Setayesh Yazdani

M.Sc. Student at Structural Genomics Consortium (SGC) Toronto

Pharmacology and Toxicology Department, University of Toronto

****

"Genetic Variability at the Remdesivir-binding Pocket of SARS-CoV-2 RNA-dependent RNA polymerase Across Coronaviruses and SARS-CoV-2 Samples"

****

Background:

SARS-CoV-2 which is a positive-strand RNA virus depends on a multi-subunit machinery for its RNA replication. The catalytic subunit of RNA-dependent RNA polymerase (RdRp) is the core component of this machinery which is also known as nsp12. Nsp12 relies on accessory subunits to have full activity. The accessory factors, nsp7 and nsp8 increase RdRp template binding and processivity. (Yin et al., 2020)

RdRp is a well-established target for a class of antiviral drugs known as nucleotide analogs. One of the known and only FDA and Health Canada approved drugs for COVID-19, remdesivir, targets RdRp.

Remdesivir is converted to its active form, remdesivir triphosphate (RTP), that gets incorporated into RNA and causes chain termination.

It is important to note that remdesivir has been tested in phase III clinical trial and the results were shared in The New England Journal of Medicine on May 27, 2020 by Goldman and his colleagues. (Goldman et al, 2020) The use of remdesivir for 5 to 10 days in patients with severe COVID-19 showed that there was no significant difference between the 5-day treatment and 10-day treatment groups. The overall effect on the survival rate was rather low. It is important to note this study did not have a placebo control arm and that means the magnitude of the benefit cannot be determined.

Yin et al solved the crystal structure of RdRp in complex with a Remdesivir-incorporated RNA strand (PDB code 7bv2). The authors co-expressed nsp7 and nsp8 with nsp12 to form the RdRp complex. (Yin et al., 2020) They showed that isolated nsp12 is much less active than the trimeric complex (nsp12-nsp7-nsp8) in terms of binding to a 50-base partial double-stranded template-primer RNA. They also showed that the RNA-polymerization activity of the complex on a poly-U template was inhibited by addition of Remdesivir triphosphate (RTP). 1 mM RTP completely inhibited RdRp polymerization activity. Remdesivir as a prodrug and in monophosphate form had no effect on polymerization activity of RdRp.

The authors published two structures of RdRp; one in apo form containing one nsp12, nsp7 and two nsp8 proteins (PDB: 7bv1) and another bound to the Remdesivir-incorporated RNA, containing nsp12, one nsp7, and one nsp8 (PDB: 7bv2).

The structure of the apo RdRp complex which contains one nsp12, one nsp7, and one nsp8 is shown in figure 1A. The structure of the template RNA-Remdesivir RdRp complex contains one nsp12, one nsp7, and one nsp8 in complex with 14-base RNA in the template primer strand and the inhibitor Remdesivir monophosphate (RMP) bound covalently to the 3' end of the primer strand as shown in figure 1B. Upon incorporation of the RTP into the primer strand, there is immediate termination of chain elongation.

This inhibitory mechanism is called nonobligate RNA chain termination by Remdesivir where the prodrug needs conversion to the active triphosphate form. The interactions between RMP and RdRp are shown in figure 2.

