Virus CKB 2.0: Viral-associated disease-specific chemogenomics knowledgebase

by

Yixuan Hao

Bachelor of Science, East China University of Science and Technology, 2020

Submitted to the Graduate Faculty of the

School of Pharmacy in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis or dissertation was presented

by

Yixuan Hao

It was defended on

March 30, 2022

and approved by

Xiang-Qun (Sean) Xie, Ph.D., EMBA, Department of Pharmaceutical Sciences

Levent Kirisci, Ph.D., Professor, Department of Pharmaceutical Sciences

Junmei Wang, Ph.D., Associate Professor, Department of Pharmaceutical Sciences

Zhiwei Feng, Ph.D., Assistant Professor, Department of Pharmaceutical Sciences

Thesis Advisor/Dissertation Director:
Zhiwei Feng, Ph.D., Assistant Professor, Department of Pharmaceutical Sciences

Virus CKB 2.0: Viral-associated disease-specific chemogenomics knowledgebase

Yixuan Hao, BS

University of Pittsburgh, 2022

ABSTRACT

Given that mutations in viruses always occur during evolution, the doctor calls for a flu vaccine every year. Vaccines obtained from diseased strains have a protective effect of 40%-60%. However, if the diseased strain is mutated, the vaccine will lose its effect. Therefore, it is very necessary to track the data and structure of different viruses for research purposes. Viruses have different domains that are related to the invasion or replication of the virus, so drugs targeting these different domains can provide therapeutic potential for the treatment of diseases caused by these viruses. To facilitate the research of viral infection disease, we updated our Virus-CKB 1.0 to 2.0 by adding structure information of another ten viruses and integrating new functions (e.g., protein-protein docking). To date, Virus-CKB 2.0 archived 65 antiviral drugs in the market, 178 viral-related targets with 291 available 3D crystal or cryo-EM structures, and 3766 chemical agents reported for these target proteins. Moreover, Virus-CKB 2.0 is implemented with web applications for the prediction of the relevant protein targets and analysis and visualization of the outputs, including Molecular Complex Characterizing System (MCCS) protocol, HTDocking, TargetHunter, blood-brain barrier (BBB) predictor, Spider Plot, Protein-Protein docking, etc. The Virus-CKB 2.0 server is accessible at https://www.cbligand.org/g/virus-ckb. By using the established chemogenomics tools and algorithms and newly developed tools, we can screen FDA-approved drugs and chemical compounds that may bind to these proteins involved in viral-associated diseases regulation. If the virus strain is mutated and the vaccine loses its effect, we can

also screen out drugs that can be used to treat the mutated virus in a short time. For some FDA-approved drugs, we can also perform drug repurposing through Virus-CKB 2.0.

Keywords: virus knowledgebase, dengue virus, online drug discovery, drug repurposing

Table of Contents

List of Tables

List of Figures

PREFACE

I would like to express my sincerest gratitude to my advisor Dr. Zhiwei Feng, an expert in computer-aided drug design, Computational Chemical Genomics Screening Center (CCGS) member, and assistant professor at the University of Pittsburgh School of Pharmacy. During my study in Pharmacometrics and Systems Pharmacology track, I have learned a lot from Dr. Feng, not only in study but also in life. Dr. Feng taught me basic software and several useful computational prediction methods. When I met problems in my research, Dr. Feng taught me to think critically and provided me with thoughtful ideas, which really encouraged me and made me more confident. More importantly, Dr. Feng always taught me to combine work to exercise, which really helped me a lot when I felt stressed.

Also, I would like to express my sincerest gratitude to Dr. Xiangqun Xie. As the director of the Computational Chemical Genomics Screening Center (CCGS), he always provided me with some thoughtful ideas and some latest findings in the scientific area, which really inspired me to find the direction of my research.

Meanwhile, I would like to thank my committee member, Dr. Junmei Wang, and Dr. Levent Kirisci. Dr. Junmei Wang is an expert in molecular dynamics simulation and force field study. The Pharmacometrics and Systems Pharmacology courses he taught gave me a general understanding of the introduction to R and Matlab Programming, Big data analysis: Data Dimension Reduction and Feature Selection, Machine Learning in Biomedical Data Analysis, and creating and simulating models using Simbiology. Dr. Levent Kirisci is the professor of the Statistical method class. The statistical method course he taught gave me a general understanding of several very useful statistical methods that really help me a lot in my research and study.

Besides, I want to express my sincere acknowledgment to all the members of the CCGS center. Especially Maozi Chen, an expert in programming, software, and algorisms development in our center. He developed the MCCS program and gave me step-by-step instructions for basic software usage, construction of the databases, and provided me with lots of technical support. I also want to thank, Dr. Terence McGuire, Dr. Jaden Jun, Dr. Ying Xue, and Dr. Lirong Wang, who are the professors in the CCGS center. They taught me much new knowledge from the other areas and gave lots of useful advice for my research. Also, I would like to thank the support and assistance from my peers during my two-year studies.

Lastly, as the old saying goes, parents are always the strongest backing for children. I would like to thank my parents for their support during my master's studies.

I offer my best regards to all of those who gave me a hand during my two-year master’s studies!

1.0 INTRODUCTION

1.1 Vaccines

The doctor calls for a flu vaccine every year due to the mutations of the virus always occur during evolution. Vaccines obtained from the diseased strains have a protective effect of 40%-60%. However, if the diseased strain is mutated, the vaccine will lose its effect. Therefore, it is very necessary to track the data and structure of different viruses for a long time to meet with the research purpose.

1.2 Virus

Viruses have different domains related to the invasion or replication of the viruses, so drugs targeting on these different domains can provide therapeutic potential for the treatment of diseases caused by these viruses.

1.2.1 SARS-CoV-2

The pandemic situation of SARS-CoV-2 infection has sparked global concern due to the disease COVID-19 caused by it. Since the first cluster of confirmed cases in China in December 2019, the infection has been reported across the continents and inflicted upon a substantial number of populations. However, the origin of this pandemic is unclear as many reports on the detection

of the coronavirus much earlier than December 2019. COVID 19, an ongoing pandemic, is a viral disease attributed to severe acute respiratory syndrome[1].

SARS-CoV-2, the causative agent of COVID-19, has four structural proteins: spike protein, envelope protein, membrane protein, and nucleocapsid protein. Spike protein plays an important role in the attachment of virus during the process of virus infection[2]. The SARS-CoV-2 genome consists of 30 kilobases of positive-sense, single-stranded linear RNA[3-5]. The SARS-CoV-2 genome encodes several polyproteins that can be processed by viral proteases and form non-structural proteins. These non-structural proteins can form a replicase complex that plays an important role during the transcription and replication of the viral genome [4]. Remdesivir has been shown to inhibit RNA-dependent RNA polymerase (RdRp) that composed of non-structural protein 12 (NS12) with high potency[6, 7]. Therefore, drugs targeting these proteins may provide therapeutic potential for the treatment of COVID-19.

1.2.2 Zika Virus

Zika virus (ZIKV) is a mosquito-borne *Flavivirus* that can cause Zika fever[8]. There were sporadic Zika virus outbreaks in Africa, Southeast Asia, and the Pacific Islands before 2015. Major outbreaks of the Zika virus have been occurring in the Americas since May 2015. Currently, 30 countries have active transmission of the Zika virus, while many other countries have reported travel-related cases. Zika virus is spread primarily through the bite of an infected mosquito, rarely from mother to child, infrequently through blood transfusion and sexual contact[9].

The Zika virus genome consists of 10.8 kilobases of positive-sense, single-stranded RNA[10, 11]. The 10.8 kilobase RNA genome is translated into a long and single polyprotein. It is post-translationally cleaved by host and viral proteases into three structural proteins: pre-

membrane protein (prM), envelope protein (E) and the capsid protein (C), and seven non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 which are involved in viral RNA replication. Zika virus non-structural protein 5 RNA-dependent PNA polymerase (RdRp) plays an important role in the replication of Zika RNA[12]. Galidesivir, a direct-acting antiviral drug, has been confirmed to treat Zika virus infection by inhibiting the non-structural protein 5 RNA-dependent PNA polymerase (RdRp)[13]. Therefore, drugs targeting these proteins may have the potential to treat Zika fever.

1.2.3 Ebola Virus

Ebola virus (EBOV) is a *filoviridae* virus that can cause Ebola virus disease, also known as Ebola hemorrhagic fever[14]. Ebola virus outbreaks have sporadically appeared in Africa. Ebola virus is spread through direct contact with infected blood, organs, or body fluids of infected people or animals, or objects contaminated by the virus.

The Ebola virus consists of a non-segmented single-stranded negative-sense RNA genome of approximately 19 kilobases, which is wrapped by nucleoprotein (NP) and other viral proteins to form a spiral nucleocapsid (NC)[15]. Ebola virus contains a viral envelope, matrix, and nucleocapsid components. Ebola virus RNA encodes seven structural proteins: nucleoprotein (NP), polymerase cofactor (VP35), (VP40), glycoprotein (GP), transcription activator (VP30), VP24, and RNA polymerase (L)[15, 16]. Nucleocapsid plays an important role in the assembly of virus, transcription, and replication of the Ebola genome. Glycoproteins are responsible for the attachment of the Ebola virus and therefore allowing them to enter new host cells. Several monoclonal antibodies such as atoltivimab, odesivimab, and maftivimab that targeting glycoprotein of Ebola virus have been approved by FDA for treatment of Ebola virus infection in

adult[17]. Therefore, drugs targeting these proteins may have the potential to treat the Ebola virus disease.

1.2.4 Dengue Virus

Dengue virus (DENV) is a mosquito-borne *Flavivirus* that can cause dengue fever, dengue hemorrhagic fever, and dengue shock syndrome[18, 19]. Dengue virus is common in more than 100 countries around the world. Each year, up to 400 million people get infected with the dengue virus. Approximately 100 million people get sick from infection, and 22,000 die from severe dengue. There are four serotypes of dengue viruses, dengue virus 1, 2, 3, and 4[20]. Each serotype is antigenically different. For this reason, a person can be infected with a dengue virus as many as four times in his or her lifetime.

The dengue virus genome consists of 10.7 kilobases of positive-sense, single-stranded RNA. This RNA encodes 3 structural proteins: pre-membrane protein (prM), envelope protein (E), and the capsid protein (C), which constitute the components of the virion, and 7 non-structural proteins: NS1, NS2A/B, NS3, NS4A/B, NS5 which are involved in viral RNA replication[21]. Fucoidan, a sulfated polysaccharide, has been confirmed to inhibit type 2 dengue virus infection by targeting the envelop protein of type 2 dengue virus and preventing the viral binding and entry into the host cells[22]. Therefore, drugs targeting these proteins may have the potential to treat dengue fever, dengue hemorrhagic fever, and dengue shock syndrome.

1.2.5 Enterovirus

Enterovirus (EV) are members of the *picornavirus* family that can cause a wide range of illnesses [23]. The fecal-oral route and the respiratory transmission are the two main routes by which enteroviruses infect humans, and they can also target gastrointestinal epithelial cells early in their life cycle[24].

The Enterovirus genome consists of 7.2-8.5 kilobases positive-sense, single-stranded RNA. The open reading frame of the Enterovirus encodes a large polyprotein, which is further hydrolyzed to form three precursor proteins, including precursor protein 1 (P1), precursor protein 2 (P2), and precursor protein 3 (P3)[25]. The capsid is made up of 60 protomers, each consisting of 4 polypeptides that comprise the structural viral proteins, viral protein 1 (VP1), viral protein 2 (VP2), viral protein 3 (VP3), and viral protein 4 (VP4). The structural proteins are encoded by the P1 region of the genome. Of all the polypeptides, VP4 is located on the internal side of the capsid while VP1, VP2, and VP3 are located on the external surface of the Enterovirus. Proteins derived from the non-structural P2 (2A-2C) and P3 (3A-3D) regions are most directly involved in virus replication, structural and biochemical changes, and apoptosis of infected cells. Capsids are promising targets for direct acting anti-Enterovirus therapy[26]. SCH 48973, a potent enterovirus capsid binder, has been proven to exhibit anti-Enterovirus efficacy[27].

1.2.6 Lassa Virus

Lassa virus (LASV) is a member of the *Arenaviridae* family that can cause Lassa fever characterized by shock, bleeding, and multiorgan failure[28, 29]. Rat is the main species that can transmit the Lassa virus to humans and cause Lassa fever.

The Lassa virus genome consists of a L segment of about 7.3 kilobases and a S segment of 3.5 kilobases. Lassa virus is a linear, segmented, single-stranded RNA virus that has a protective envelope. Lassa virus consists of glycoprotein, matrix protein, nucleoprotein, ribonucleocapsid, polymerase, and genomic RNA[30].

1.2.7 Norovirus

Norovirus (NORV) is a member of the *Caliciviridae* family that can cause gastroenteritis[31, 32]. The murine and canine norovirus are two main noroviruses that can cause enteric disease in humans. Fecal-oral spread is the main transmission of norovirus among humans, infectious vomitus can also participate in the transmission of norovirus.

Norovirus genome contains a linear, non-segmented, positive-sense RNA genome of approximately 7.5 kilobases[33], encoding three open reading frames. The genome is protein-linked at the 5′ end and polyadenylated at the 3′ end. Open reading frames 1 encodes the nonstructural proteins including NS1 to NS7 that are processed by the viral 3C-like protease (3CLpro). Open reading frames 2 encodes the major capsid protein VP1 and Open reading frames 3 encodes the minor structural capsid protein VP2. The norovirus NS6 protease is a promising antiviral target that play an important role in viral replication. Rupintrivir, a protease inhibitor, has been conformed to exhibit anti-norovirus activity[34].

1.2.8 Andes Orthohantavirus

Andes orthohantavirus (ADNV) is a species of *Orthohantavirus*, and a member of the *Hantaviridae* family, which is the principal pathogen of respiratory diseases such as hantavirus

pulmonary syndrome (HPS) and hantavirus cardiopulmonary syndrome (HCPS) in South America[35]. Andes orthohantavirus is mainly transmitted by direct contact with infected rodents or their fecal waste and is the only hantavirus that can be transmitted from infected humans to healthy humans.

Andes orthohantavirus genome consists of enveloped, negative-sense single-stranded RNA. The segmented genome includes small segment (S), medium segment (M), and large segment (L) encoding nucleocapsid (N), glycoproteins G1 and G2, and L protein respectively. The M segment generates glycoprotein precursor (GPC), which can be cleaved into envelope Gn and Gc proteins[36]. The L protein encoded by the L segment has enzymatic functions involved in transcription and replication[36].

1.2.9 Herpes Simplex Virus

Herpes simplex virus (HSV) belongs to the *Herpesviridae* family, and herpes simplex virus 1 (HSV-1) and herpes simplex virus 2 (HSV-2) are the most typical herpes simplex viruses that can produce viral infections among humans [37]. HSV-1 is commonly associated with epithelial lesions, orofacial infections, and encephalitis, while HSV-2 is the common causing agent of genital infection and can be transmitted from infected mothers to infants[38].

HSV-1 and HSV-2 genomes contain linear, double-stranded DNA. The DNA molecule of HSV-1 and HSV-2 consists of a long segment and a short segment covalently bonded with each other. The HSV virion has four parts: DNA, capsid, tegument, and envelope[39]. The linear DNA genome is encased in the capsid. The capsid is encapsulated in a lipid bilayer called the envelope. The envelope is connected to the capsid through tegument. The whole particle is the virion of HSV. HSV attachment and penetration into host cells are mediated by viral surface glycoproteins,

which then induce immune responses. Replication of HSV involves the docking of the HSV capsid into the host cell's nuclear pore complex (NPC), and then the HSV genome enters the nucleoplasm through the nuclear pore and replicates in the nucleoplasm[40]. Thymol-related monoterpenoids are shown to directly affect the virion of HSV-1 and exhibit significant anti-HSV-1 activity[41].

1.2.10 Hepatitis C Virus

Hepatitis C virus (HCV) belongs to the genus *Hepacivirus*, a member of the family *Flaviviridae*. Hepatitis C virus can cause cirrhosis, hepatocellular carcinoma, and lymphomas[42, 43]. Contamination of blood and other body fluids from injection drugs has been the predominant risk factor for the transmission of HCV[44].

Hepatitis C Virus genome consists of 9.6 kilobase positive-sense, single-stranded RNA, and the genome of HCV contains only one single open reading frame that can be translated to produce a single active protein[45]. The hepatitis C virus particle consists of RNA, capsid, and a lipid membrane envelope. RNA materials are wrapped in the capsid. Two viral envelope glycoproteins, E1 and E2 are covalently bond with each other when they are embedded in the lipid envelope[46]. These glycoproteins play an important role in the viral attachment, entry into the host cell, and interactions of hepatitis C with the immune system[47]. The hepatitis C virus also contains a set of non-structural proteins that play important role in the replication of HCV. Sofosbuvir, a NS5B nucleotide polymerase inhibitor, has been conformed to treat infections caused by HCV-genotype 1, 2, 3, and 4[48].

