One Target and three drugs for SARS-CoV-2

By

Jeremy Singer

# Abstract

SARS-CoV-2, also known as COVID-19, is a virus that causes flu like symptoms including respiratory distress, in many cases requiring respirators to maintain oxygenation in patients. It is highly contagious, and is currently causing pandemic infection, with a fatality rate estimated between 2% and 3% [1]. Persons over 60 have may have much higher fatality rates [2].

A target repurposing strategy can provide drugs more quickly and cheaply than creating new drugs and finding new targets [3]. This strategy can produce treatments to ameliorate the disease until a vaccine becomes available, or in addition to the vaccine.

Using a paralog search pipeline, the author searched the ChEMBL 25 database, screening it against the SARS-CoV-2 genome and found a high scoring target that has three known drugs [4-6]. The target that was found is an RNA polymerase from the viral genome used by the virus to reproduce.

# Introduction

ChEMBL provides a downloadable database that includes drug targets and drug information for those targets, as well as amino acid sequences of the protein targets [5, 6]. Drug targets tend to be proteins that are important enough to the organism to which they belong that they tend to be conserved [7].

Using **jackhmmer** to provide similarity reports, a pipeline imported scores showing sequence similarity. The target scores were loaded into a PostgreSQL database that also contains the ChEMBL data [8].

The data were analyzed, producing a report showing the target that had high similarity to the SARS-CoV-2 genome, and drugs associated with that target.

# Materials and Methods

The nucleotide genome of SARS-CoV-2 was downloaded as MN908947.3.FASTA [9].

ORFs were translated using EMBOSS tools [10].

This command creates a file containing all the open reading frames (ORFs) found and translates the nucleotide sequences into amino acid sequences:

**getorf MN908947.3.FASTA**

This creates file MN908947.3.orf, which contains all the ORFs found for the .FASTA file.

Commands run in R Studio quantify how many ORFs are contained:

|  |
| --- |
| > aa=read.table(file="mn908947.orf",header = FALSE, sep='~', stringsAsFactors = FALSE)  > aa=aa[!is.na(aa[,1]),] # filter out NA  > aa=data.frame(lines=aa, stringsAsFactors = FALSE)  > orf\_headers=aa[substr(aa[,1],1,1)=='>' ,]  > length(orf\_headers)  [1] 1572 |
|  |
| |  | | --- | |  |   1572 ORFs were found. |

**Jackhmmer** was used to create reports and summaries of similarities with targets [8].

[osboxes@osboxes ~/genomes/MN908947.3 ] jackhmmer --domtblout orf.summary -o orf.hmm.txt mn908947.orf ~/hmmer\_targets/component\_sequences.fa

From psql, the data were imported into the **chembl\_25** database:

[postgres@osboxes /home/osboxes/genomes] **psql -U postgres -d chembl\_25**

psql (9.2.24)

Type "help" for help.

chembl\_25=# **\i import\_hmmer\_statistics.sql**

TRUNCATE TABLE

INSERT 0 49

chembl\_25=# **update hmmer\_statistics set tax\_id=2697049, organism='SARS-CoV-2’ where tax\_id is null;**

49 ORFs had enough similarity to targets to participate in our analysis.

This histogram shows the distribution of scores:

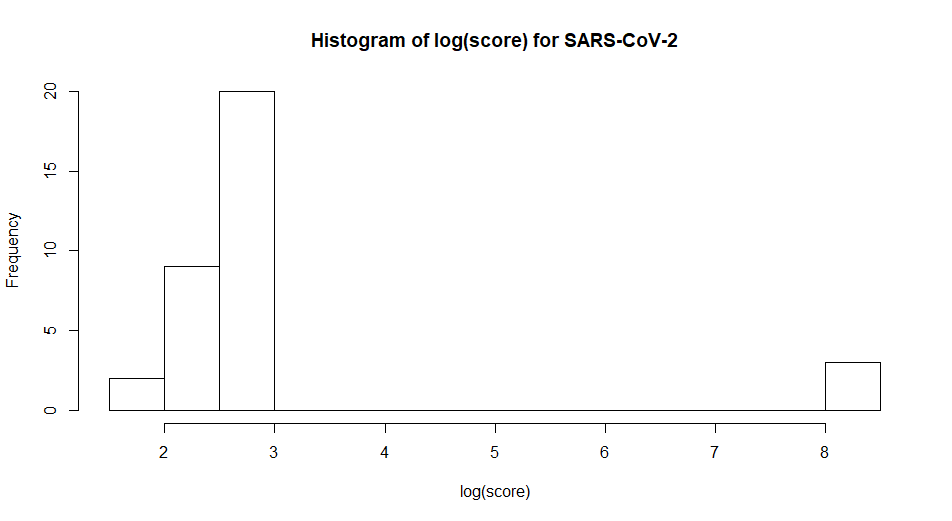


Figure 1: Target similarity score distribution of ORFs from SARS-CoV-2 genome.