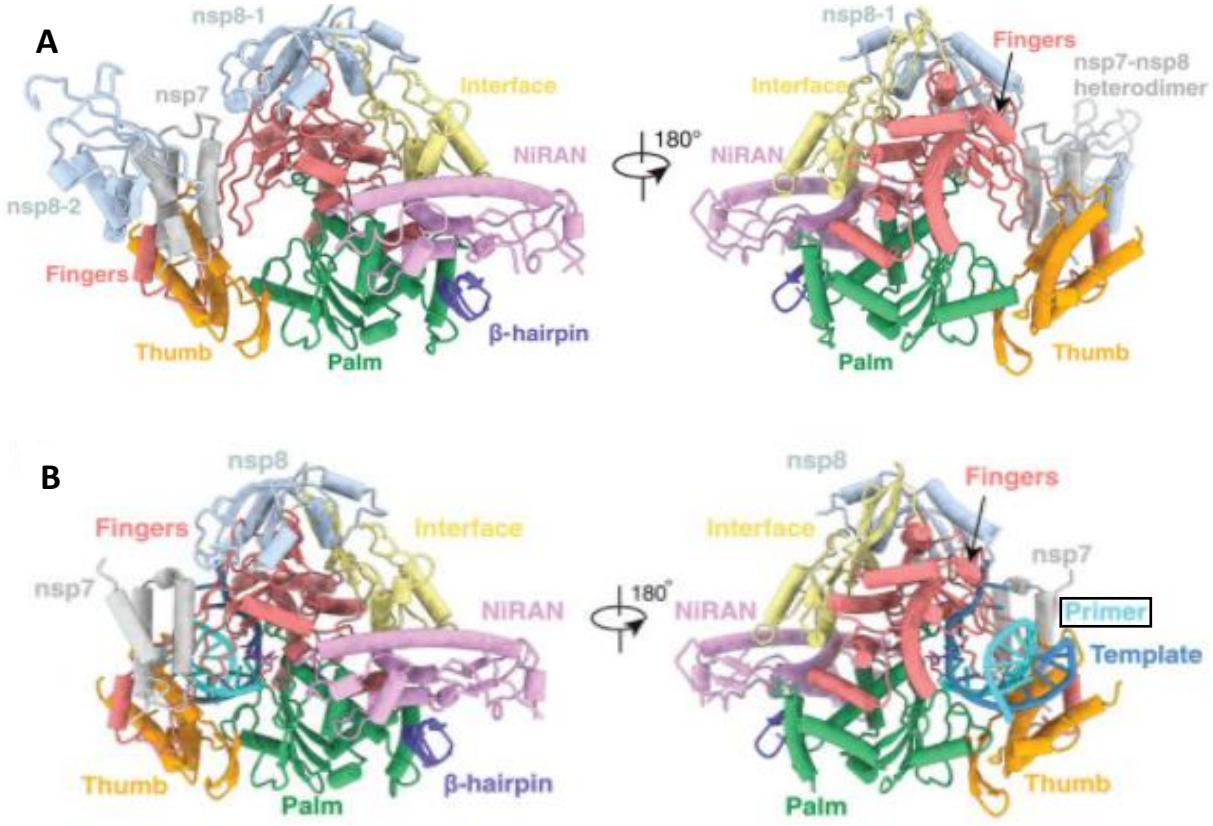


Figure 1. (A) Views of Cryo-electron microscopy structure of nsp12-nsp-7-nsp8 RdRp complex (B) and nsp12-nsp7-nsp8 in complex with the primer RNA template and remdesivir (Yin et al., 2020)

Once RdRp is in the template-RTP form, it goes into a closed conformation where the 11 base-pairs template primer is held by the finger-palm-thumb subdomains. There are also extensive RNA-protein interactions which involve the RNA phosphate-ribose backbone. In the absence of specific interactions between any specific base-pairs and RdRp, the binding of RNA by RdRp becomes independent of the RNA sequence which explain why no specific sequence is needed for RdRp's enzymatic activity.

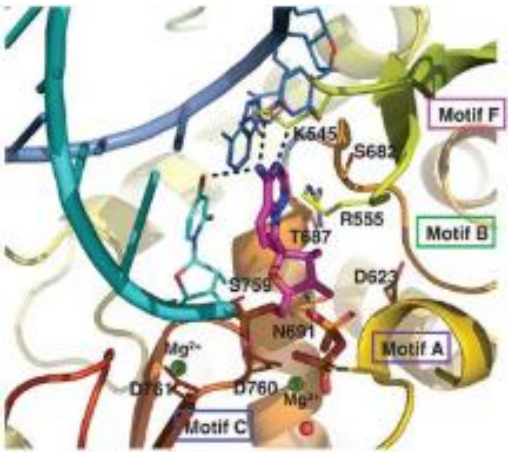


Figure 2. The chemical interactions between the RdRp catalytic site residues and RMP (purple). (Yin et al., 2020)

Goal:

To analyze the genetic diversity of RdRp catalytic site across coronaviruses and SARS-CoV-2 samples**.**

Method:

Step 1: Assessing the druggability of SARS-CoV-2 RdRp active site with SiteMap (PDB: 7bv2).

Using SiteMap (Schrodinger, NY) we assessed the druggability of the binding sites using the druggability score (Dscore). The active site of RdRp in the vicinity of bound remdesivir has a Dscore of 0.859 indicative of a difficult druggable site.