1.2.11 Human Immunodeficiency Virus

Human Immunodeficiency Virus (HIV) is a member of the genus *Lentivirus*, part of the family *Retroviridae* that can attack the body's immune system[49, 50]. If HIV is not treated, it can lead to AIDS (acquired immunodeficiency syndrome)[51]. Sexual contact across body fluids, maternal-infant exposure, and percutaneous inoculation are the main modes of HIV transmission[52].

The HIV genome consists of two copies of positive-sense, single-stranded RNA that can encode nine genes of the HIV [53]. The reverse transcriptase and RNA of HIV are wrapped in the protective protein called the capsid. The enzyme surrounding the capsid are the enzymes involved in HIV infection and replication. These structures are surrounded by an envelope that coats the glycoproteins, which are involved in the identification and binding of the HIV to its targets[54]. The HIV also contains a set of non-structural proteins that play important role in the replication of HIV. Boceprevir, a first-generation NS3/4A protease inhibitors, has been authorized for their use in the treatment in HIV infections[48].

1.2.12 Influenza Virus

Influenza virus comprises four species: Influenza A virus (IAV) belonging to genus Alphainfluenzavirus, Influenza B virus (IBV) belonging to genus Betainfluenzavirus, Influenza C virus (ICV) belonging to genus Gammainfluenzavirus, Influenza D virus (IDV) belonging to genus Deltainfluenzavirus[55, 56]. IAV is the predominant pathogen for most severe illnesses as well as seasonal epidemics and occasional pandemics[57].

The influenza virus genome consists of segmented, negative-sense, single-stranded RNA. IAV and IBV have 8 genomic segments encoding 10 major proteins. ICV and IDV have seven genome segments encoding nine major proteins[58]. The influenza virion is an enveloped virus that derives its lipid bilayer from the plasma membrane of a host cell. Two different varieties of glycoprotein spike are embedded in the envelope.

1.3 Virus-CKB 1.0

1.3.1 The Current Version of Virus-CKB 1.0

To facilitate the COVID-19 research, we have established an integrated viral-associated disease-specific chemogenomics knowledgebase published in July 2020[59]. Currently, our former version Virus-CKB 1.0 has totally archived 65 antiviral drugs in the market and 107 viral-related targets, including 6 HIV-related targets, 7 BCV/HCV-related targets, 49 influenza-related targets, 39 coronavirus-related targets with 189 available 3D crystal or cryo-EM structures. Moreover, 2,698 chemical agents reported for these target proteins are also included in Virus-CKB 1.0.

1.3.2 Disadvantages of Virus-CKB 1.0 that Need Improvement

Up to date, Virus-CKB 1.0 was published for more than a year to facilitate the research of COVID-19, and many targets collected at that time have not been incorporated into Virus-CKB 1.0. In addition, many target information, 3D structures, and related active compounds of SARS-

CoV-2 have been released during the past one year. For example, the 3CL protease crystal structure of SARS-CoV-2 is over 240 with more than 8000 reported compounds until now[60], the nonstructural protein crystal structure of SARS-CoV-2 is over 250 with more than 800 reported compounds until now[61], the spike protein crystal structure of SARS-CoV-2 is over 170 with more than 800 reported compounds until now[62].

Virus-CKB can also be applied to the research of other viruses that can cause viral infectious diseases. As the years go by, information and related structures of many other viruses have also been published, such as Dengue virus[18, 21], Zika virus[10, 11], Ebola virus[15, 16], Enterovirus[63, 64], Lassa virus[65, 66], Norovirus[67, 68], Herpes simplex virus[37-40], and Andes orthohantavirus. These viruses have not been included in Virus-CKB 1.0. Integrating information and structures of different viruses into Virus-CKB 1.0 can facilitate the research of virus and viral infectious diseases. For example, Zika virus (ZIKV) is a mosquito-borne Flavivirus that can cause Zika fever[10]. Zika virus is spread mainly through the bite of an infected mosquito. In addition, Ebolavirus (EBOV) is a filoviridae virus that can cause Ebola virus disease, also known as Ebola hemorrhagic fever[15]. Ebola virus outbreaks have sporadically appeared in Africa. Ebola virus is spread mainly through direct contact with infected blood or body fluids.

1.4 Aim

In this project, we updated Virus-CKB 1.0 to Virus-CKB 2.0 to facilitate the research of the viral infectious diseases. Using the established chemogenomics tools and algorithms and newly developed tools, we can screen FDA-approved drugs and chemical compounds that can bind to these proteins involved in viral-associated diseases regulation. If the virus strain is mutated and

the vaccine/drug loses its effect, we can also screen out drugs that can be used to treat the mutated virus in a short time. For some FDA-approved drugs, we can also perform drug repurposing through Virus-CKB 2.0.

2.0 MATERIAL AND METHODS

2.1 Genes and Domain Structures

The genome of a virus can encode the proteins of the virus. Viral proteins undergo post-translational modifications to form different domain proteins of the virus with different functions. Direct-acting antiviral agents are drugs that act directly on viral proteins, which are considered the most promising and effective antiviral treatment[69]. Screening small molecules that can directly act on viral proteins with different functions is a common method for antiviral drug discovery. Therefore, we focus on the interaction between small molecules and viral proteins and collect protein structures based on the different domain information of viruses. For example, the dengue virus genome encodes 3 structural proteins: pre-membrane (prM), envelope (E), and the capsid (C), which constitute the components of the virion, and 7 non-structural proteins: NS1, NS2A/B, NS3, NS4A/B, NS5 which involved in viral RNA replication[21]. There are four serotypes of the dengue viruses dengue virus 1, 2, 3, and 4. Each serotype is antigenically different[19]. When we collect the domain structure of the dengue virus from Protein Data Bank[70] (https://www.rcsb.org/), there are more than 24000 PDB files. First, we divided the PDB files according to the domain information of the dengue viruses such as pre-membrane (prM), envelope (E), capsid (C), and 7 non-structural proteins: NS1, NS2A/B, NS3, NS4A/B, NS5. Then we selected three 3D crystal or cryo-EM structures with the highest resolution and most diverse structure from each domain group and incorporated them into Virus-CKB 2.0.

We collect data such as chemical molecules, antibodies, genes, and proteins involved in viral-associated diseases regulation from several public databases such as Protein Data Bank[70]

(https://www.rcsb.org/), UniProt[71] (https://www.uniprot.org/), ChEMBL[72] (https://www.ebi.ac.uk/chembl/), therapeutic target database[73-77] (http://db.idrblab.net/ttd/), and recent literature. In the current version, we collected 178 viral-related targets with 291 available 3D crystal or cryo-EM structures that could be the potential targets for viral-related diseases, and 3766 chemical agents reported for these target proteins.

2.2 Workflow of Virus-CKB 2.0

2.2.1 Web Applications and Existing Algorithms Implemented in Virus-CKB 2.0

The Virus-CKB 2.0 server is implemented with several web applications and existing algorithms including Molecular Complex Characterizing System (MCCS) protocol[78, 79], High-throughput docking (HTDocking)[80, 81], TargetHunter[82], blood-brain barrier (BBB) predictor[83-85], Spider Plot[86], Protein-Protein docking, and other third-party software including Open Babel[87, 88], JSME Molecular Editor[89], idock[90], and NGL viewer[91], which enable the prediction of the protein-ligand interaction and the visualization of the results.

2.2.2 Workflow of Virus-CKB 2.0

Our platform mainly focuses on the 3D crystal or cryo-EM structures of viral proteins and their related compounds. Workflow of the Virus-CKB 2.0 server consists of three major steps. The first step is to input compounds or chemical agents that may have the potential to bind with the protein structures of the viruses. Figure 1 shows the job creation page of Virus-CKB 2.0. The user

can submit up to five compounds or chemical agents through the job creation page. The user can define the name of the job and the name of the molecule. The user can define the molecule structure by drawing it with JSME Molecular Editor[92] or paste any pre-existing molecule structure in MOL, SDF, or SMILES format via the dropdown menu of the blue double-triangle icon on the toolbar. In this way, user need to be aware that the pasted code should not contain any leading or trailing spaces. If the user does not want to publish the test results of the query compounds on Virus-CKB 2.0, they can choose to click Private job to hide the test results. After inputting the structures of query compounds, our platform will automatically convert the input format from MOL, SDF, or SMILES into PDB and PDBQT formats with the help of Open Babel and conduct further computational prediction.

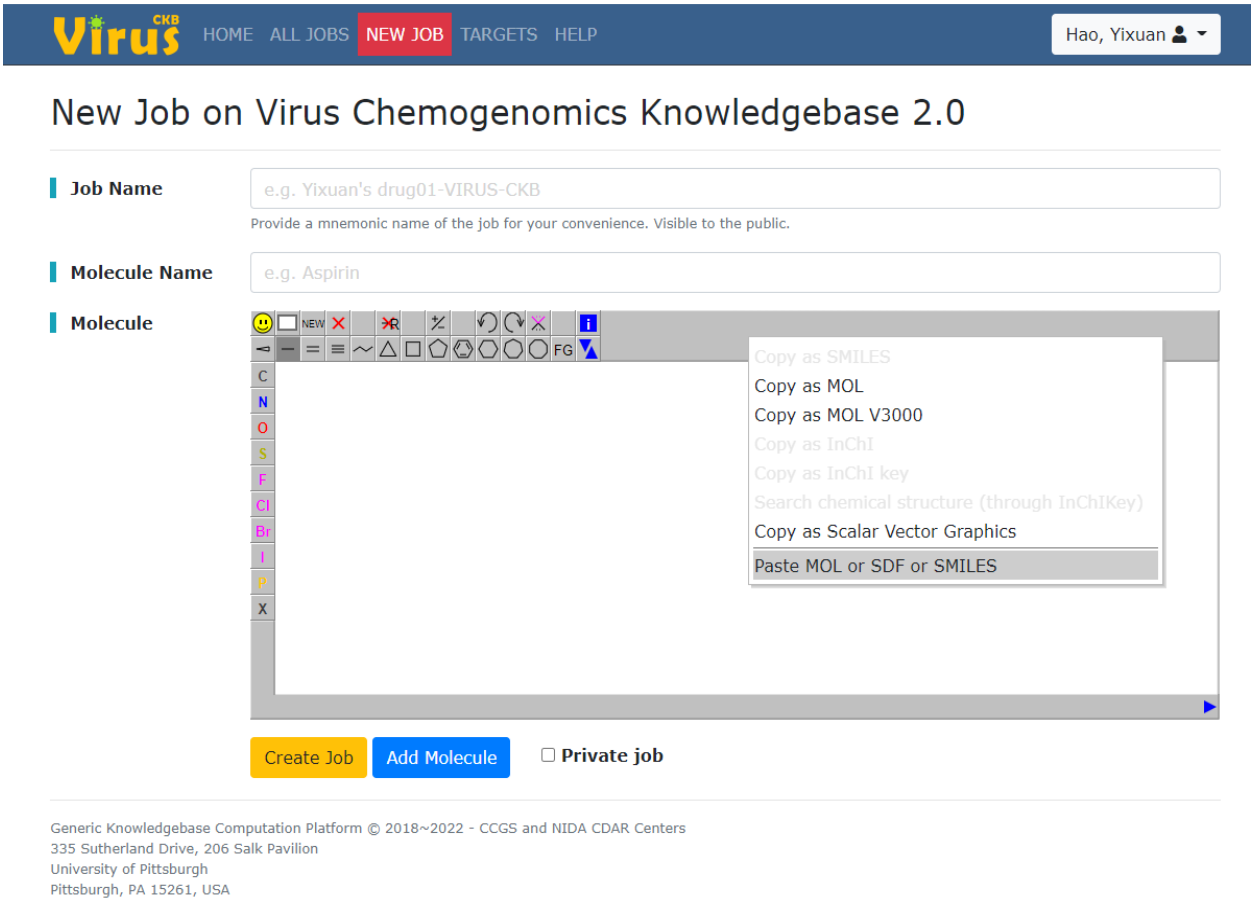


Figure 1 The job creation page

From top to bottom are job name input box, molecule name input box, molecule 2D structure editor, create job button, add molecule button, and private job option.

The second step is to conduct Molecular Complex Characterizing System (MCCS) protocol, in-silico blood-brain barrier (BBB) prediction, High-throughput docking (HTDocking), and fingerprints-based similarity search simultaneously by our established algorithms implemented in Virus-CKB 2.0. MCCS is an in-silico scoring function-based algorism that can recognize and manifest the characterization of the receptor-ligand binding[79]. The MCCS algorithm can calculate the residue energy contribution, generate the binding recognition vectors, perform docking, and calculate the similarity of binding features between two complexes[78]. The in silico blood-brain barrier (BBB) prediction can be used to predict whether the query compounds can pass through the blood-brain barrier and therefore cause side effects in the central nervous system (CNS)[83-85]. High-throughput docking (HTDocking) is a web-based high-throughput computing method that can predict the potential interactions between the user-inputted compounds and target proteins. In Virus CKB 2.0, each virus domain includes up to three conformations with the highest resolution and the most diverse structures. HTDocking will dock up to five user-inputted compounds into binding pockets of up to three different conformations and generate docking scores respectively. HTDocking can provide up to nine potential docking scores with different binding poses of the same protein pocket and generate the average docking score. According to the convention, the docking score is negative and the smaller value (larger absolute value) represents the better binding pose[80]. TargetHunter is a web-based target prediction tool that can predict the potential biotargets of submitted compounds based on targets associated with its most similar counterparts (TAMOSIC) algorithm[82]. The prediction is based on an important principle of medicinal chemistry: Structurally similar compounds have similar physicochemical properties and may cause similar biological effects. When the query compound is uploaded, TargetHunter

will first search the similar compounds in the public database such as the ChEMBL library, then compares similar compounds with the query compound, generates molecular fingerprints using Tanimoto coefficients, and calculates the similarity scores based on Tanimoto coefficients, and finally list the potential biotargets.

The third step is to conduct systems pharmacology target mapping for potential drug repurposing, and drug combination. After finishing all the in-silico computation processes, we can analyze the docking scores, similarity scores, which can be integrated to generate the Spider Plot for the visualization of the compound-targets interaction network. Spider Plot is an online tool that can visualize the molecule-protein interaction network based on the target classification[86]. Using spider plot, users can intuitively see the in-silico computational prediction results generated by Virus-CKB 2.0.

2.3 Drugs and Chemicals

Using “Dengue”, “Zika” or any other virus’s name as keywords, we searched the DrugBank[93] database and ChEMBL[72] database and retrieved 3766 antiviral drugs or viral-related chemical agents and incorporated them into Virus-CKB 2.0. Antiviral drugs or chemical agents with IC50 lower than 1 μM toward a viral-related target are defined as active chemical agents, while those chemical agents with IC50 larger than 10 μM are defined as inactive chemical agents. Therefore, we only retrieved chemical agents with IC50 lower than 10 μM into Virus-CKB 2.0.

3.0 RESULTS

3.1 Virus-CKB 2.0 Overview

Currently, we have collected 178 viral-related targets with 291 available 3D crystal or cryo-EM structures, which include 6 HIV-related targets, 5 Herpes simplex virus-related targets, 3 Andes orthohantavirus-related targets, 4 Lassa virus-related targets, 15 Dengue virus-related targets, 4 Norovirus-related targets, 11 Ebolavirus-related targets, 9 Zika virus-related targets, 13 Hepatitis C virus-related targets, 12 Enterovirus-related targets, 37 Coronavirus-related targets that include 6 targets from SARS, 8 targets from SARS-CoV-2, and 23 targets from other coronaviruses, and 49 Influenza virus-related targets as shown in Figure 2. Since the dengue virus has four serotypes and each serotype is antigenically different, so we divided the targets based on the serotypes of the dengue virus. We collected 15 dengue virus related targets including Capsid protein if dengue virus type 2 (PDB ID: 1R6R), Envelope protein of flavivirus (PDB ID: 6DFJ), Envelope protein of dengue virus type 1 (PDB ID: 3UZQ), Envelope protein of dengue virus type 2 (PDB ID: 1OK8), Envelope protein of dengue virus type 3 (PDB ID: 3VTT), Membrane protein of dengue virus type 2 (PDB ID: 3C6E), NS1 of dengue virus type 2 (PDB ID: 7BSC), NS1 of dengue virus type 1 (PDB ID: 4OIG), NS2A of dengue virus type 2 (PDB ID: 2M0S), NS2B of dengue virus type 1 (PDB ID: 3LKW), NS2B of dengue virus type 2 (PDB ID: 4M9K, 4M9T), NS3 of flavivirus (PDB ID: 5YVU), NS3 of dengue virus type 4 (PDB ID: 2JLQ, 5YVW), NS5 of dengue virus type 2 (PDB ID: 5ZQK), NS5 of dengue virus type 3 (PDB ID: 2J7U, 5JJS). These target proteins are directly involved in the process of dengue virus infection or replication. Therefore, drugs targeting on these proteins may have the potential to produce therapeutic effects

towards dengue virus infection. Figure 3 shows all the viruses and their related target protein names archived in Virus-CKB 2.0. The specific name of each target, original organism, Gene, UniProt ID, and ChEMBL ID are listed in Appendix Table 1.