Kmeans was used to identify the high scoring threshold.

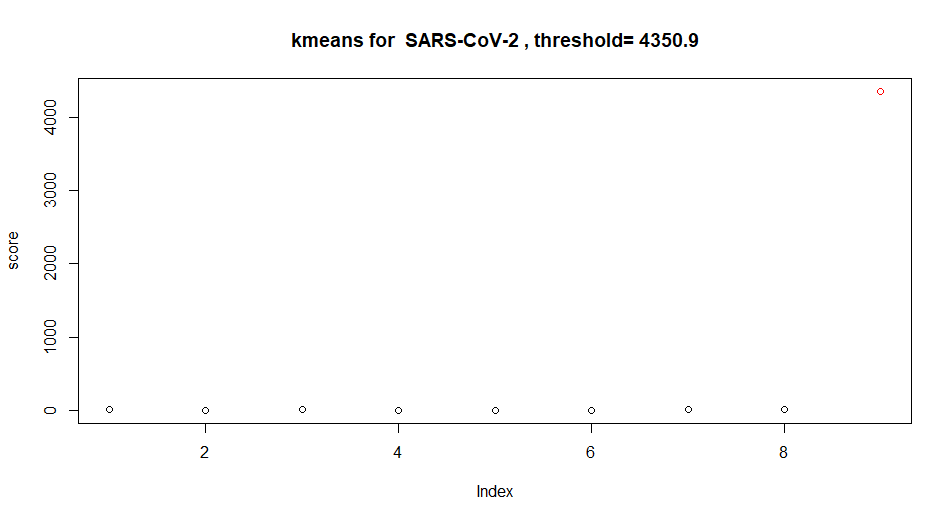


Figure 2: Kmeans showing high scoring group in red.

The threshold was used to select a report from the database identifying the target having high similarity to the SARS-CoV-2 genome, and drugs associated with it.

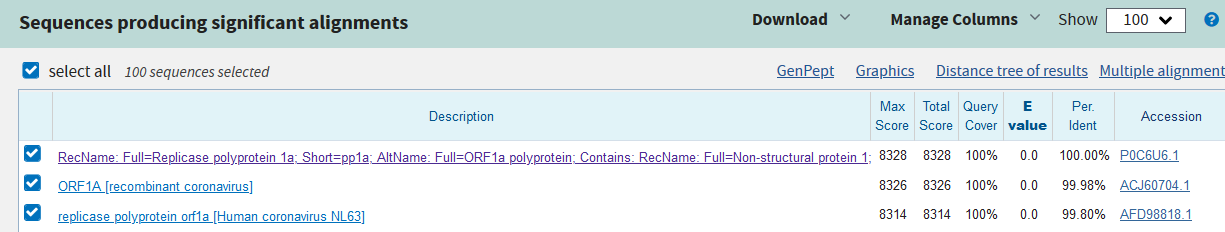
# Results and discussion

One target and three drugs were found.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **score** | **original tax id** | **orig\_organism** | **target\_type** | **target\_name** | **target\_chembl\_id** | **drug\_name** | **drug\_chembl\_id** |
| 4350.9 | 1773 | Mycobacterium tuberculosis | PROTEIN NUCLEIC-ACID COMPLEX | 70S ribosome | CHEMBL2363965 | VIOMYCIN SULFATE | CHEMBL3989823 |
| 4350.9 | 1773 | Mycobacterium tuberculosis | PROTEIN NUCLEIC-ACID COMPLEX | 70S ribosome | CHEMBL2363965 | CAPREOMYCIN SULFATE | CHEMBL2218913 |
| 4350.9 | 1773 | Mycobacterium tuberculosis | PROTEIN NUCLEIC-ACID COMPLEX | 70S ribosome | CHEMBL2363965 | PYRAZINAMIDE | CHEMBL614 |

In the ChEMBL database, a target may have multiple sequences. CHEMBL2363965 has 59 sequences. The high scoring sequence we chose has component\_id 8515.