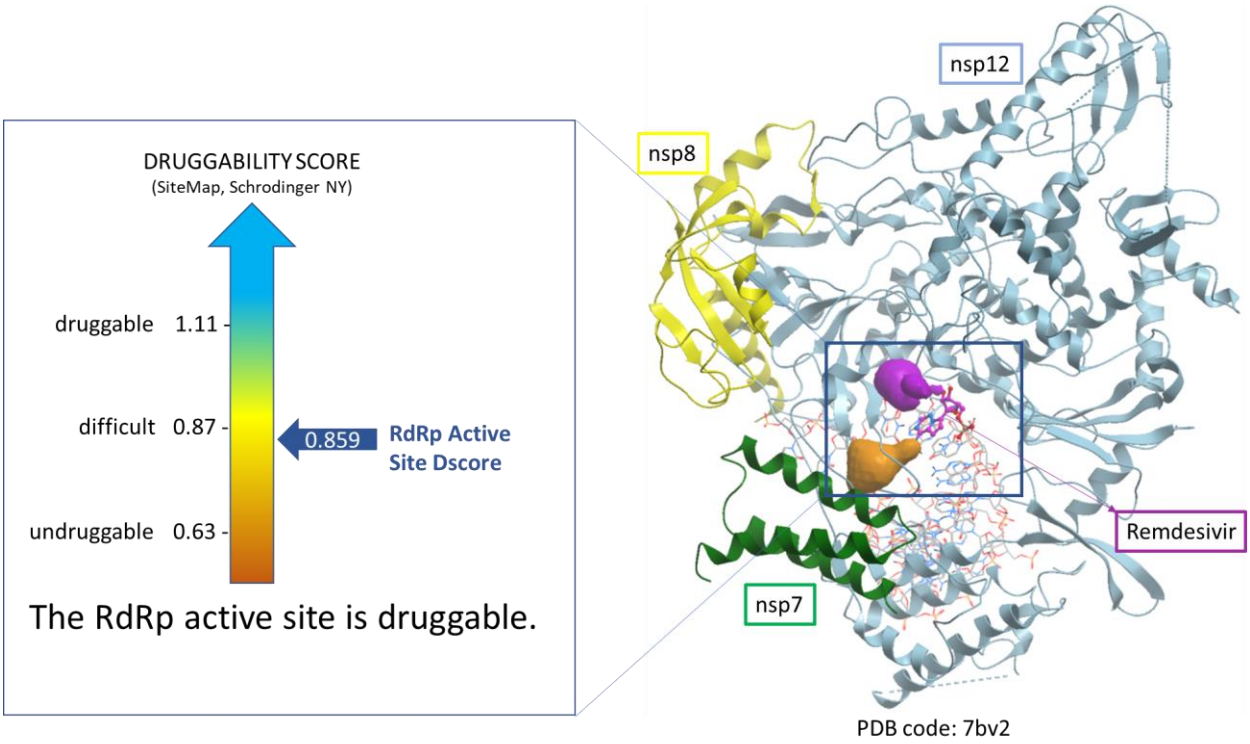


Figure 3. Druggability analysis of SARS-CoV-2 RdRp active site. Cavities at the catalytic site (orange and magenta blobs) are found next to the position occupied by remedesivir monophosphate incorporated into the RNA template **(purple stick).**

Step 2. Determine the residues that line RdRp active site (PDB: 7bv2).

Using ICM PocketFinder, we find that RMP is not occupying a clearly defined binding pocket but is juxtaposed to two cavities. This probably reflects the fact that we are looking at a structure where Remdesivir was already incorporated into RNA. The site is structurally conserved in the apo form (PDB: 7bv1). We hypothesize that compounds occupying these cavities would block the catalytic activity of RdRp. We identify 48 amino acid sidechains lining the pockets juxtaposed to Remdesivir shown in Figures 3 and 4 (within 2.8 Å of the pocket): V410, F440, F441, F442, A443, Q444, D452, Y453, Y455, Y456, Y458, T540, M542, N543, L544, K545, Y546, A547, I548, S549, K551, N552, R553, A554, R555, T556, V557, A558, G559, Y619, P620, K621, C622, D623, R624, V667, K676, G679, T680, S681, S682, G683, D684, T687, A688, N691, S759, D760.

