted to pursue an effective therapeutic agent utilizing drug-repurposing methods. However, several controversial results have arisen, causing discussions within the scientific community. Scientific research needs time to design experiments carefully to understand the results. A strategy integrating rigorous peer review and the use of artificial intelligence was proposed to prioritize therapeutic research [66]. Finally, there are limited data on superinfection and re-infection associated with SAR-CoV-2. Severe infection leads to immune system dysregulation, which can leave patients vulnerable to bacterial or fungal proliferation. A handful of cases of re-infection have been reported and, thus, research is need to determine out how likely and how severe re-infection is.

Concluding remarks

Researchers have made great progress in the development of potential medical treatments in response to the COVID-19 pandemic. Rapid and open sharing of data greatly accelerates drug research and discovery. Here, we have integrated and organized the available public research resources for the development of anti-COVID-19 drugs, vaccines, and mAbs. Abundant sequence and structure data have been obtained and are maintained on several public platforms. Moreover, drugs against COVID-19 and drug targets are recorded in representative pharmaceutical databases. Specialist data repositories cover epitope information for vaccines and mAb development. On the one hand, the data accumulated about the virus have contributed to chemotherapy and immunotherapy. On the other hand, we should realize the accompanying challenges including viral genome variability, the risks of drug repurposing, and superinfection and re-infection. We still need to learn more about CoV. We anticipate that our review will contribute to stimulating therapeutic opportunities and provides a comprehensive information source for a general research audience.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (21772059, 91853127, and 31960548). Program of Introducing Talents of Discipline to Universities of China (111 Program, D20023). Frontiers Science Center for Asymmetric Synthesis and Medicinal Molecules, Department of Education, Guizhou Province [Qianjiaohe KY number (2020) 004].