NCBI BLASTP finds a 100% match:



[11, 12]

Although ChEMBL says that this is a Tuberculosis target, no match was found in the Tuberculosis genome. It may be that as part of treatment of Tuberculosis, a concurrent infection by a Carona virus also was treated by the drugs associated with this target.

The distance tree shows relations of this protein:

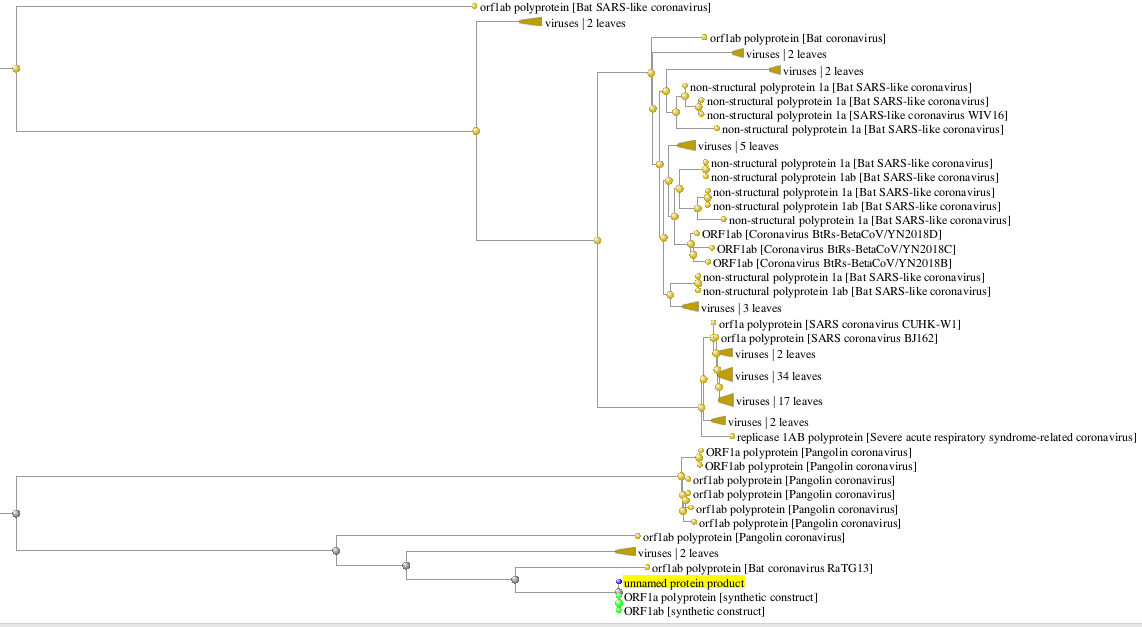
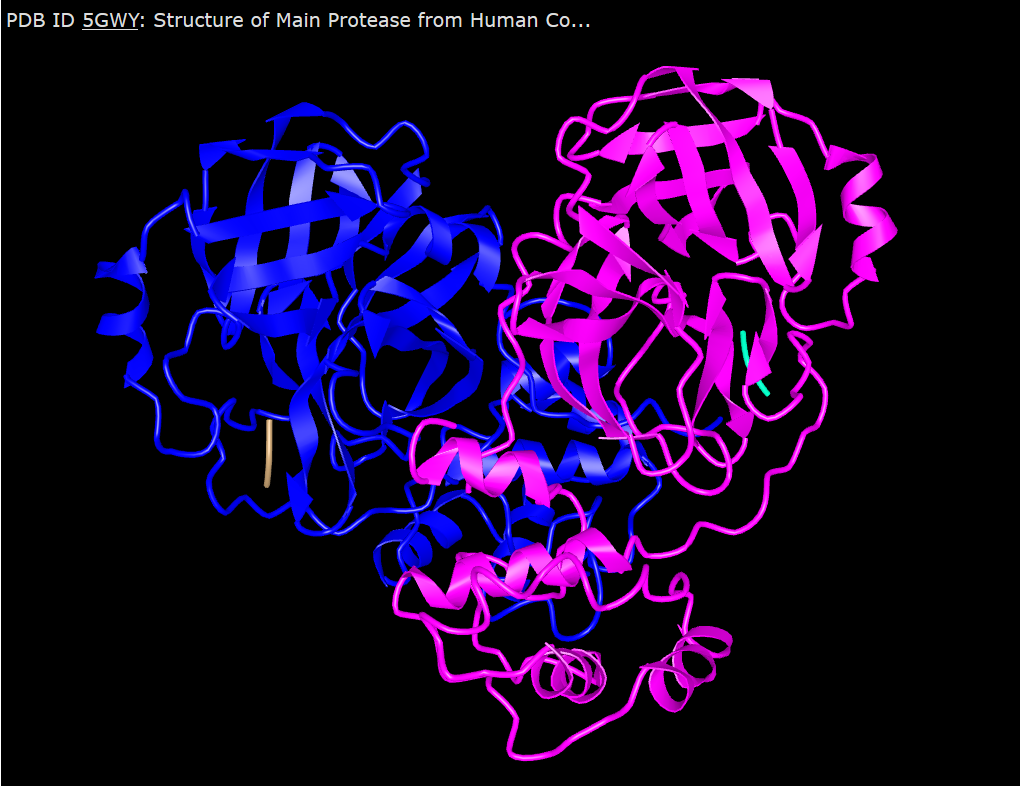