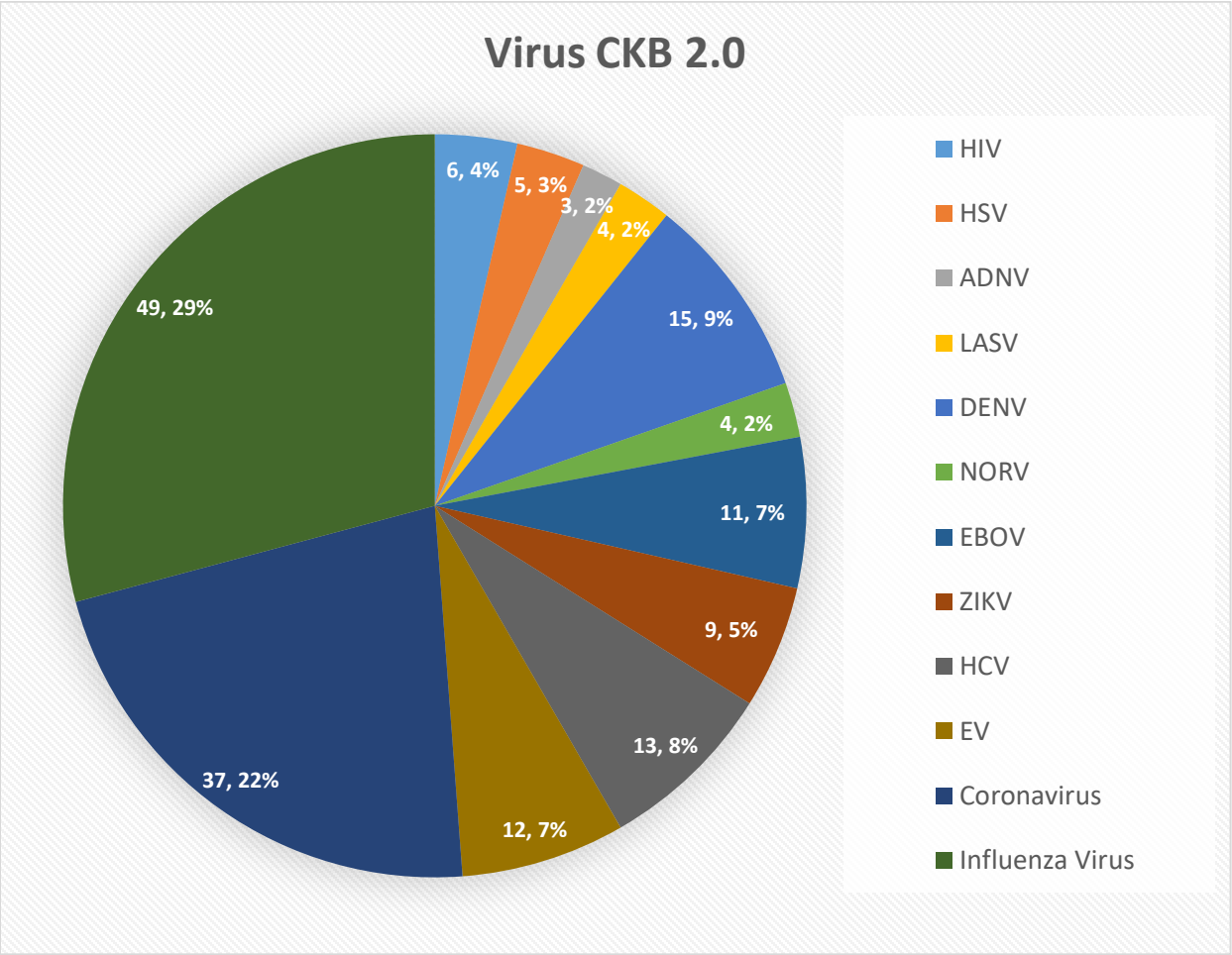


Figure 2 The number of targets contained in each virus in Virus-CKB 2.0

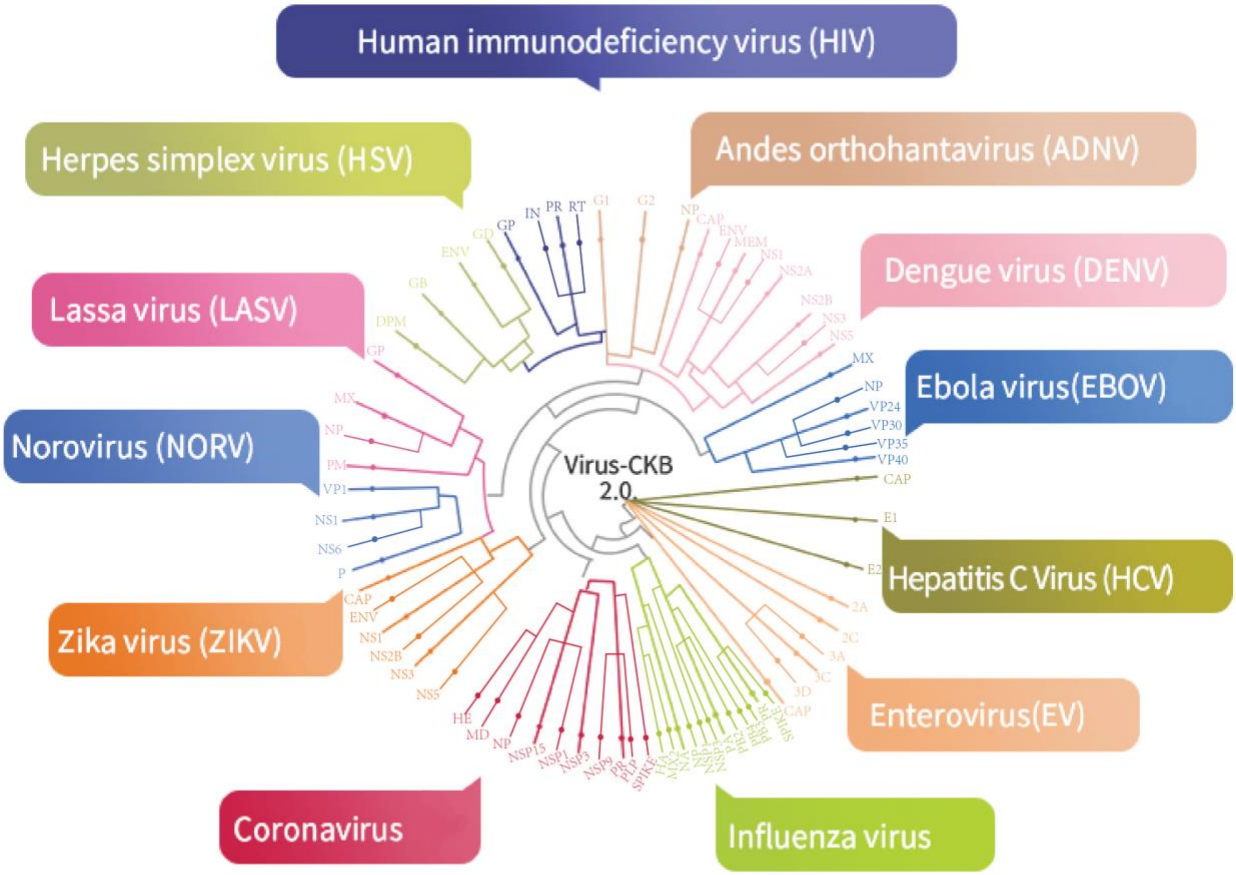


Figure 3 Spider Plot of viral-related targets of the Virus-CKB 2.0

Viruses and their target protein names are listed above.

3.2 Virus-CKB 2.0 Web Interface

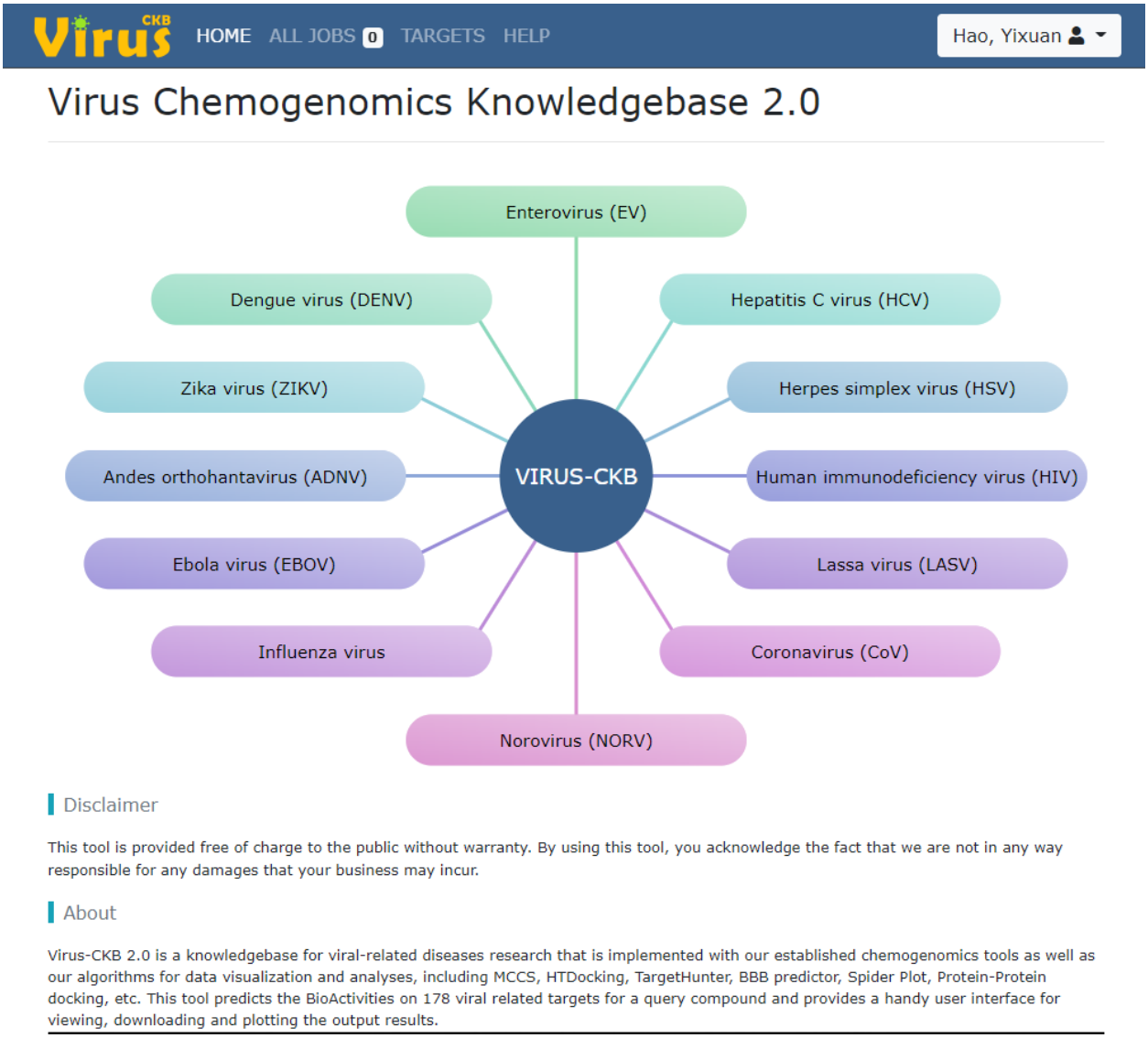


Figure 4 The Virus-CKB 2.0 home page

The Spider Plot of all virus species in Virus-CKB 2.0. Click on one of the viruses to display all the targets of the specific virus in the Virus-CKB 2.0. A brief introduction to the function of Virus-CKB 2.0 and its underlying technology and architecture.

The menu bar contains six buttons, including the icon of Virus-CKB 2.0, HOME, ALL JOBS, TARGETS, HELP and Sign in. For the convenience of users, the menu bar will always stay

at the top of the web page. Clicking on the icon of Virus-CKB 2.0 or HOME button can lead the user back to the home page (Figure 4) of Virus-CKB 2.0, which will display the spider plot of all virus species in Virus-CKB 2.0 and the introduction to the function and underlying technology of Virus-CKB 2.0. Clicking on the ALL JOBS button will bring the user to the job listing page (Figure 5), where the user can search for all jobs submitted to our platform by typing the job name or the molecule name, except for the private jobs. Clicking on the Create a new job button will bring the user to the job creation page (Figure 1), where the user can define the name of the job, the name of the molecules, and the structures of the molecules. Clicking on the TARGETS button will lead the user to the viral-related targets listing page, where the user can see all the viral-related targets archived in Virus-CKB 2.0. The user can also search the protein targets of a specific virus through the search function by typing the keywords such as “Dengue” or “Zika”. Another way to bring the user to a page that displays all the proteins targets of a specific virus is to click on the name of a specific virus on the spider plot located on the home page (Figure 4), clicking on the button of dengue virus will bring the user to the protein targets listing page of dengue virus (Figure 6). Clicking on a specific protein name will bring the user to a popup window that displays the structure, external resources, and different binding pockets of a specific viral protein target. For example, Figure 7 shows the information of non-structural protein 3 of dengue virus type 4. In the structure display area on the left side of the popup window, hold down the left mouse button and move the mouse can rotate the protein structure, hold down the right mouse button and move the mouse can move the protein structure, and rotate the mouse wheel can zoom in or zoom out the protein structure. Clicking on the HELP button will bring the user to a tutorial page (Figure 8), where the user can learn how to create a job and visualize the results step-by-step. On the right of the menu bar is the sign-in button, where the user can sign in through a third-party website. After

signing in, the sign-in button will automatically change to the user's name, clicking on this button and the My jobs button can direct the user to the jobs that the user has created (Figure 9).

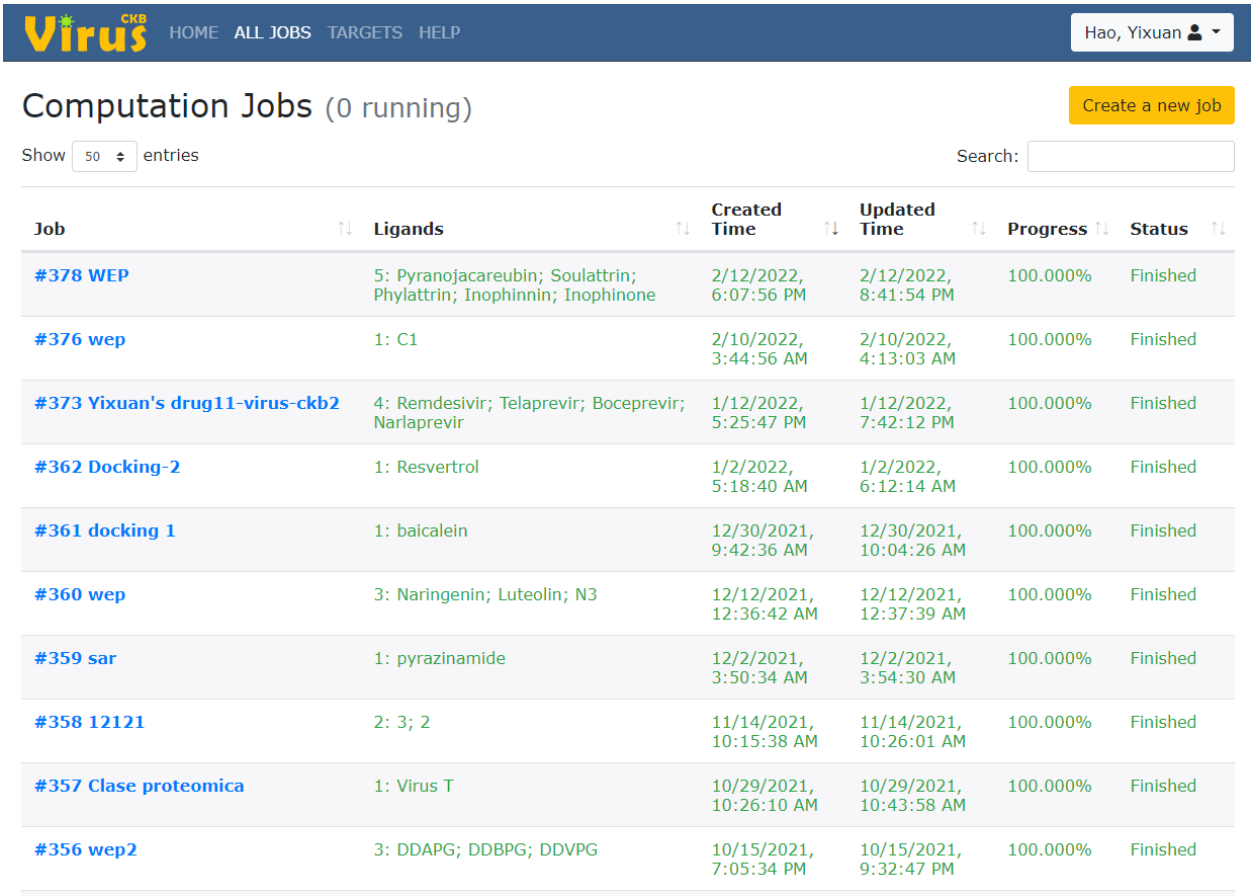


Figure 5 The job listing page

The users can search the job using the search function according to the name of the job or the name of the ligand. If the user selects a private job button when creating a job, unauthorized users cannot search for that job through the search function. The number displayed before the colon on the ligands column is the number of the ligands in that job.