References

[1] J. Zheng, SARS-CoV-2: an emerging coronavirus that causes a global threat, Int J Biol Sci 16 (2020) 1678–1685.
[2] V.M. Corman, D. Muth, D. Niemeyer, C. Drosten, Hosts and sources of endemic human coronaviruses, Adv Virus Res 100 (2018) 163–188.
[3] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: origin, transmission, and characteristics of human coronaviruses, J Adv Res 24 (2020) 91–98.
[4] M.A. Ahsan, Y. Liu, C. Feng, Y. Zhou, G. Ma, Y. Bai, et al., Bioinformatics resources facilitate understanding and harnessing clinical research of SARS-CoV-2, Brief Bioinform 22 (2021) 714–725.
[5] D. Mercatelli, A.N. Holding, F.M. Giorgi, Web tools to fight pandemics: the COVID-19 experience, Brief Bioinform 22 (2021) 690–700.
[6] F. Hufsky, K. Lamkiewicz, A. Almeida, A. Aouacheria, C. Arighi, A. Bateman, et al., Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research, Brief Bioinform 21 (2021) 642–663.
[7] L. van Dorp, M. Acman, D. Richard, L.P. Shaw, C.E. Ford, L. Ormond, et al., Emergence of genomic diversity and recurrent mutations in SARS-CoV–2, Infection Genetics Evol 83 (2020) 104351.
[8] R.A. Khailany, M. Safdar, M. Ozaslan, Genomic characterization of a novel SARS-CoV-2, Gene Rep 19 (2020) 100682–100782.
[9] B. Dearlove, E. Lewitus, H. Bai, Y. Li, D.B. Reeves, M.G. Joyce, et al., A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants, Proc Natl Acad Sci USA 117 (2020) 23652–23662.
[10] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, Viruses 12 (2020) 254.
[11] Mei LC, Wang YL, Wu FX, Wang F, Hao GF, Yang GF. HISNAPI: a bioinformatic tool for dynamic hot spot analysis in nucleic acid–protein interface with a case study. Brief Bioinform. Published January 7, 2021. https://doi.org/10.1093/bib/bbaa373.
[12] C. Wang, W. Li, D. Drabek, N.M.A. Okba, R. van Haperen, A.D.M.E. Osterhaus, et al., A human monoclonal antibody blocking SARS-CoV-2 infection, Nature Commun 11 (2020) 2251.
[13] C.O. Barnes, A.P. West Jr, K.E. Huey-Tubman, M.A.G. Hoffmann, N.G. Sharaf, P. R. Hoffman, et al., Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies, Cell 182 (2020) 828–842.
[14] UniProt Consortium, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res 47 (2019) D506–D515.
[15] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob Chall 1 (2017) 33–46.
[16] B.E. Pickett, E.L. Sadat, Y. Zhang, J.M. Noronha, R.B. Squires, V. Hunt, et al., ViPR: an open bioinformatics database and analysis resource for virology research, Nucleic Acids Res (2012). D593–8.
[17] C. Hulo, E. de Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, et al., ViralZone: a knowledge resource to understand virus diversity, Nucleic Acids Res 39 (2011) D576–D582.
[18] S. Patiyal, D. Kaur, H. Kaur, N. Sharma, A. Dhall, S. Sahai, et al., A web-based platform on Coronavirus Disease-19 to maintain predicted diagnostic, drug, and vaccine candidates, Monoclon Antib Immunodiagn Immunother 39 (2020) 204–216.
[19] Zhu Z, Meng K, Liu G, Meng G. A database resource and online analysis tools for coronaviruses on a historical and global scale. Database. Published online October 3, 2020. https://doi.org/10.1093/database/baaa070.
[20] Y. Huang, S.K. Lau, P.C. Woo, K.Y. Yuen, CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes, Nucleic Acids Res 36 (2008) D504–D511.
[21] ww PDBc, Protein Data Bank: the single global archive for 3D macromolecular structure data, Nucleic Acids Res 47 (2019) D520–D528.
[22] R. Gowthaman, J.D. Guest, R. Yin, J. Adolf-Bryfogle, W.R. Schief, B.G. Pierce, CoV3D: a database of high resolution coronavirus protein structures, Nucleic Acids Res 49 (2021) D282–D287.
[23] T. Schmidt, A. Bergner, T. Schwede, Modelling three-dimensional protein structures for applications in drug design, Drug Discov Today 19 (2014) 890–897.
[24] C. Zhang, W. Zheng, X. Huang, E.W. Bell, X. Zhou, Y. Zhang, Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1, J Proteome Res 19 (2020) 1351–1360.
[25] X. Huang, R. Pearce, Y. Zhang, De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2, Aging 12 (2020) 11263–11276.
[26] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, et al., SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res 46 (2018) W296–W303.
[27] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, et al., A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269.
[28] M. Sedova, L. Jaroszewski, A. Alisoltani, A. Godzik, Coronavirus 3D: 3D structural visualization of COVID-19 genomic divergence, Bioinformatics 36 (2020) 4360–4362.
[29] G. Ciliberto, L. Cardone, Boosting the arsenal against COVID-19 through computational drug repurposing, Drug Discov Today 25 (2020) 946–948.
[30] S. Pushpakom, F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, et al., Drug repurposing: progress, challenges and recommendations, Nat Rev Drug Discov 18 (2019) 41–58.
[31] Y. Wang, S. Zhang, F. Li, Y. Zhou, Y. Zhang, Z. Wang, et al., Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics, Nucleic Acids Res 48 (2020) D1031–D1041.
[32] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res 46 (2018) D1074–D1082.
[33] Martin R, Lochel HF, Welzel M, Hattab G, Hauschild AC, Heider D. CORDITE: The Curated CORona Drug InTERactions Database for SARS-CoV-2. iScience 2020; 23: 101297.
[34] K.R. Brimacombe, T. Zhao, R.T. Eastman, X. Hu, K. Wang, M. Backus, et al., An OpenData portal to share COVID–19 drug repurposing data in real time, bioRxiv (2020). 2020.06.04.135046.
[35] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, et al., PubChem substance and compound databases, Nucleic Acids Res 44 (2016) D1202–D1213.
[36] M. Duran-Frigola, E. Pauls, O. Guitart-Pla, M. Bertoni, V. Alcalde, D. Amat, et al., Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker, Nat Biotechnol 38 (2020) 1087–1096.
[37] J.F. Yang, F. Wang, Y.Z. Chen, G.F. Hao, G.F. Yang, LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor, Brief Bioinform 21 (2020) 2206–2218.
[38] T.F. Chen, Y.C. Chang, Y. Hsiao, K.H. Lee, Y.C. Hsiao, Y.H. Lin, et al., DockCoV2: a drug database against SARS-CoV-2, Nucleic Acids Res 49 (2021) D1152–D1159.
[39] D. Pinto, Y.-J. Park, M. Beltramello, A.C. Walls, M.A. Tortorici, S. Bianchi, et al., Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody, Nature 583 (2020) 290–295.
[40] W.S. Lee, A.K. Wheatley, S.J. Kent, B.J. DeKosky, Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies, Nat Microbiol 5 (2020) 1185–1191.
[41] S.F. Ahmed, A.A. Quadeer, M.R. McKay, COVIDep: a web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2, Nat Protoc 15 (2020) 2141–2142.
[42] S. Sahoo, S.R. Mahapatra, B.K. Parida, S. Rath, B. Dehury, V. Raina, et al., DBCOVP: a database of coronavirus virulent glycoproteins, Comput Biol Med 129 (2020) 104131.
[43] H. Singh, H.R. Ansari, G.P.S. Raghava, Improved method for linear B-cell epitope prediction using antigen's primary sequence, PLoS ONE 8 (2013) e62216.
[44] M. Bhasin, G.P.S. Raghava, Prediction of CTL epitopes using QM, SVM and ANN techniques, Vaccine 22 (2004) 3195–3204.
[45] H. Singh, G.P.S. Raghava, ProPred: prediction of HLA-DR binding sites, Bioinformatics 17 (2001) 1236–1237.