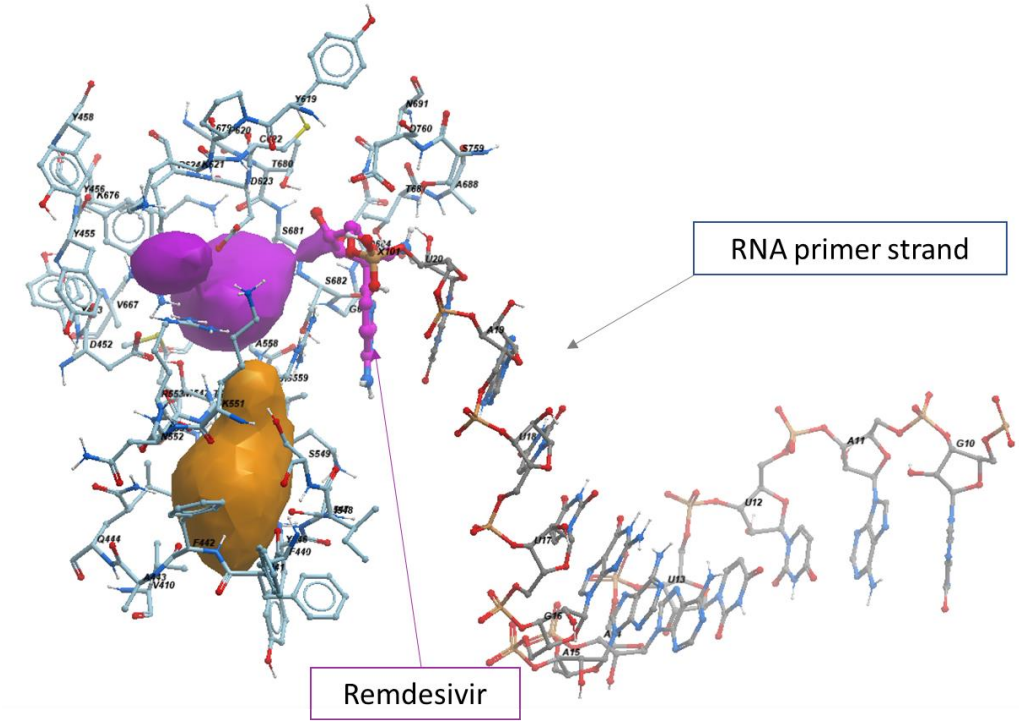
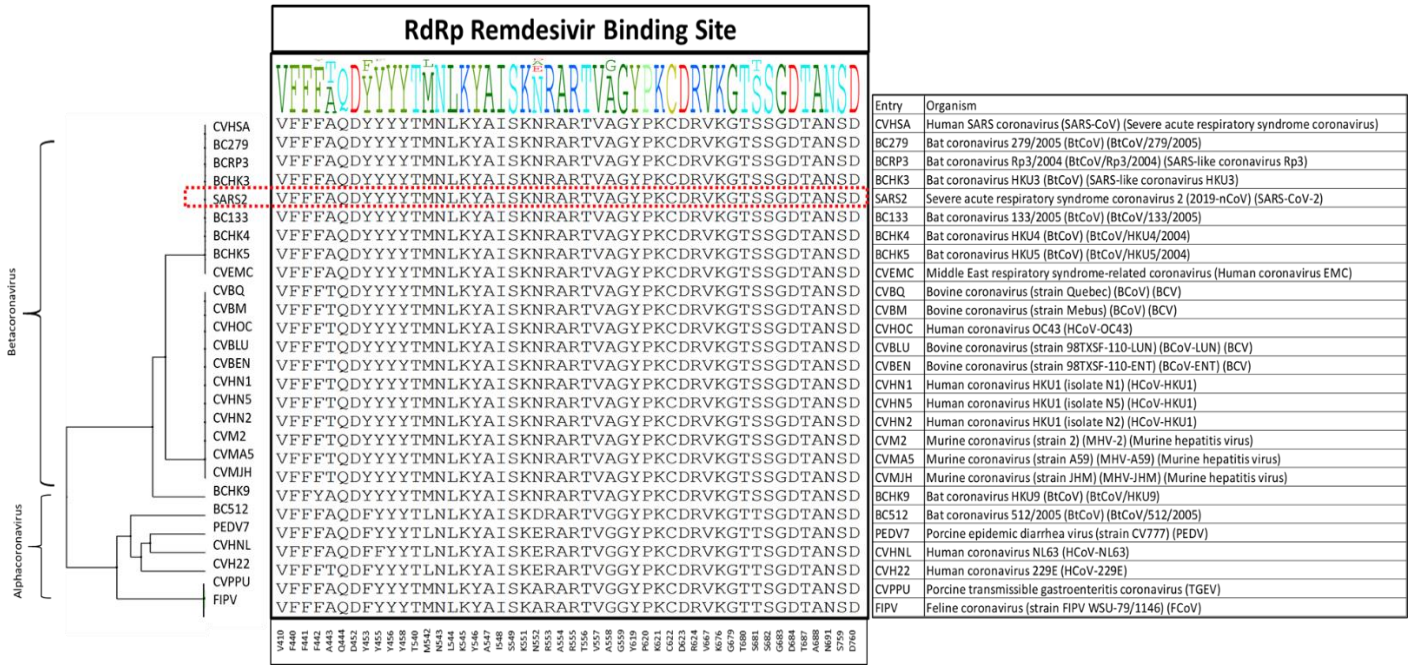