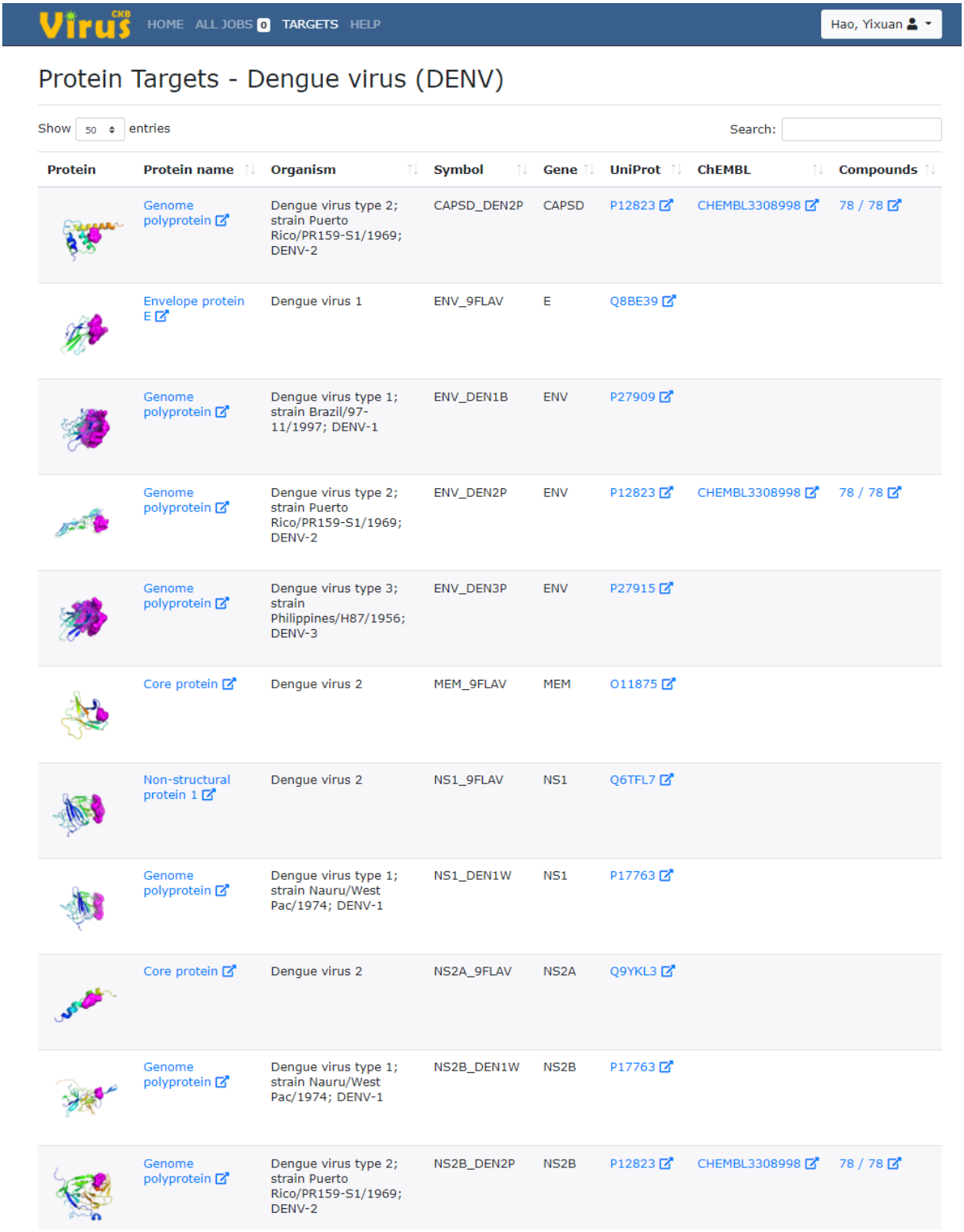


Figure 6 The Protein Targets of Dengue virus (DENV)

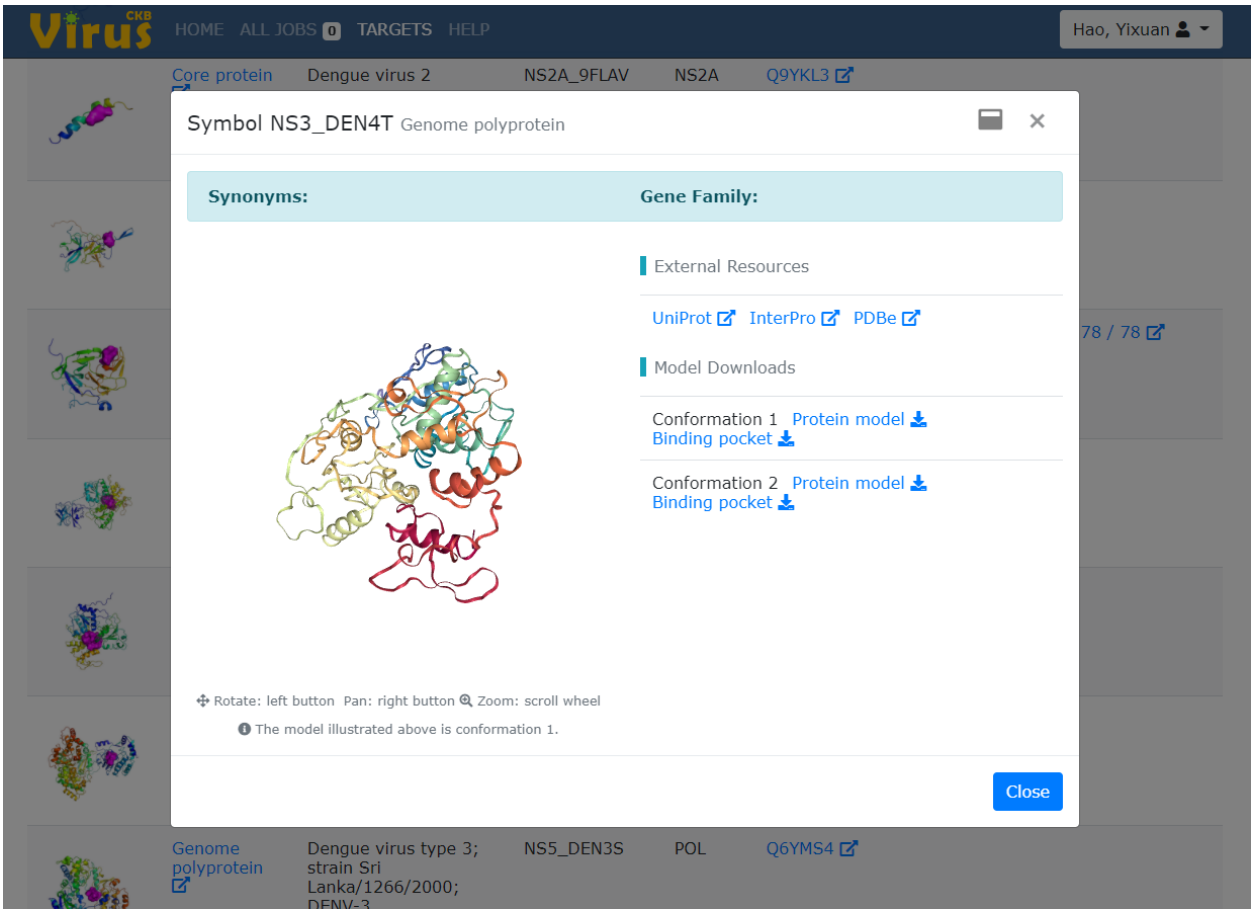


Figure 7 Information of a selected target.

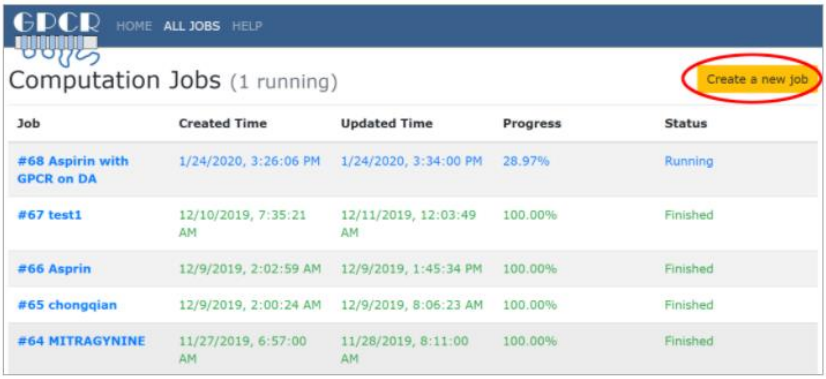
Popup window displays the structure, external resources, and different binding pockets of non-structural protein 3 of dengue virus type 4.

Virus CKB HOME ALL JOBS 0 TARGETS HELP Hao, Yixuan

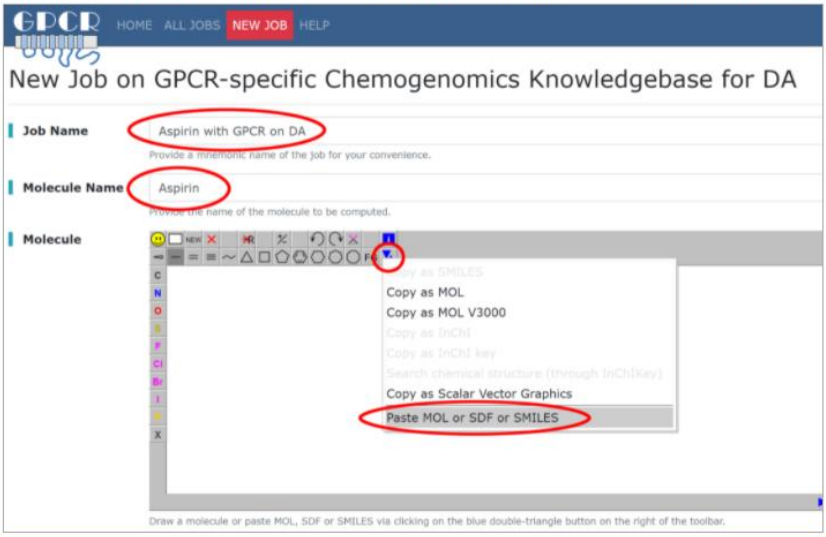
User Guide

How to create a computation job for Virus Chemogenomics Knowledgebase 2.0?

1. To issue a new computation job, you first need to click on the yellow Create a new job button from the ALL JOBS page. You can find the link to **ALL JOBS** wherever you are on this web site.



2. Name your job and draw your molecule with the Molecular Editor. You can also paste in any pre-existing molecule structure in either *MOL*, *SDF* or *SMILES* format via the dropdown menu of the blue double-triangle icon on the toolbar. In this way, please be aware that the pasted code should **NOT** contain any leading or trailing spaces. Hit **Accept** to continue.



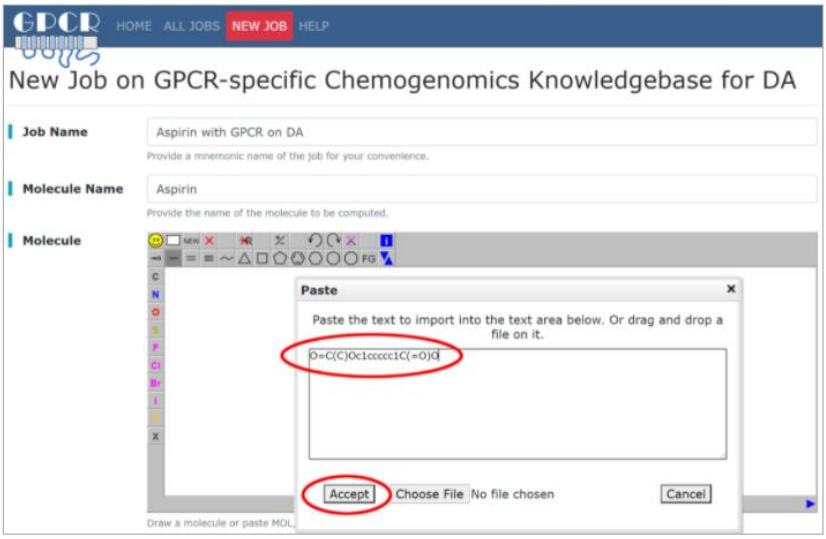


Figure 8 User Guide page

Figure 9 My Computation Jobs page

3.3 Job Output

After creating the job, the platform will be displayed on the job listing page (Figure 5). If the user wants to view the details of the job, the user can click on the job. After clicking on a

specific job, four buttons will be added in the menu bar between the ALL JOBS button and the TARGETS button, including JOB, BBB, OUTPUT, and SPIDER PLOT. Clicking on the BBB button will bring the user to the Blood-Brain Barrier Predictor page (Figure 10), which displays the two kinds of BBB algorithm predictions including the AdaBoost algorithm prediction and the SVM algorithm prediction with in total eight algorithm-fingerprint combinations for each query compound[83]. Clicking on the OUTPUT button will bring the user to the output page (Figure 11), which displays the structure of viral protein targets and their binding pockets, the submitted ligand name, the viral protein target name, the docking scores calculated by HTDocking between the input compounds and the protein target conformations, similarity scores calculated by TargetHunter represent the similarity between the input compounds and the most similar active ChEMBL compound, and the best match compound's ChEMBL ID. Clicking on the ChEMBL ID can bring the user to a comparison page (Figure 12) that provides a detailed comparison between the structural information of the input compound and the most similar compound. The two sets of numbers at the bottom of the comparison page are generated by the algorithm implemented in our platform to characterize the structure of the input compound and the most similar compound in the database. Numbers marked in red indicate structural differences between the two compounds, so the more numbers marked in red, the lower the structural similarity between the two compounds. The output page has two display modes: graphic-based mode and text-based mode. The user can switch the display mode by clicking on the  button in the upper right corner. For a specific job, clicking on the SPIDER PLOT will bring the user to the spider plot page (Figure 13), which displays the predicted interactions network between the input compounds and the protein targets in Virus-CKB 2.0. Both the text-based output page and the spider plot page have a "Gene/Protein"

switch button in the upper right corner. Clicking on this button can enable the switch between gene name and protein name displayed on the text-based output page and spider plot page.

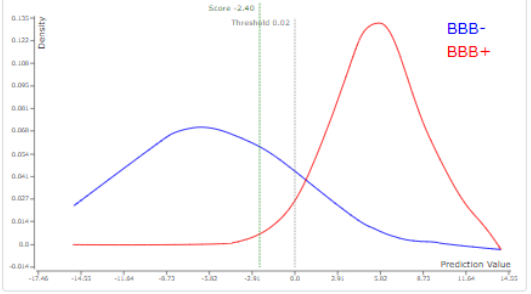
As we can see from the spider plot generated by our platform (Figure 13), we submitted four antiviral drugs in one job, including Telaprevir, Boceprevir, Narlaprevir, and Remdesivir. The structure and name of the query compounds defined by the user are placed in a rectangular box. The discs around the query compound represent the predicted viral protein targets with which the query compound has the potential to bind. The number on the dotted line connecting the predicted viral protein targets and the query compound represents its average docking score. The user can set a critical point, docking scores above this critical point will not be displayed. If the query compound has been verified by bioassay to bind with the predicted viral protein target and can produce corresponding therapeutic effects or the query compound has a highly similar compound (similarity score >= 0.95) that has been proved to bind with the predicted viral protein target and can produce corresponding therapeutic effects, the predicted viral protein target will be marked as green discs, while the other predicted targets will be marked as purple discs. For example, our results indicated that the antiviral drug Telaprevir can bind to non-structural protein 3 protease (Discs: NS3_HCV1B) of the Hepatitis C virus, which is consistent with a bioassay that verifies Telaprevir can be used to treat HCV infection[94].