[46] J. Wu, W. Chen, J. Zhou, W. Zhao, Y. Sun, H. Zhu, et al., COVIEdb: a database for potential immune epitopes of coronaviruses. Data report, Front Pharmacol 11 (2020) 572249.
[47] M.V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, M. Nielsen, Large-scale validation of methods for cytotoxic T–lymphocyte epitope prediction, BMC Bioinformatics 8 (2007) 424.
[48] R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, et al., The immune epitope database (IEDB) 3.0, Nucleic Acids Res 43 (2015) D405–D412.
[49] Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the Coronavirus Antibody Database. Bioinformatics. Published online August 17; 2020. http://dx.doi.org/10.1093/bioinformatics/btaa739.
[50] L. Farrera-Soler, J.P. Daguer, S. Barluenga, O. Vadas, P. Cohen, S. Pagano, et al., Identification of immunodominant linear epitopes from SARS-CoV–2 patient plasma, PLoS ONE 15 (2020) e0238089.
[51] M. Yarmarkovich, J.M. Warrington, A. Farrel, J.M. Maris, Identification of SARS-CoV-2 vaccine epitopes predicted to induce long-term population-scale immunity, Cell Rep Med 1 (2020) 100036.
[52] W. Yin, C. Mao, X. Luan, D.D. Shen, Q. Shen, H. Su, et al., Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir, Science 368 (2020) 1499–1504.
[53] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, et al., Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (2020) 221–224.
[54] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (2020) 215–220.
[55] H.S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, P. Cramer, Structure of replicating SARS-CoV-2 polymerase, Nature 584 (2020) 154–156.
[56] Y.-Z. Zhang, E.C. Holmes, A genomic perspective on the origin and emergence of SARS-CoV-2, Cell 181 (2020) 223–227.
[57] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574.
[58] T.R.F. Smith, A. Patel, S. Ramos, D. Elwood, X. Zhu, J. Yan, et al., Immunogenicity of a DNA vaccine candidate for COVID-19, Nat Commun 11 (2020) 2601.
[59] W. Chour, A.M. Xu, A.H.C. Ng, J. Choi, J. Xie, D. Yuan, et al., Shared antigen-specific CD8+ T cell responses against the SARS-COV-2 spike protein in HLA A*02: 01 COVID-19 participants, medRxiv (2020). 2020.05.04.20085779.
[60] S. Duffy, Why are RNA virus mutation rates so damn high?, PLoS Biol 16 (2018) e3000003
[61] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182 (2020) 812–827.
[62] J. Chen, R. Wang, M. Wang, G.W. Wei, Mutations Strengthened SARS-CoV-2 Infectivity, J Mol Biol 432 (2020) 5212–5226.
[63] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, et al., The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity, Cell 182 (2020) 1284–1294.
[64] F.X. Wu, F. Wang, J.F. Yang, W. Jiang, M.Y. Wang, C.Y. Jia, et al., AIMMS suite: a web server dedicated for prediction of drug resistance on protein mutation, Brief Bioinform 21 (2018) 318–328.
[65] T.I. Oprea, J.E. Bauman, C.G. Bologa, T. Buranda, A. Chigaev, B.S. Edwards, et al., Drug repurposing from an academic perspective, Drug Discov Today Ther Strateg 8 (2011) 61–69.
[66] J.M. Levin, T.I. Oprea, S. Davidovich, T. Clozel, J.P. Overington, Q. Vanhaelen, et al., Artificial intelligence, drug repurposing and peer review, Nat Biotechnol 38 (2020) 1127–1131.