Figure 4. The neighboring residues lining the active site of RdRp (PDB: 7bv2).

Step 3. Make diversity dendrograms for sequences of the Alpha- and Betacoronavirus genera entries in UniProt database.

In the context of the emergence of future MERS-like or SARS-like coronaviruses from the bat strains circulating in bat reservoir species, it is imperative to do a broad survey of viral proteins to identify the best strategies for the development of broad-spectrum viral inhibitors. (Sheahan et al., 2020) For this exact reason, we are looking at twenty-seven reviewed sequences from Uniprot, where six belong to entries of Alphacoronavirus genus and twenty-one belong to the Betacoronavirus genus. We then assess the variability of amino acid lining the binding pockets found at the RdRP active site (PDB: 7bv2).

I made a diversity dendrogram focusing on the sidechains from step2. Shown below in figure 5, the consensus profile at each of the 48 sidechains is shown. Strikingly, 83% of residues, 40 out of 48, are conserved across these 27 entries. This is one of the most conserved SARS-CoV-2 binding sites that we have seen so far (I will post a systematic comparison once I have reviewed all sites on all SARS-CoV-2 proteins in the PDB). Here are the 8 non-conserved residues: F442, A443, Y453, Y455, M543, N552, A558, S681.

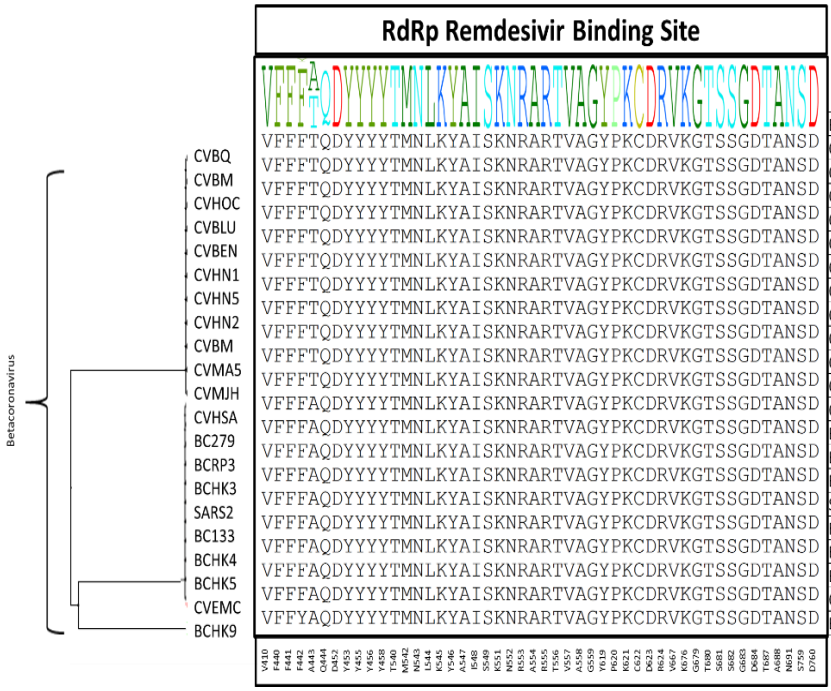


Entry	Organism
CVHSA	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)
BC279	Bat coronavirus 279/2005 (BtCoV) (BtCoV/279/2005)
BCRP3	Bat coronavirus Rp3/2004 (BtCoV/Rp3/2004) (SARS-like coronavirus Rp3)
BCHK3	Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)
SARS2	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)
BC133	Bat coronavirus 133/2005 (BtCoV) (BtCoV/133/2005)
BCHK4	Bat coronavirus HKU4 (BtCoV) (BtCoV/HKU4/2004)
BCHK5	Bat coronavirus HKU5 (BtCoV) (BtCoV/HKU5/2004)
CVEMC	Middle East respiratory syndrome-related coronavirus (Human coronavirus EMC)
CVBQ	Bovine coronavirus (strain Quebec) (BCoV) (BCV)
CVBM	Bovine coronavirus (strain Mebus) (BCoV) (BCV)
CVHOC	Human coronavirus OC43 (HCoV-OC43)
CVBLU	Bovine coronavirus (strain 98TXSF-110-LUN) (BCoV-LUN) (BCV)
CVBEN	Bovine coronavirus (strain 98TXSF-110-ENT) (BCoV-ENT) (BCV)
CVHN1	Human coronavirus HKU1 (isolate N1) (HCoV-HKU1)
CVHN5	Human coronavirus HKU1 (isolate N5) (HCoV-HKU1)
CVHN2	Human coronavirus HKU1 (isolate N2) (HCoV-HKU1)
CVM2	Murine coronavirus (strain 2) (MHV-2) (Murine hepatitis virus)