Yixuan's drug11-virus-ckb2 #373 Blood-Brain Barrier Permeation Prediction

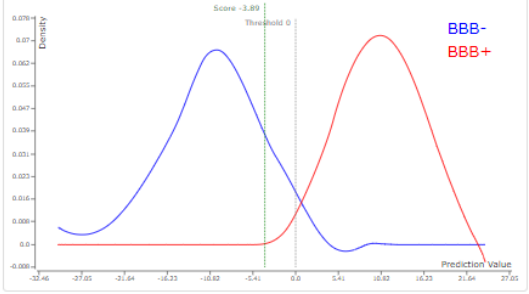
The job was finished at 1/12/2022, 7:42:12 PM.
This BBB permeation predictor is specially designed to classify whether a compound can cross the blood–brain barrier (BBB+) or not (BBB−).
For more information on BBB and the prediction algorithm, please refer to Wikipedia and our paper.

Ligand 1 Remdesivir - AdaBoost Algorithm

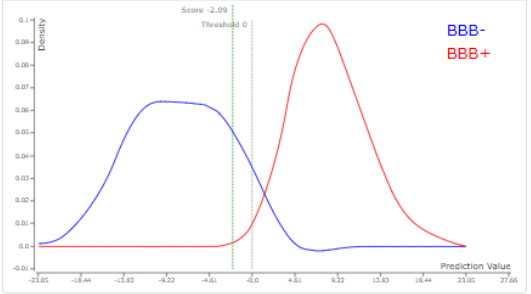
With **MACCS** Fingerprint | Score: **-2.40**



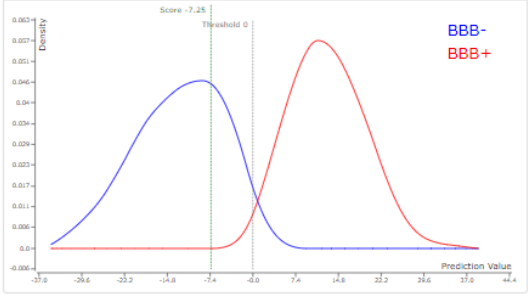
With **PubChem** Fingerprint | Score: **-3.89**



With **Molprint 2D** Fingerprint | Score: **-2.09**

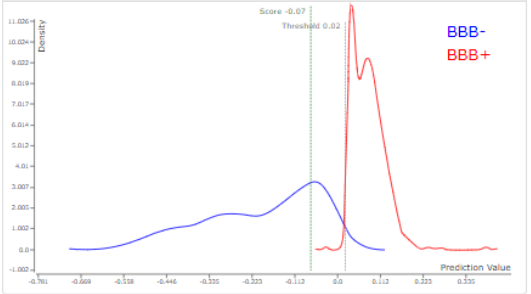


With **FP2** Fingerprint | Score: **-7.25**

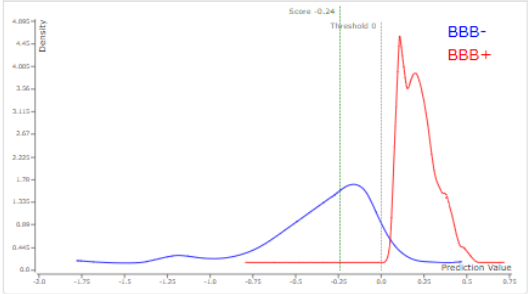


Ligand 1 Remdesivir - SVM Algorithm

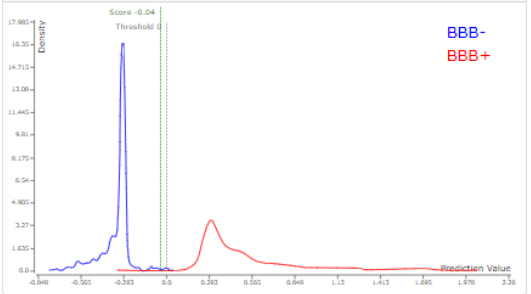
With **MACCS** Fingerprint | Score: **-0.07**



With **PubChem** Fingerprint | Score: **-0.24**



With **Molprint 2D** Fingerprint | Score: **-0.04**



With **FP2** Fingerprint | Score: **-0.13**

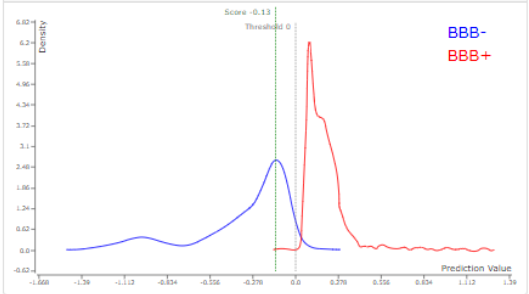


Figure 10 Blood-Brain Barrier Predictor page

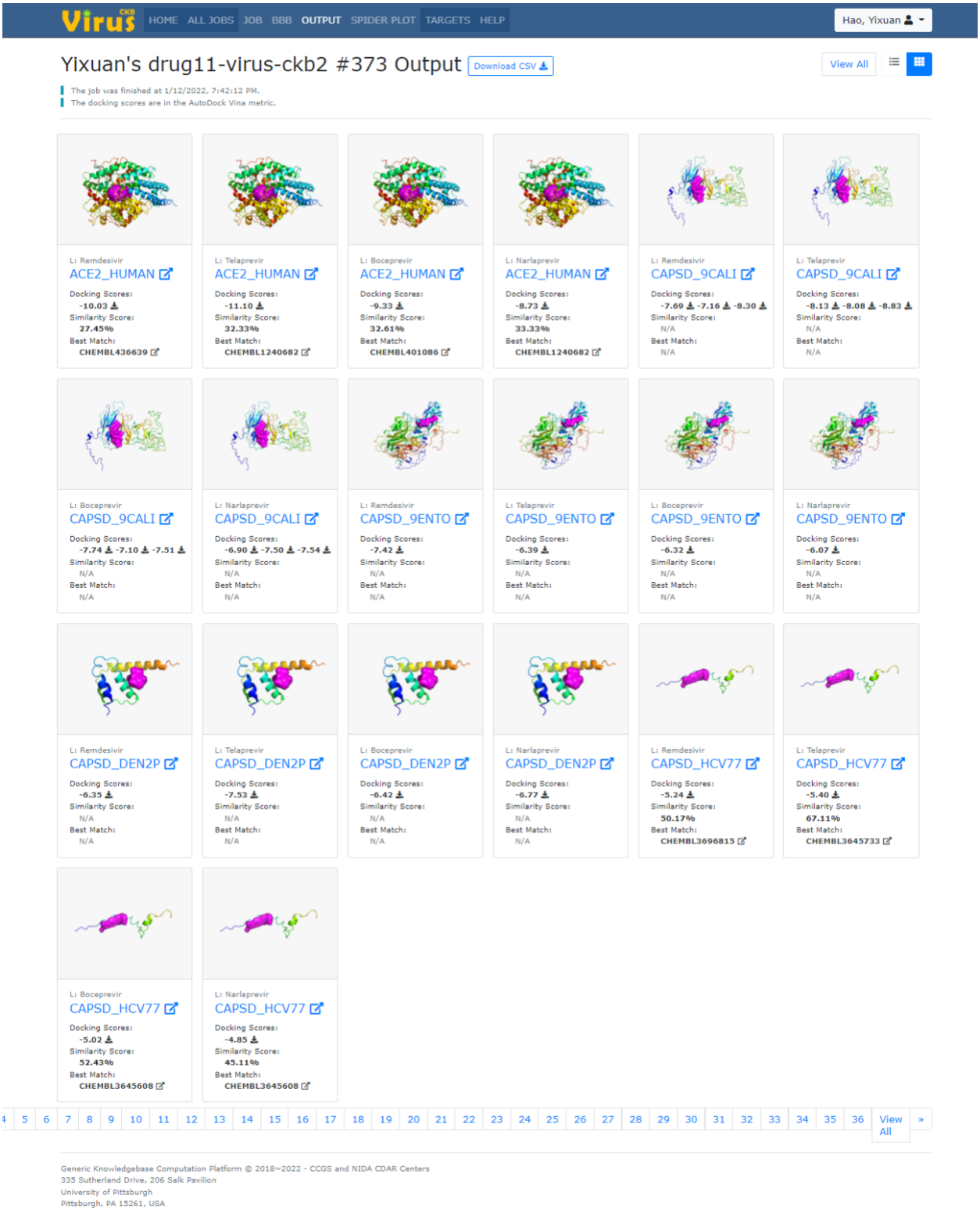
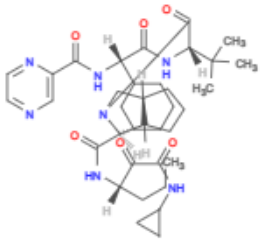


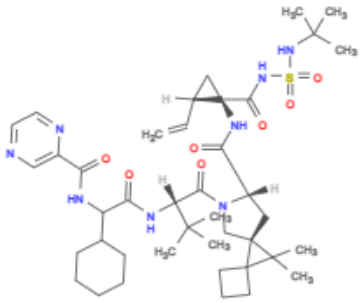
Figure 11 Output page

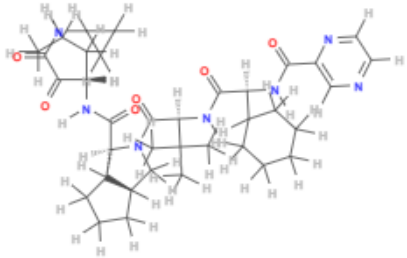
On Target CAPSD_HCV77 0.67105263/CHEMBL3645733

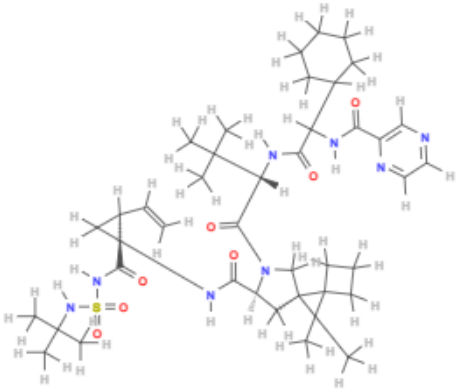
For the given ligand **Telaprevir**, we have found in our database that, being scored **0.67105263**, the most similar ligand is **CHEMBL3645733**. Check out the elaboration below.

Telaprevir (Given Ligand)	CHEMBL3645733 (Similar Ligand)
CCC[C@H](NC(=O)[C@@H]2[C@H]1CCC[C@H]1CN2C(=O)[C@@H](NC(=O)[C@@H](NC(=O)c3cnccn3)C4CCCCC4)C(C)(C)C)C(=O)C(=O)NC5CC5	C=C[C@@H]1C[C@]1(NC(=O)[C@@H]1C[C@@]2(CN1C(=O)[C@@H](NC(=O)C(NC(=O)c1cnccn1)C1CCCCC1)C(C)(C)C)C(C)(C)C21CCC1)C(=O)NS(=O)(=O)NC(C)(C)C









10018002 01803400 00100800 14000c10 00000040	100180a2 09803000 06100800 15200c10 00000040
00100010 0000c000 20003009 14000000 000101a0	00100010 000ac000 20003409 10000000 0011c188
01290000 40013000 0e00801a 01102001 80820400	01692000 40013000 0601801a 01102011 a0c20400
00040104 08020000 02000003 02002010 00000201	00040104 0c024000 02000002 46012000 00004201
80000000 300f8e30 21004003 02100018 00141000	80000000 30078e30 21004003 32900038 00140000
00000000 08000000 90000004 08000004 01060811	00000000 08000000 90000014 08000004 c0060881
a0024000 80002200	e4024004 80012260

Close

Figure 12 Comparison page

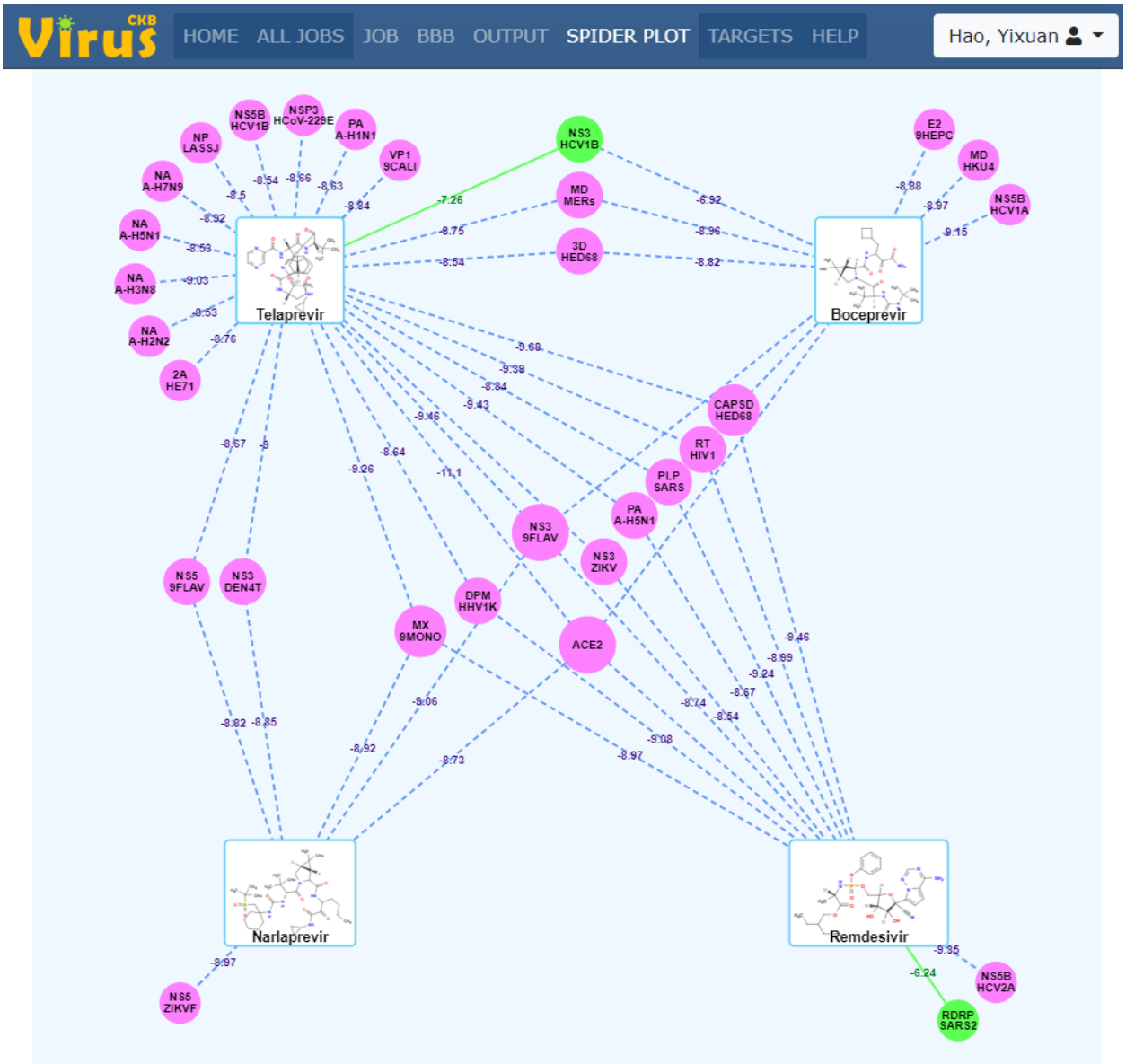


Figure 13 Spider Plot for the analysis of four antiviral drugs, including Telaprevir, Boceprevir, Narlaprevir, and Remdesivir.

Potentially active binding sites (viral protein targets) for the query compounds are shown as purple discs, and the binding sites that have been validated by bioassay are shown as green discs. The average docking scores are marked on the line connecting the query compounds to the viral protein targets.

3.4 Drug Repurposing for Dengue Virus Infection

To further validate our platform, we predicted several FDA-approved non-viral drugs that can bind with the dengue virus and may have the potential to produce therapeutic effects towards dengue virus infections.

According to statistics, antiviral research on the dengue virus mainly focused on targeting structural and non-structural proteins. Among them, the research on structural proteins has mainly focused on the envelope protein because it plays an important role in the process of virus entry into cells. Research on non-structural proteins has mainly focused on NS3 and NS5, because NS5 is the largest and most conserved non-structural protein and has viral RNA-dependent RNA polymerase and methyltransferase activity[95, 96], and NS3 has serine protease activity[97]. We conducted computational prediction for FDA-approved non-viral drugs Ivermectin as shown in Figure 14, our results showed that Ivermectin was predicted to bind with NS3 of Dengue virus (Discs: NS3_DEN4T) and NS3 of Zika virus (Discs: NS3_ZIKV), which are consistent with a study that conformed Ivermectin is a potent inhibitor of flavivirus replication specifically targeting NS3 helicase activity[98]. As shown in Figure 15, we also conducted computational prediction for Policresulen, which is commonly used for the treatment of gynecological infections[99], our prediction results showed that Policresulen was predicted to bind with viral NS3 of Dengue virus (Discs: NS3_DEN4T), which are consistent with a cell assay that conformed Policresulen is a NS2B/NS3 protease inhibitor and can inhibit the replication of dengue virus[100]. All these findings could provide computational prediction-based guidance for repurposing FDA-approved drugs to deal with sudden outbreaks of viral infections. We also recommend that results based on computational predictions need to be validated by biological experiments.