Figure 3: Distance tree for orf1a [11]

Link from accession [P0C6U6.1](https://www.ncbi.nlm.nih.gov/protein/P0C6U6.1?report=genbank&log$=prottop&blast_rank=1&RID=DCW398Y4014) gets us to the structural information:



[13, 14]

“The replicase polyprotein of coronaviruses is a multifunctional protein: it contains the activities necessary for the transcription of negative stranded RNA, leader RNA, sub genomic mRNAs and progeny virion RNA as well as proteinases responsible for the cleavage of the polyprotein into functional products [15].”

# Conclusions

Paralog searching the CHEMBL\_25 database with ORFs from the SARS-CoV-2 genome has found a target and three promising drugs that have already been used in the treatment of Tuberculosis.

Success with these drugs in Tuberculosis treatment may have been due to their affect in combination with other drugs to quell concurrent Corona virus infection.

Computational studies could be used to assess whether these drugs can dock with the target.

Next, in vitro screens should be tried using these drugs.

# References

1. **Early Release - Case-Fatality Risk Estimates for COVID-19 Calculated by Using a Lag Time for Fatality - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases journal - CDC**. 2020.

2. **Coronavirus Age, Sex, Demographics (COVID-19) - Worldometer**. 2020.

3. Pollastri MP, Campbell RK: **Target repurposing for neglected diseases**. *Future Med Chem* 2011, **3**(10):1307-1315.

4. Singer JB: **Finding Paralog Targets for Neglected Diseases**. In*.* Edited by paralog\_targets.pdf; 2020: Methods for finding targets and drugs for a diisease caused by a pathogen searching the ChEMBL database for existing targets with protein sequences similar to ones in the pathogen genome.

5. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B *et al*: **ChEMBL: a large-scale bioactivity database for drug discovery**. *Nucleic Acids Res* 2012, **40**(Database issue):D1100-1107.

6. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E *et al*: **The ChEMBL database in 2017**. *Nucleic Acids Res* 2017, **45**(Database issue):D945-954.

7. Lv W, Xu Y, Guo Y, Yu Z, Feng G, Liu P, Luan M, Zhu H, Liu G, Zhang M *et al*: **The drug target genes show higher evolutionary conservation than non-target genes**. In: *Oncotarget.* vol. 7; 2016: 4961-4971.

8. Wheeler TJ, HHMI Janelia Farm Research Campus A, VA 20147, USA, Eddy SR, HHMI Janelia Farm Research Campus A, VA 20147, USA: **nhmmer: DNA homology search with profile HMMs**. *Bioinformatics* 2013, **29**(19):2487-2489.

9. **Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co - Nucleotide - NCBI** [<https://www.ncbi.nlm.nih.gov/pubmed/>]

10. **EMBOSS: The European Molecular Biology Open Software Suite (2000)** [<http://emboss.sourceforge.net>]

11. **NCBI Blast:(2) - CHEMBL2363965\_8515** [<https://www.ncbi.nlm.nih.gov/pubmed/>]

12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. In: *Nucleic Acids Res.* vol. 25; 1997: 3389-3402.

13. **RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1 - Protein - NCBI** [<https://www.ncbi.nlm.nih.gov/pubmed/>]

14. group NCS: **5GWY: Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design**. 2020.

15. **RCSB PDB - Protein Feature View - Replicase polyprotein 1ab - P0C6X9 (R1AB\_CVMA5)** [<http://www.rcsb.org/pdb/protein/P0C6X9>]