CVMA5	Murine coronavirus (strain A59) (MHV-A59) (Murine hepatitis virus)
CVMJH	Murine coronavirus (strain JHM) (MHV-JHM) (Murine hepatitis virus)
BCHK9	Bat coronavirus HKU9 (BtCoV) (BtCoV/HKU9)
BC512	Bat coronavirus 512/2005 (BtCoV) (BtCoV/512/2005)
PEDV7	Porcine epidemic diarrhea virus (strain CV777) (PEDV)
CVHNL	Human coronavirus NL63 (HCoV-NL63)
CVH22	Human coronavirus 229E (HCoV-229E)
CVPPU	Porcine transmissible gastroenteritis coronavirus (TGEV)
FIPV	Feline coronavirus (strain FIPV WSU-79/1146) (FCoV)

Figure 5. The diversity dendrogram of the active site of RdRp for the entries of Alpha- and Betacoronavirus genera.

Step 4: Make diversity dendrograms for reviewed sequences of the Betacoronavirus genus entries in UniProt database.

I made a diversity dendrogram of the 21 entries of the Betacoronavirus genus (including SARS-CoV-2). As shown in figure 6, 46 out of 48 residues are conserved at RdRp active site. Here are the non-conserved residues: F442 (almost perfectly conserved), and A443.



Entry	Organism
CVBQ	Bovine coronavirus (strain Quebec) (BCoV) (BCV)
CVBM	Bovine coronavirus (strain Mebus) (BCoV) (BCV)
CVHOC	Human coronavirus OC43 (HCoV-OC43)
CVBLU	Bovine coronavirus (strain 98TXSF-110-LUN) (BCoV-LUN) (BCV)
CVBEN	Bovine coronavirus (strain 98TXSF-110-ENT) (BCoV-ENT) (BCV)
CVHN1	Human coronavirus HKU1 (isolate N1) (HCoV-HKU1)
CVHN5	Human coronavirus HKU1 (isolate N5) (HCoV-HKU1)
CVHN2	Human coronavirus HKU1 (isolate N2) (HCoV-HKU1)
CVM2	Murine coronavirus (strain 2) (MHV-2) (Murine hepatitis virus)
CVMA5	Murine coronavirus (strain A59) (MHV-A59) (Murine hepatitis virus)
CVMJH	Murine coronavirus (strain JHM) (MHV-JHM) (Murine hepatitis virus)
CVHSA	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)
BC279	Bat coronavirus 279/2005 (BtCoV) (BtCoV/279/2005)
BCRP3	Bat coronavirus Rp3/2004 (BtCoV/Rp3/2004) (SARS-like coronavirus Rp3)
BCHK3	Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)
SARS2	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)
BC133	Bat coronavirus 133/2005 (BtCoV) (BtCoV/133/2005)
BCHK4	Bat coronavirus HKU4 (BtCoV) (BtCoV/HKU4/2004)
BCHK5	Bat coronavirus HKU5 (BtCoV) (BtCoV/HKU5/2004)
CVEMC	Middle East respiratory syndrome-related coronavirus (Human coronavirus EMC)
BCHK9	Bat coronavirus HKU9 (BtCoV) (BtCoV/HKU9)

Figure6. The diversity dendrogram of RdRp active site residues among the UniProt entries of the Betacoronavirus genus.