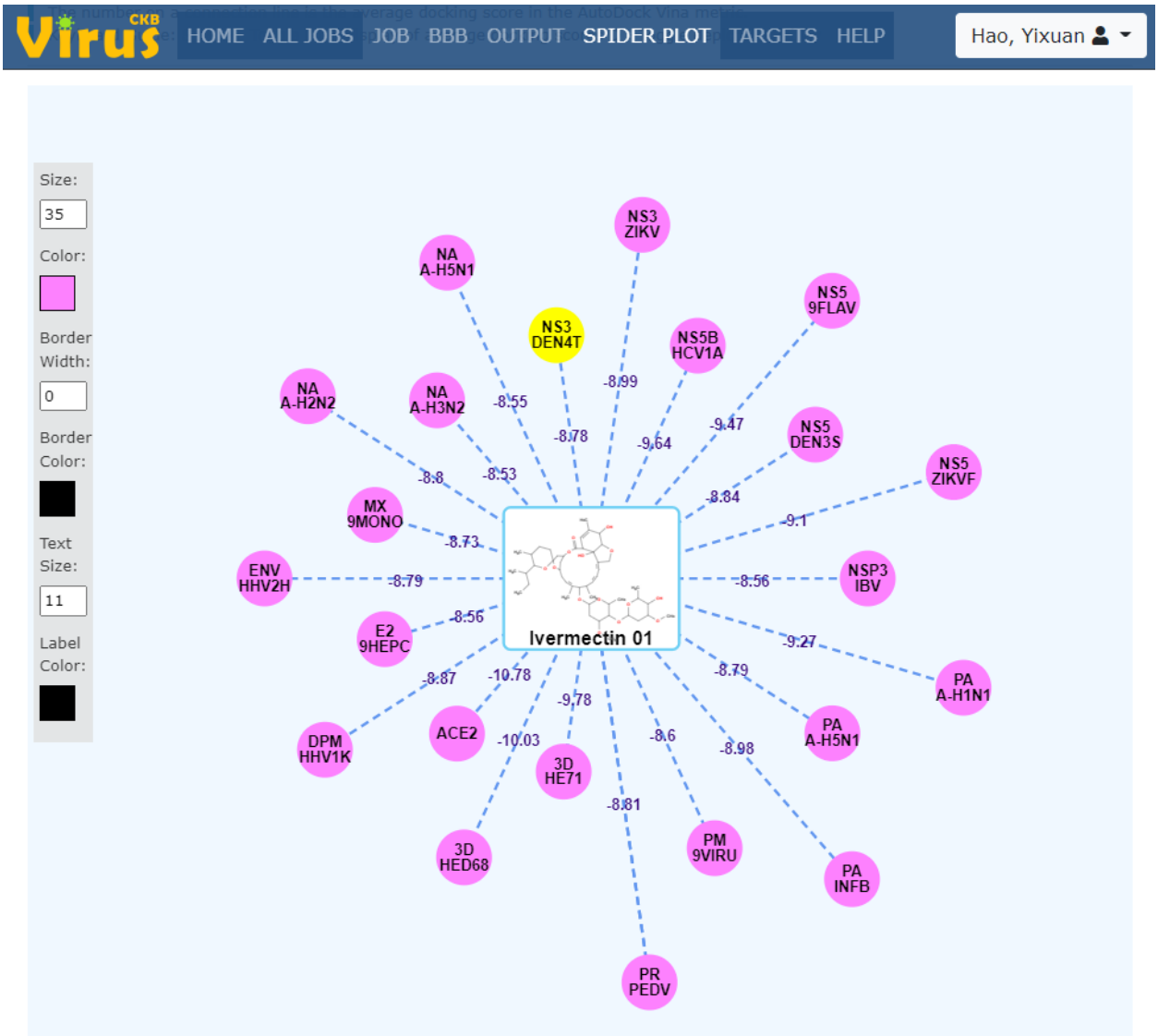


Figure 14 Spider Plot of FDA-approved non-viral drugs Ivermectin.

The prediction results show that Ivermectin can bind with NS3 of the Dengue virus and NS3 of the Zika virus.

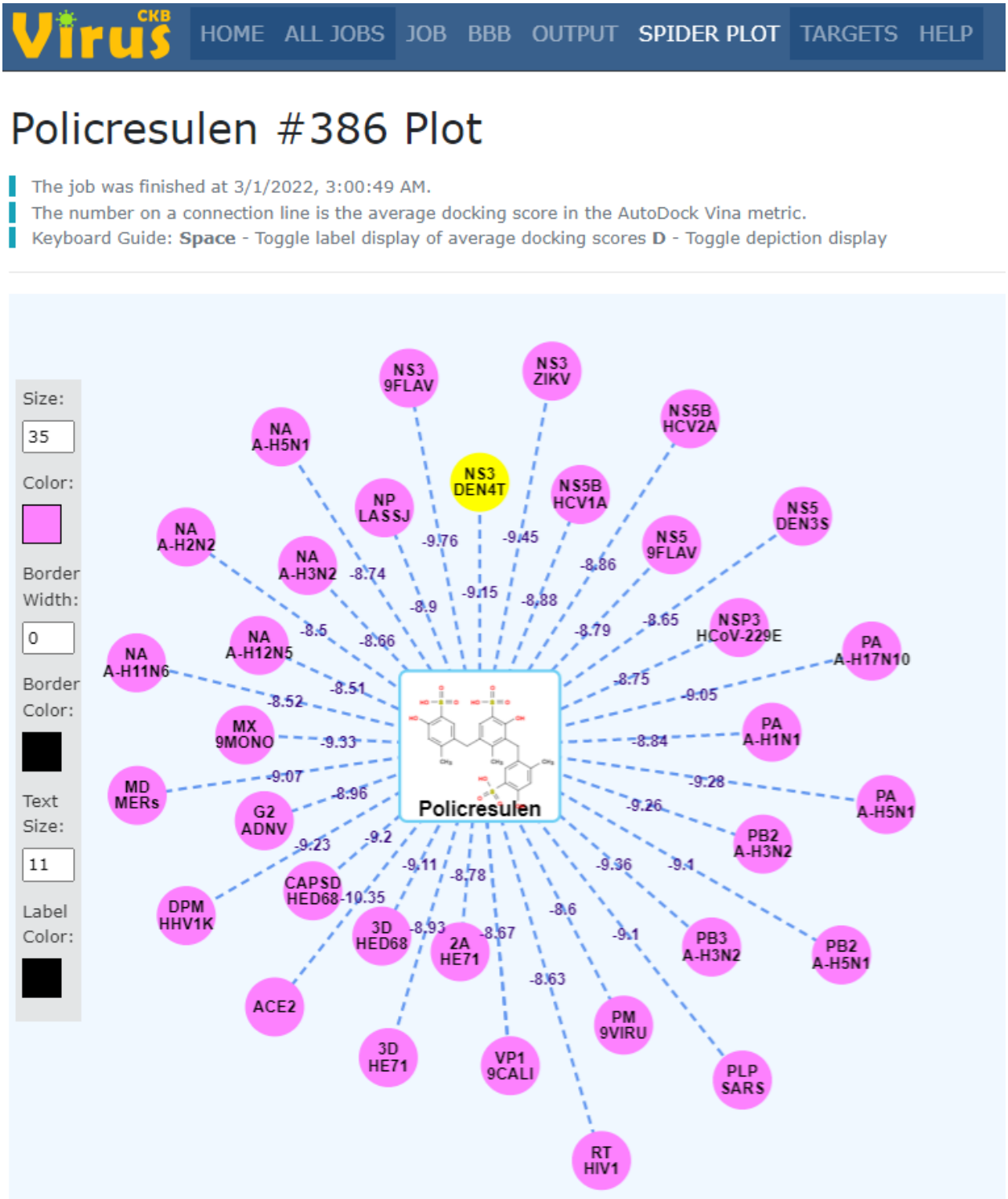


Figure 15 Spider Plot of FDA-approved non-viral drugs Policresulen.

The prediction results show that Policresulen can bind with NS3 of Dengue virus and NS3 of Zika virus.

3.5 Drug Combination for Treatment of Viral-related Disease

Combination therapies are becoming more and more popular in treating viral-related diseases because combination therapies have the advantages of high efficacy and low toxicity and can prevent the development of drug resistance among viruses. In addition to guiding drug repurposing, our platform can also provide computational prediction-based recommendations for drug combinations to treat viral-related diseases. For example, as shown in Figure 16, our prediction results showed that Brequinar (anti-cancer) and Sofosbuvir (anti-HCV) can bind with HIV (Discs: RT_HIV1) and HCV (Discs: NS5B_HCV1A, NS5B_HCV2A), which is consistent with a cell assay that conformed the combination of Sofosbuvir and Brequinar can produce significant therapeutic effect towards HCV infection with low toxicity[101]. The findings could provide computational prediction-based guidance for combination therapies to improve efficacy, lower toxicity, and deal with drug resistance issues of viral-related diseases. We also recommend that results based on computational predictions need to be validated by biological experiments.

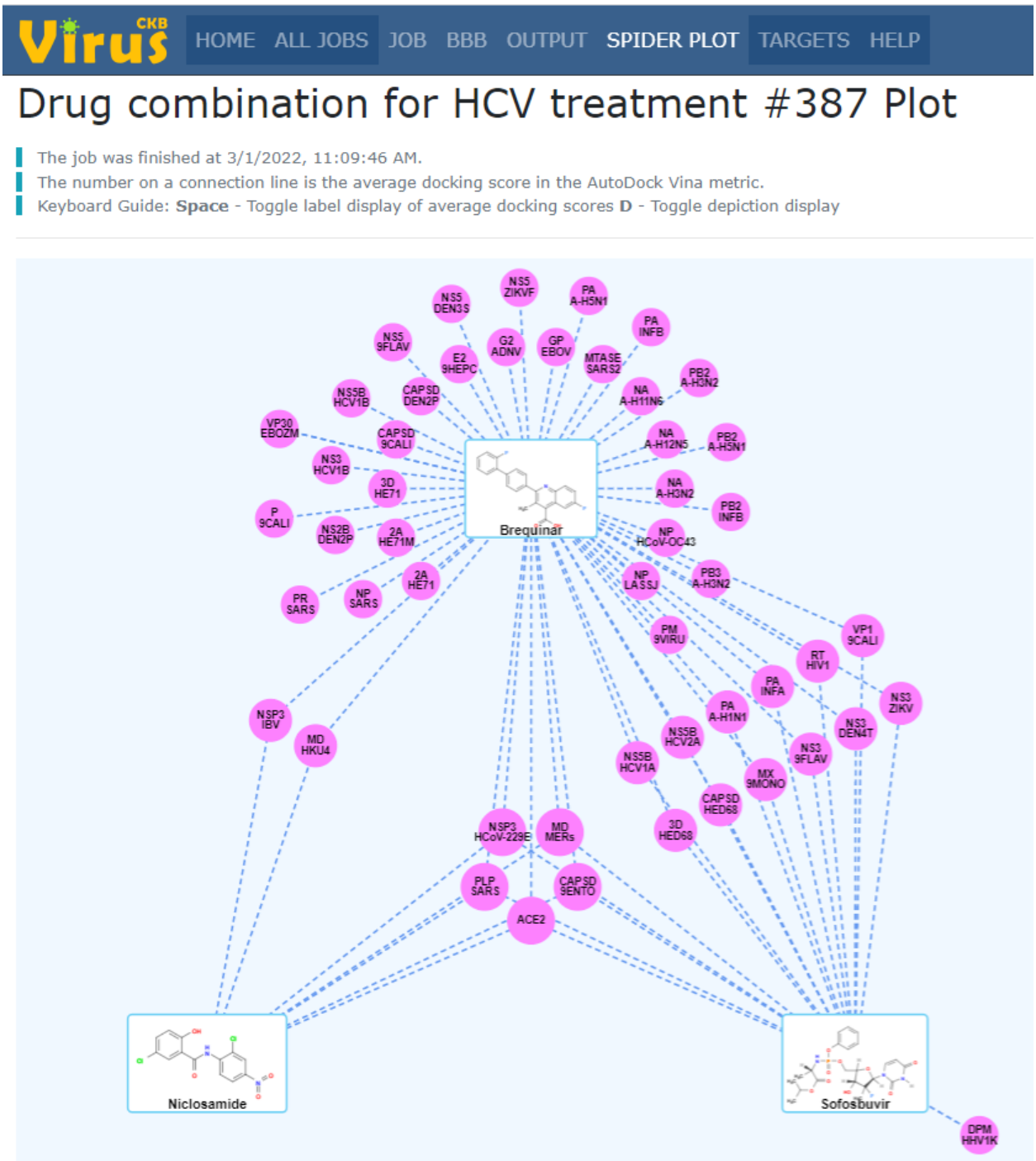


Figure 16 Spider Plot for the analysis of the drug combination, including Brequinar, Sofosbuvir, and Niclosamide.

The prediction results showed that Brequinar and Sofosbuvir can bind with HIV and HCV.

4.0 CONCLUSION

In the present work, we developed a Viral-associated disease-specific chemogenomics knowledgebase, Virus-CKB 2.0, which archived 65 antiviral drugs in the market, 178 viral-related targets with 291 available 3D crystal or cryo-EM structures, and 3766 chemical agents reported for these target proteins. Virus-CKB 2.0 provides the user with a multifunctional platform, which can conduct computational prediction of direct-acting antiviral effects of submitted compounds. Our platform first conducted Molecular Complex Characterizing System (MCCS) protocol, in-silico blood-brain barrier (BBB) prediction, High-throughput docking (HTDocking), and fingerprints-based similarity search simultaneously by our established algorithms implemented in Virus-CKB 2.0. Then the docking scores, similarity scores, and all the other computation results were analyzed and used to generate the Spider Plot for the visualization of the compound-targets interaction network. Finally, the Spider Plot can be used to guide drug repurposing and drug combination.

To validate our platform, we applied Virus-CKB 2.0 to predict the direct-acting antiviral effects of FDA-approved non-viral drugs Ivermectin and Policresulen. The computational prediction results showed that Ivermectin and Policresulen can bind with dengue virus proteins and exhibit significant anti-dengue virus effects, which have been confirmed by relevant bioassays. We also predicted that the drug combination Brequinar and Sofosbuvir can bind with HIV and HCV and exhibit anti-HIV and anti-HCV effects, which have also been confirmed by relevant experimental data. Therefore, our platform can provide very meaningful computationally based recommendations for drug repurposing and drug combination therapy.

In the future, we plan to develop more techniques such as protein-protein docking, which can enable the docking of proteins into virus protein targets and also continue to collect targets information and their protein structures of different viruses and classify proteins structures according to different domains and screen out the protein with the highest resolution to integrate into Virus CKB 2.0 and make Virus-CKB 2.0 a very useful knowledgebase for the research of viral related disease and drug repurposing.

One limitation of our platform is that we can't provide computational prediction of detailed drug-drug interaction so the user can't predict the toxicity of the query compounds because we don't have data on how the drug is metabolized in the body and what kind of enzymes are involved. In order to overcome this limitation, we plan to build an enzyme-related database in the future and develop an algorithm to predict the metabolic mechanism of drugs in vivo and related enzyme species. Another limitation is that when we collect virus structure information, many viruses are undergoing mutation and have produced many diverse variants, so we will also focus on the collection and classification of virus variants in the future.