Step 5: Map variations observed among Alpha- and Betacoronavirus genera entries and the variations within only the Betacoronavirus entries onto SARS-CoV-2 RdRp crystal structure (PDB: 7bv2).

We mapped the variations shown in step 3 and 4 onto SARS-CoV-2 RdRp color-coded crystal structure in figure 7 such that sidechains that are variable among the entries are highlighted in orange and the conserved sidechains are shown in green.

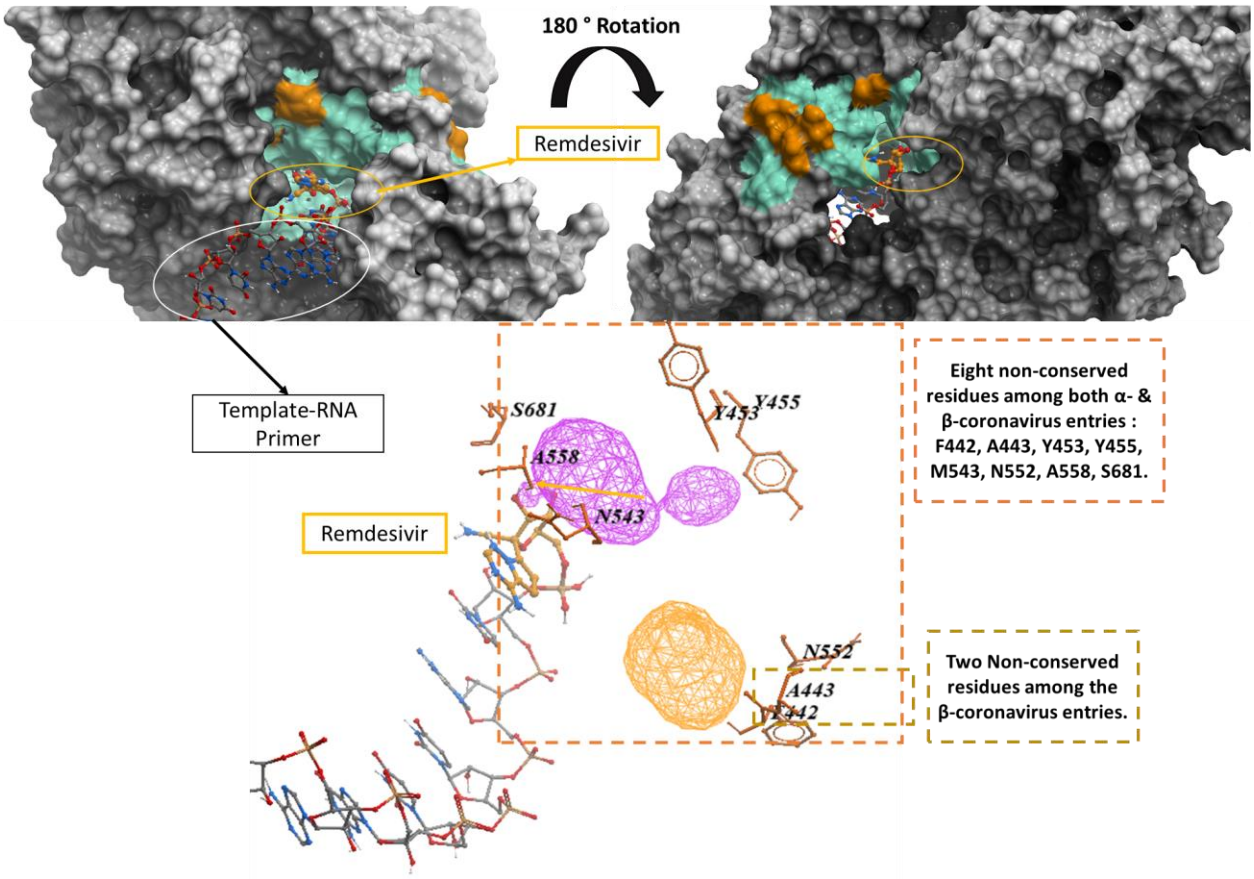


Figure 7. Mapping the genetic variation of UniProt entries belonging to Alpha- and Betacoronavirus genera onto SARS-CoV-2 RdRp Crystal structure (PDB: 7bv2). The eight non-conserved residues among the entries of both Alpha- and Betacoronavirus include: F442, A443, Y453, Y455, M543, N552, A558, S681 while the two non-conserved residues among the Betacoronavirus entries include F442, A443.