APPENDIX

Appendix Table 1 Protein Targets of Virus-CKB 2.0

Target name	Organism	Gene	UniProt Ids	ChEMBL ID
2A_HE71	Human enterovirus 71; EV71; EV-71	2A	A9XG43	
2A_HE71M	Human enterovirus 71; strain 7423/MS/87; EV71; EV-71	2A	Q66479	
2C_HE71	Human enterovirus 71; EV71; EV-71	2C	B9VUU3	
3A_9ENTO	Enterovirus F	3A	Q2LKY9	
3A_HE71M	Human enterovirus 71; strain 7423/MS/87; EV71; EV-71	3A	Q66479	
3A_HED68	Human enterovirus D68; EV68; EV-68	3A	A0A2K9Y515	
3C_HE71	Human enterovirus 71; EV71; EV-71	3C	B8YLV9	
3C_HED68	Human enterovirus D68; EV68; EV-68	3C	A1E4A3	
3D_HE71	Human enterovirus 71; EV71; EV-71	3D	I3UIB4	
3D_HED68	Human enterovirus D68; EV68; EV-68	3D	F1T146	
ACE2_HUMAN	Homo sapiens; Human	ACE2	Q9BYF1	CHEMBL3736

CAPSD_9CALI	Feline calicivirus	CAPSD	A2T4Q0	
CAPSD_9ENTO	Enterovirus F	CAPSD	Q2LKZ0	
CAPSD_DEN2P	Dengue virus type 2; strain Puerto Rico/PR159-S1/1969; DENV-2	CAPSD	P12823	CHEMBL3308998
CAPSD_HCV77	Hepatitis C virus genotype 1a; isolate H77; HCV	CAPSD	P27958	CHEMBL3638344
CAPSD_HED68	Human enterovirus D68; EV68; EV-68	VP1	Q9YLJ3	
CAPSD_ZIKV	Zika virus; ZIKV	CAPSD	A0A0X8GJ44	
DPM_HHV1K	Human herpesvirus 1; strain KOS; HHV-1; Human herpes simplex virus 1	UL30	P04292	CHEMBL5944
E1_9HEPC	Hepacivirus C	E1	H9XGD6	
E1_HCV77	Hepatitis C virus genotype 1a; isolate H77; HCV	E1	P27958	CHEMBL3638344
E1_HCVJ1	Hepatitis C virus genotype 1b; isolate HC-J1; HCV	E1	Q03463	
E2_9HEPC	Hepacivirus C	E2	C1KH25	
E2_HCV77	Hepatitis C virus genotype 1a; isolate H77; HCV	E2	P27958	CHEMBL3638344
ENV_9FLAV	Dengue virus 1	E	Q8BE39	

ENV_DEN1B	Dengue virus type 1; strain Brazil/97-11/1997; DENV-1	ENV	P27909	
ENV_DEN2P	Dengue virus type 2; strain Puerto Rico/PR159-S1/1969; DENV-2	ENV	P12823	CHEMBL3308998
ENV_DEN3P	Dengue virus type 3; strain Philippines/H87/1956; DENV-3	ENV	P27915	
ENV_HHV2H	Human herpesvirus 2; strain HG52; HHV-2; Human herpes simplex virus 2	GH	P89445	
ENV_ZIKV	Zika virus; ZIKV	ENV	A0A0X8GJ44	
ENV_ZIKVK	Zika virus; isolate ZIKV/Human/French Polynesia/10087PF/2013; ZIKV	ENV	A0A024B7W1	
G1_ADNV	Andes orthohantavirus	G1	Q99BV0	
G2_ADNV	Andes orthohantavirus	M	Q9E006	
GB_HHV11	Human herpesvirus 1; strain 17; HHV-1; Human herpes simplex virus 1	GB	P10211	CHEMBL2364696
GB_HHV1K	Human herpesvirus 1; strain KOS; HHV-1; Human herpes simplex virus 1	GB	P06437	
GD_9ALPH	Suid alphaherpesvirus 1	US6	G3G933	
GD_HHV1	Human herpesvirus 1; HHV-1; Human herpes simplex virus 1	GD	Q991M3	

GP160_HIV1	Human immunodeficiency virus 1	ENV	C6G099	
GP_EBOV	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	GP	Q05320	CHEMBL4105829
GP_LASSJ	Lassa virus; strain Mouse/Sierra Leone/Josiah/1976; LASV	GPC	P08669	
HA_A-H10N7	Influenza A virus; strain A/Chicken/Germany/n/1949 H10N7	HA	P12581	
HA_A-H10N8	Influenza A virus; A/Jiangxi/IPB13/2013; H10N8	HA	A0A059T4A1	
HA_A-H13N6	Influenza A virus; strain A/Gull/Maryland/704/1977 H13N6	HA	P13103	
HA_A-H14N6	Influenza A virus; strain A/Mallard/Astrakhan/244/1982 H14N6; Influenza A virus; strain A/Mallard/Gurjev/244/1982 H14N6	HA	P26137	
HA_A-H15N9	Influenza A virus; A/shearWater/Australia/2576/1979; H15N9	HA	L0L3X3	
HA_A-H17N10	Influenza A virus; A/little yellow-shouldered bat/Guatemala/060/2010; H17N10	HA	H6QM93	

HA_A-H18N11		HA		
HA_A-H1N1	Influenza A virus; A/WDK/JX/12416/2005; H1N1	HA	C7C6F1	
HA_A-H2N2	Influenza A virus; strain A/Japan/305/1957 H2N2	HA	P03451	
HA_A-H3N2	Influenza A virus	HA	K7N5L2	CHEMBL2366448
HA_A-H3N8	Influenza A virus; A/eq/Newmarket/93/; H3N8	HAI	Q82847	
HA_A-H5N1	Influenza A virus; A/Vietnam/1194/2004; H5N1	HA	Q6DQ34	
HA_A-H5N3	Influenza A virus; strain A/Duck/Singapore/3/1997 H5N3; Influenza A virus; strain A/Duck/Malaysia/F119-3/1997 H5N3	HA	A5Z226	
HA_A-H6N1	H6N1 subtype	HA	A0A0J9X268	
HA_A-H6N6	Influenza A virus; A/chicken/Guangdong/S1414/2010; H6N6	HA	A0A067YZ73	
HA_A-H7N2	Influenza A virus; A/chicken/New Jersey/Sg-00421/2004; H7N2	HA	B7NY59	
HA_A-H7N7	Influenza A virus; A/Netherlands/219/2003; H7N7	HA	Q6VMK1	

HA_A-H7N9	Influenza A virus; A/Hangzhou/1/2013; H7N9	HA	M4YV75	
HA_INFB	Influenza B virus; strain B/Hong Kong/8/1973	HA	P03462	
HA_INFD	Influenza D virus; D/swine/Oklahoma/1334/2011	HEF	K9LG83	
HE_BCV	Bovine coronavirus; strain Mebus; BCoV; BCV	HE	P15776	
HE_HCoV-OC43	Human coronavirus OC43; HCoV-OC43	HE	Q4VID6	
HE_MHV	Murine coronavirus; strain DVIM; MHV-DVIM; Murine hepatitis virus	HE	O92367	
IN_HIV1	Human immunodeficiency virus type 1 group M subtype B; isolate NY5; HIV-1	GAG-POL	P12497	
MD_HKU4	Bat coronavirus HKU4; BtCoV; BtCoV/HKU4/2004	REP	P0C6W3	
MD_MERs	Middle East respiratory syndrome-related coronavirus; isolate United Kingdom/H123990006/2012; MERS-CoV; Betacoronavirus England 1	1A	K9N638	
MEM_9FLAV	Dengue virus 2	MEM	O11875	
MTASE_SARS2	Severe acute respiratory	MTASE		

	syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2			
MTASE_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	MTASE		
MX2_A-H5N1	Influenza A virus; strain A/Goose/Guangdong/1/1996 H5N1 genotype Gs/Gd	M	Q9Q0L9	
MX_9MONO	Sudan ebolavirus	VP40	B0LPL6	
MX_LASSJ	Lassa virus; strain Mouse/Sierra Leone/Josiah/1976; LASV	Z	O73557	
NA_A-H11N6	Influenza A virus; strain A/Duck/England/1/1956 H11N6	NA	Q6XV27	
NA_A-H11N9	Influenza A virus; strain A/Tern/Australia/G70C/1975 H11N9	NA	P03472	
NA_A-H12N5	Influenza A virus; strain A/Duck/Alberta/60/1976 H12N5	NA	A1ILL9	
NA_A-H13N9	Influenza A virus; strain A/Whale/Maine/1/1984 H13N9	NA	P05803	
NA_A-H17N10	Influenza A virus; A/little yellow-shouldered bat/Guatemala/164/2009; H17N10	NA	H6QM85	

NA_A-H18N11		NA		
NA_A-H1N1	Influenza A virus; A/Texas/17/2009; H1N1	NA	C6KP13	
NA_A-H2N2	Influenza A virus; strain A/Tokyo/3/1967 H2N2	NA	P06820	
NA_A-H3N2	Mus musculus; Mouse	NA	P84751	
NA_A-H3N8	Influenza A virus; strain A/Duck/Ukraine/1/1963 H3N8	NA	Q07599	CHEMBL4295592
NA_A-H5N1	Influenza A virus; A/American green-winged teal/Washington/195750/2014; H5N1	NA	A0A0C5BL75	
NA_A-H7N9	Influenza A virus; A/Shanghai/02/2013; H7N9	NA	R4NFR6	
NA_INFB	Influenza B virus; strain B/Lee/1940	NA	P03474	CHEMBL3377
NP_9ADNV	Andes orthohantavirus	NP	Q80DP9	
NP_A-H1N1	Influenza A virus; strain A/Wilson-Smith/1933 H1N1; Influenza A virus; strain A/WS/1933 H1N1	NP	Q1K9H2	
NP_A-H5N1	Influenza A virus; A/Chicken/Hong Kong/786/97; H5N1	NP	Q9PX50	

NP_EBOZM	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	NP	P18272	
NP_HCoV-NL63	Human coronavirus NL63; HCoV-NL63	N	Q6Q1R8	
NP_HCoV-OC43	Human coronavirus OC43; HCoV-OC43	N	P33469	CHEMBL3232681
NP_INFB	Influenza B virus; B/Managua/4577.01/2008	NP	C4LQ26	
NP_INFD	Influenza D virus; D/bovine/France/2986/2012	NP	A0A0E3VZU8	
NP_LASSJ	Lassa virus; strain Mouse/Sierra Leone/Josiah/1976; LASV	N	P13699	
NP_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	N	P59595	
NS1_9CALI	Murine norovirus 1	NS1	Q80J95	
NS1_9FLAV	Dengue virus 2	NS1	Q6TFL7	
NS1_DEN1W	Dengue virus type 1; strain Nauru/West Pac/1974; DENV-1	NS1	P17763	
NS1_ZIKV	Zika virus; ZIKV	NS1	A0A109PRQ3	

NS2A_9FLAV	Dengue virus 2	NS2A	Q9YKL3	
NS2B_DEN1W	Dengue virus type 1; strain Nauru/West Pac/1974; DENV-1	NS2B	P17763	
NS2B_DEN2P	Dengue virus type 2; strain Puerto Rico/PR159-S1/1969; DENV-2	NS2B	P12823	CHEMBL3308998
NS2B_ZIKV	Zika virus; ZIKV	NS2B	Q32ZE1	CHEMBL4523307
NS2B_ZIKVF	Zika virus; isolate ZIKV/Human/French Polynesia/10087PF/2013; ZIKV	NS2B	A0A024B7W1	
NS3_9FLAV	Dengue virus 4	NS3	F8TEL4	
NS3_DEN4T	Dengue virus type 4; strain Thailand/0348/1991; DENV-4	NS3	Q2YHF0	
NS3_HCV1A	Hepatitis C virus genotype 1a; isolate H77; HCV	NS3	P27958	CHEMBL3638344
NS3_HCV1B	Hepacivirus C	NS3	Q91RS4	CHEMBL3988605
NS3_HCV3A	Hepacivirus C	NS3	A0A0B4WYC6	
NS3_ZIKV	Zika virus; ZIKV	NS3	Q32ZE1	CHEMBL4523307

NS5B_HCV1A	Hepatitis C virus genotype 1a; isolate 1; HCV	NS5B	P26664	CHEMBL4620
NS5B_HCV1B	Hepatitis C virus genotype 1b; isolate BK; HCV	NS5B	P26663	CHEMBL6040
NS5B_HCV2A	Hepatitis C virus genotype 2a; isolate JFH-1; HCV	NS5B	Q99IB8	CHEMBL4295932
NS5_9FLAV	Dengue virus 2	NS5	H9M652	
NS5_DEN3S	Dengue virus type 3; strain Sri Lanka/1266/2000; DENV-3	POL	Q6YMS4	
NS5_ZIKV	Zika virus; ZIKV	NS5	Q32ZE1	CHEMBL4523307
NS5_ZIKVF	Zika virus; isolate ZIKV/Human/French Polynesia/10087PF/2013; ZIKV	NS5	A0A024B7W1	
NS6_9CALI	Murine norovirus 1	NS6	Q80J95	
NSP15_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	REP	P0C6X7	CHEMBL5118
NSP15_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	NSP15		
NSP1_INFB	Influenza B virus; strain B/Lee/1940	NS	P03502	

NSP1_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	REP	P0C6X7	CHEMBL5118
NSP3_HCoV-229E	Human coronavirus 229E; HCoV-229E	1A	P0C6U2	
NSP3_IBV	Avian infectious bronchitis virus; strain M41; IBV	1A	P0C6V5	
NSP3_MHV	Murine coronavirus; strain A59; MHV-A59; Murine hepatitis virus	REP	P0C6X9	
NSP3_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	NSP3		
NSP9_HCoV-229E	Human coronavirus 229E; HCoV-229E	REP	P0C6X1	
PA_A-H17N10	Influenza A virus; A/little yellow-shouldered bat/Guatemala/060/2010; H17N10	PA	H6QM92	
PA_A-H1N1	Influenza A virus; strain swl A/California/04/2009 H1N1	PA	C3W5S0	
PA_A-H5N1	Influenza A virus; strain A/Goose/Guangdong/1/1996 H5N1 genotype Gs/Gd	PA	Q9Q0U9	
PA_INFA	Influenza A virus; strain A/Puerto Rico/8/1934 H1N1	PA	P03433	CHEMBL1169598
PA_INFB	Influenza B virus; B/Memphis/13/2003	PA	Q5V8Z9	

PB2_A-H3N2	Influenza A virus; strain A/Beijing/39/1975 H3N2	PB2	Q30NP1	
PB2_A-H5N1	Influenza A virus; A/duck/Shantou/4610/2003; H5N1	PB2	Q2LG68	
PB2_INFB	Influenza B virus; strain B/Lee/1940	PB2	Q9QLL6	
PB3_A-H3N2	Influenza A virus; strain A/Beijing/39/1975 H3N2	PB2	Q30NP1	
PB4_A-H3N2	Influenza A virus; strain A/Beijing/39/1975 H3N2	PB2	Q30NP1	
PLP_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	1A	P0C6U8	CHEMBL3927
PLP_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	PLP		
PM_9VIRU	Lassa mammarenavirus	L	A0A097F4L1	
PR_FCoV	Feline coronavirus; strain FIPV WSU-79/1146; FCoV	REP	Q98VG9	CHEMBL4295624
PR_HCoV-229E	Human coronavirus 229E; HCoV-229E	1A	P0C6U2	
PR_HCoV-NL63	Human coronavirus NL63; HCoV-NL63	1A	P0C6U6	CHEMBL3232683

PR_HIV1	Human immunodeficiency virus type 1 group M subtype B; isolate ARV2/SF2; HIV-1	GAG-POL	P03369	CHEMBL3638331
PR_HIV2	Human immunodeficiency virus type 2 subtype A; isolate ROD; HIV-2	GAG-POL	P04584	
PR_HKU1	Human coronavirus HKU1; isolate N1; HCoV-HKU1	1A	P0C6U3	
PR_HKU4	Bat coronavirus HKU4; BtCoV; BtCoV/HKU4/2004	REP	P0C6W3	
PR_IBV	Avian infectious bronchitis virus; strain M41; IBV	1A	P0C6V5	
PR_MERs	Middle East respiratory syndrome-related coronavirus; MERS-CoV	ORF1A	V9TU12	
PR_PEDV	Porcine epidemic diarrhea virus	ORF1A	U6BPB2	
PR_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	REP	P0C6X7	CHEMBL5118
PR_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	PR		
PR_TGEV	Porcine transmissible gastroenteritis coronavirus; strain Purdue; TGEV	1A	P0C6V2	
P_9CALI	Alphatron virus	P	A0A509GV46	

P_NVN68	Norwalk virus; strain GI/Human/United States/Norwalk/1968; Hu/NV/NV/1968/US	ORF2	Q83884	
RDRP_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	RDRP		
RT_HIV1	Human immunodeficiency virus type 1 group M subtype B; isolate HXB2; HIV-1	GAG-POL	P04585	CHEMBL3638360
SPIKE_HCoV-229E	Human coronavirus 229E; HCoV-229E	S	P15423	
SPIKE_HCoV-NL63	Human coronavirus NL63; HCoV-NL63	S	Q6Q1S2	
SPIKE_HCoV-OC43	Human coronavirus OC43; HCoV-OC43	S	Q696P8	
SPIKE_HKU4	Bat coronavirus HKU4; BtCoV; BtCoV/HKU4/2004	S	A3EX94	
SPIKE_IBV	Infectious bronchitis virus	SPIKE	F4MIW6	
SPIKE_MERs	Middle East respiratory syndrome-related coronavirus; isolate United Kingdom/H123990006/2012; MERS-CoV; Betacoronavirus England 1	S	K9N5Q8	
SPIKE_MHV	Murine hepatitis virus	ORF3	Q9J3E7	

SPIKE_PDCoV	Deltacoronavirus PDCoV/USA/Ohio137/2014	S	A0A075E3D7	
SPIKE_SARS	Severe acute respiratory syndrome coronavirus; SARS-CoV	S	P59594	
SPIKE_SARS2	Severe acute respiratory syndrome coronavirus 2; 2019-nCoV; SARS-CoV-2	S	P0DTC2	
VP1_9CALI	Murine norovirus 1	VP1	Q2V8W4	
VP1_NVN68	Norwalk virus; strain GI/Human/United States/Norwalk/1968; Hu/NV/NV/1968/US	ORF2	Q83884	
VP24_EBORE	Reston ebolavirus; strain Philippines-96; REBOV; Reston Ebola virus	VP24	Q91DD5	
VP24_EBOSU	Sudan ebolavirus; strain Human/Uganda/Gulu/2000; SEBOV; Sudan Ebola virus	VP24	Q5XX02	
VP24_EBOZM	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	VP24	Q05322	
VP30_9MONO	Mengla dianlovirus	VP30	A0A3S8UVN5	

VP30_EBOZ5	Zaire ebolavirus; strain Kikwit-95; ZEBOV; Zaire Ebola virus	VP30	Q77DJ5	
VP30_EBOZM	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	VP30	Q05323	
VP35_EBORR	Reston ebolavirus; strain Reston-89; REBOV; Reston Ebola virus	VP35	Q8JPY0	
VP35_EBOZM	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	VP35	Q05127	
VP40_EBOZM	Zaire ebolavirus; strain Mayinga-76; ZEBOV; Zaire Ebola virus	VP40	Q05128	

Bibliography

1. Wu, Y.-C., C.-S. Chen, and Y.-J. Chan, *The outbreak of COVID-19: An overview.* 2020. **83**(3): p. 217-220.
2. Wu, C., et al., *Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods.* Acta Pharm Sin B, 2020. **10**(5): p. 766-788.
3. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.
4. Douangamath, A., et al., *Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease.* Nature Communications, 2020. **11**(1): p. 5047.
5. V'Kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2.* Nat Rev Microbiol, 2021. **19**(3): p. 155-170.
6. Gordon, C.J., et al., *Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency.* Journal of Biological Chemistry, 2020. **295**(20): p. 6785-6797.
7. Hasan, A., et al., *A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin.* Journal of Biomolecular Structure and Dynamics, 2021. **39**(8): p. 3025-3033.
8. Yun, S.-I. and Y.-M. Lee, *Zika virus: An emerging flavivirus.* Journal of Microbiology, 2017. **55**(3): p. 204-219.
9. Atif, M., et al., *Zika virus disease: a current review of the literature.* Infection, 2016. **44**(6): p. 695-705.
10. Sirohi, D. and R.J. Kuhn, *Zika Virus Structure, Maturation, and Receptors.* J Infect Dis, 2017. **216**(suppl_10): p. S935-s944.
11. Sirohi, D., et al., *The 3.8 Å resolution cryo-EM structure of Zika virus.* Science, 2016. **352**(6284): p. 467-70.
12. Lin, Y., et al., *Identification and characterization of Zika virus NS5 RNA-dependent RNA polymerase inhibitors.* International Journal of Antimicrobial Agents, 2019. **54**(4): p. 502-506.
13. Lim, S.-Y., et al., *A direct-acting antiviral drug abrogates viremia in Zika virus–infected rhesus macaques.* Science translational medicine, 2020. **12**(547): p. eaau9135.
14. Beer, B., R. Kurth, and A. Bukreyev, *Characteristics of Filoviridae: Marburg and Ebola viruses.* Die Naturwissenschaften, 1999. **86**(1): p. 8-17.
15. Wan, W., et al., *Structure and assembly of the Ebola virus nucleocapsid.* Nature, 2017. **551**(7680): p. 394-397.
16. Cantoni, D. and J.S. Rossman, *Ebolaviruses: New roles for old proteins.* PLoS Negl Trop Dis, 2018. **12**(5): p. e0006349.
17. Hansen, F., H. Feldmann, and M.A. Jarvis, *Targeting Ebola virus replication through pharmaceutical intervention.* Expert Opinion on Investigational Drugs, 2021. **30**(3): p. 201-226.
18. Morens, D.M. and A.S. Fauci, *Dengue and Hemorrhagic FeverA Potential Threat to Public Health in the United States.* JAMA, 2008. **299**(2): p. 214-216.

19. Solomon, T. and M. Mallewa, *Dengue and Other Emerging Flaviviruses.* Journal of Infection, 2001. **42**(2): p. 104-115.
20. Murugesan, A. and M. Manoharan, *Dengue Virus.* Emerging and Reemerging Viral Pathogens, 2020: p. 281-359.
21. Perera, R. and R.J. Kuhn, *Structural proteomics of dengue virus.* Current Opinion in Microbiology, 2008. **11**(4): p. 369-377.
22. Hidari, K.I.P.J., et al., *Structure and anti-dengue virus activity of sulfated polysaccharide from a marine alga.* Biochemical and Biophysical Research Communications, 2008. **376**(1): p. 91-95.
23. Tapparel, C., et al., *Picornavirus and enterovirus diversity with associated human diseases.* Infection, Genetics and Evolution, 2013. **14**: p. 282-293.
24. Wells, A.I. and C.B. Coyne, *Enteroviruses: A Gut-Wrenching Game of Entry, Detection, and Evasion.* 2019. **11**(5): p. 460.
25. Yuan, J., et al., *Enterovirus A71 proteins: structure and function.* Frontiers in Microbiology, 2018. **9**: p. 286.
26. Bauer, L., et al., *Direct-acting antivirals and host-targeting strategies to combat enterovirus infections.* Current Opinion in Virology, 2017. **24**: p. 1-8.
27. Buontempo, P.J., et al., *SCH 48973: a potent, broad-spectrum, antienterovirus compound.* Antimicrobial agents and chemotherapy, 1997. **41**(6): p. 1220-1225.
28. Safronetz, D., et al., *Detection of Lassa virus, Mali.* Emerging infectious diseases, 2010. **16**(7): p. 1123-1126.
29. Bowen, M.D., et al., *Genetic Diversity among Lassa Virus Strains.* 2000. **74**(15): p. 6992-7004.
30. Brunotte, L., et al., *Domain Structure of Lassa Virus L Protein.* 2011. **85**(1): p. 324-333.
31. Glass, R.I., U.D. Parashar, and M.K. Estes, *Norovirus Gastroenteritis.* 2009. **361**(18): p. 1776-1785.
32. Robilotti, E., S. Deresinski, and B.A. Pinsky, *Norovirus.* 2015. **28**(1): p. 134-164.
33. Jiang, X., et al., *Sequence and Genomic Organization of Norwalk Virus.* Virology, 1993. **195**(1): p. 51-61.
34. Netzler, N.E., D. Enosi Tuipulotu, and P.A. White, *Norovirus antivirals: Where are we now?* Medicinal research reviews, 2019. **39**(3): p. 860-886.
35. Fulhorst, C.F., et al., *CHAPTER 71 - Hantavirus Infections*, in *Tropical Infectious Diseases: Principles, Pathogens and Practice (Third Edition)*, R.L. Guerrant, D.H. Walker, and P.F. Weller, Editors. 2011, W.B. Saunders: Edinburgh. p. 470-480.
36. Heinemann, P., J. Schmidt-Chanasit, and S. Günther, *The N terminus of Andes virus L protein suppresses mRNA and protein expression in mammalian cells.* Journal of virology, 2013. **87**(12): p. 6975-6985.
37. Enright, A.M. and C.G. Prober, *Herpesviridae infections in newborns: varicella zoster virus, herpes simplex virus, and cytomegalovirus.* Pediatric Clinics of North America, 2004. **51**(4): p. 889-908.
38. Whitley, R.J. and B. Roizman, *Herpes simplex virus infections.* The Lancet, 2001. **357**(9267): p. 1513-1518.
39. McElwee, M., et al., *Structure of the herpes simplex virus portal-vertex.* PLOS Biology, 2018. **16**(6): p. e2006191.

40. Copeland, A.M., W.W. Newcomb, and J.C. Brown, *Herpes Simplex Virus Replication: Roles of Viral Proteins and Nucleoporins in Capsid-Nucleus Attachment.* 2009. **83**(4): p. 1660-1668.
41. Lai, W.L., et al., *Inhibition of herpes simplex virus type 1 by thymol-related monoterpenoids.* Planta Med, 2012. **78**(15): p. 1636-8.
42. Ferri, C., et al., *Hepatitis C virus syndrome: A constellation of organ- and non-organ specific autoimmune disorders, B-cell non-Hodgkin's lymphoma, and cancer.* World journal of hepatology, 2015. **7**(3): p. 327-343.
43. Rusyn, I. and S.M. Lemon, *Mechanisms of HCV-induced liver cancer: what did we learn from in vitro and animal studies?* Cancer letters, 2014. **345**(2): p. 210-215.
44. Alter, M.J., *HCV Routes of Transmission: What Goes Around Comes Around.* Semin Liver Dis, 2011. **31**(04): p. 340-346.
45. *Genome of Human Hepatitis C Virus (HCV): Gene Organization, Sequence Diversity, and Variation.* 2000. **5**(3): p. 129-151.
46. Op De Beeck, A. and J. Dubuisson, *Topology of hepatitis C virus envelope glycoproteins.* 2003. **13**(4): p. 233-241.
47. Castelli, M., et al., *A Biologically-validated HCV E1E2 Heterodimer Structural Model.* Scientific reports, 2017. **7**(1): p. 214-214.
48. Bryan-Marrugo, O.L., et al., *History and progress of antiviral drugs: From acyclovir to direct-acting antiviral agents (DAAs) for Hepatitis C.* Medicina Universitaria, 2015. **17**(68): p. 165-174.
49. Ferguson, M.R., et al., *HIV-1 replication cycle.* Clinics in Laboratory Medicine, 2002. **22**(3): p. 611-635.
50. Barré-Sinoussi, F., *HIV as the cause of AIDS.* The Lancet, 1996. **348**(9019): p. 31-35.
51. Weiss, R.A., *How Does HIV Cause AIDS?* 1993. **260**(5112): p. 1273-1279.
52. Shaw, G.M. and E. Hunter, *HIV Transmission.* 2012. **2**(11).
53. Carla Kuiken, T., et al., *HIV Sequence Compendium 2008 Introduction.* 2008.
54. Wilkins, T.J.N.T., *HIV 1: epidemiology, pathophysiology and transmission.* 2020. **116**(7): p. 40-42.
55. Krammer, F., et al., *Influenza.* Nature reviews. Disease primers, 2018. **4**(1): p. 3-3.
56. Ghebrehewet, S., P. MacPherson, and A. Ho, *Influenza.* BMJ (Clinical research ed.), 2016. **355**: p. i6258-i6258.
57. Joseph, U., et al., *The ecology and adaptive evolution of influenza A interspecies transmission.* Influenza and other respiratory viruses, 2017. **11**(1): p. 74-84.
58. Sederdahl, B.K. and J.V. Williams, *Epidemiology and Clinical Characteristics of Influenza C Virus.* Viruses, 2020. **12**(1): p. 89.
59. Feng, Z., et al., *Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research.* Briefings in Bioinformatics, 2020. **22**(2): p. 882-895.
60. Liu, Y., et al., *The development of Coronavirus 3C-Like protease (3CL(pro)) inhibitors from 2010 to 2020.* Eur J Med Chem, 2020. **206**: p. 112711.
61. Littler, D.R., et al., *Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9.* iScience, 2020. **23**(7): p. 101258.
62. Sternberg, A. and C. Naujokat, *Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination.* Life Sci, 2020. **257**: p. 118056.
63. Yee, P.T.I. and C.L. Poh, *Development of Novel Vaccines against Enterovirus-71.* 2016. **8**(1): p. 1.

64. Yi, E.-J., et al., *Enterovirus 71 infection and vaccines.* Clin Exp Vaccine Res, 2017. **6**(1): p. 4-14.
65. Li, S., et al., *Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike.* PLoS Pathog, 2016. **12**(2): p. e1005418.
66. Tang-Huau, T.L., H. Feldmann, and K. Rosenke, *Animal models for Lassa virus infection.* Curr Opin Virol, 2019. **37**: p. 112-117.
67. Graziano, V.R., J. Wei, and C.B. Wilen, *Norovirus Attachment and Entry.* Viruses, 2019. **11**(6).
68. de Graaf, M., J. van Beek, and M.P. Koopmans, *Human norovirus transmission and evolution in a changing world.* Nat Rev Microbiol, 2016. **14**(7): p. 421-33.
69. Kim, J., et al., *Identification of a Direct-Acting Antiviral Agent Targeting RNA Helicase via a Graphene Oxide Nanobiosensor.* 2021. **13**(22): p. 25715-25726.
70. Burley, S.K., et al., *RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy.* Nucleic Acids Res, 2019. **47**(D1): p. D464-d474.
71. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
72. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1100-7.
73. Zhu, F., et al., *Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1128-36.
74. Qin, C., et al., *Therapeutic target database update 2014: a resource for targeted therapeutics.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1118-23.
75. Yang, H., et al., *Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information.* Nucleic Acids Res, 2016. **44**(D1): p. D1069-74.
76. Li, Y.H., et al., *Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics.* Nucleic Acids Res, 2018. **46**(D1): p. D1121-d1127.
77. Wang, Y., et al., *Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics.* Nucleic Acids Research, 2019. **48**(D1): p. D1031-D1041.
78. Feng, Z., et al., *Binding Characterization of GPCRs-Modulator by Molecular Complex Characterizing System (MCCS).* ACS Chemical Neuroscience, 2020. **11**(20): p. 3333-3345.
79. Chen, M., et al., *MCCS, a novel characterization method for protein–ligand complex.* Briefings in Bioinformatics, 2020. **22**(4).
80. Chen, M., et al., *DAKB-GPCRs: An Integrated Computational Platform for Drug Abuse Related GPCRs.* Journal of chemical information and modeling, 2019. **59**(4): p. 1283-1289.
81. Cheng, J., et al., *Computational systems pharmacology-target mapping for fentanyl-laced cocaine overdose.* 2019. **10**(8): p. 3486-3499.
82. Wang, L., et al., *TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database.* The AAPS Journal, 2013. **15**(2): p. 395-406.

83. Liu, H., et al., *AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research.* Journal of Chemical Information and Modeling, 2014. **54**(4): p. 1050-1060.
84. Zhang, Y., et al., *StemCellCKB: An Integrated Stem Cell-Specific Chemogenomics KnowledgeBase for Target Identification and Systems-Pharmacology Research.* Journal of Chemical Information and Modeling, 2016. **56**(10): p. 1995-2004.
85. Zhang, H., et al., *Cardiovascular Disease Chemogenomics Knowledgebase-guided Target Identification and Drug Synergy Mechanism Study of an Herbal Formula.* Scientific Reports, 2016. **6**(1): p. 33963.
86. Gareau, T.P., et al., *Spider plots: a tool for participatory extension learning.* 2010. **48**(5): p. 5TOT8.
87. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* Journal of Cheminformatics, 2011. **3**(1): p. 33.
88. O'Boyle, N.M., et al., *Open babel.* 2013. **3**(1): p. 33-2011.
89. Ertl, P., *Molecular structure input on the web.* Journal of Cheminformatics, 2010. **2**(1): p. 1.
90. Li, H., et al., *istar: A web platform for large-scale protein-ligand docking.* 2014. **9**(1): p. e85678.
91. Rose, A.S., et al., *NGL viewer: web-based molecular graphics for large complexes.* 2018. **34**(21): p. 3755-3758.
92. Ertl, P.J.J.o.c., *Molecular structure input on the web.* 2010. **2**(1): p. 1-9.
93. Wishart, D.S., et al., *DrugBank 5.0: a major update to the DrugBank database for 2018.* Nucleic Acids Research, 2017. **46**(D1): p. D1074-D1082.
94. Zeuzem, S., et al., *Telaprevir for Retreatment of HCV Infection.* 2011. **364**(25): p. 2417-2428.
95. Kaptein, S.J.F. and J. Neyts, *Towards antiviral therapies for treating dengue virus infections.* Current Opinion in Pharmacology, 2016. **30**: p. 1-7.
96. Chan, C.Y. and E.E. Ooi, *Dengue: an update on treatment options.* 2015. **10**(12): p. 2017-2031.
97. Troost, B. and J.M. Smit, *Recent advances in antiviral drug development towards dengue virus.* Current Opinion in Virology, 2020. **43**: p. 9-21.
98. Mastrangelo, E., et al., *Ivermectin is a potent inhibitor of flavivirus replication specifically targeting NS3 helicase activity: new prospects for an old drug.* Journal of Antimicrobial Chemotherapy, 2012. **67**(8): p. 1884-1894.
99. Renner, A., *[Albothyl, a substance with a new mechanism of action in the treatment of gynecological diseases].* Med Klin, 1954. **49**(50): p. 1998-9.
100. Wu, D.-w., et al., *Policresulen, a novel NS2B/NS3 protease inhibitor, effectively inhibits the replication of DENV2 virus in BHK-21 cells.* Acta Pharmacologica Sinica, 2015. **36**(9): p. 1126-1136.
101. Ianevski, A., et al., *Identification and Tracking of Antiviral Drug Combinations.* 2020. **12**(10): p. 1178.