Step 6: Assessing the genetic variability of catalytic site of RdRp among SARS-CoV-2 samples.

Nicola De Maio, our collaborator from Nick Goldman's lab at European Bioinformatics Institute, looked at more than 15000 SARS-CoV-2 samples and identified all mutations at the RdRp catalytic site pockets. He identified 6 non-synonymous variants as shown in table 1. For example, at residue number 443, the wild-type sidechain in SARS-CoV-2 RdRp is alanine. In the sample batch Nicola has looked at, he saw that 15870 samples had alanine at that position while there were 2 samples with valine at this exact position.

Index	Non-synonymous Variants	variants	codon one	codon two	codon three
1	A443	A(15870), V(2)	g(15872)	c(15870), t(2)	t(15872)
2	Q444	Q(15870), H(2)	c(15872)	a(15872)	t(2), g(15870)
3	M542	M(15873), L(1)	a(15874), c(1)	t(15875)	g(15874)
4	A547	A(15874), V(1)	g(15875)	c(15874), t(1)	c(15875)
5	A554	A(15874), V(1)	g(15875)	c(15874), t(1)	c(1), t(15874)
6	V667	V(15870), I(5)	a(5), g(15870)	t(15875)	c(15875)

Table 1. The non-synonymous variants at RdRp active site across more than 15000 SARS-CoV-2 samples.

We could not predict the effect of these mutations on remdesivir binding due to the lack of RdRp structure where remdesivir has not been incorporated into the RNA primer sequence while it is sitting in a pocket in RdRp. Therefore, we can only comment on how the nature of the variant sidechains differ from the wild-type sidechains according to their physiochemical properties: the size and polarity of mutated residues are largely conserved in all cases (A->V, Q->H, V->I), suggesting high chances of a light effect on ligand binding.

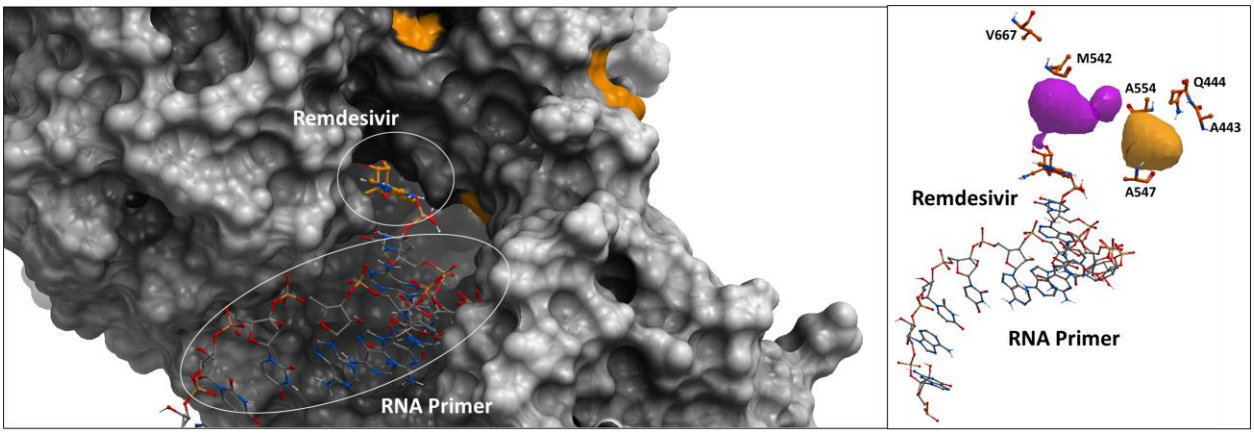


Figure 8. Left Panel: The surface representation of SARS-CoV-2 RdRp crystal structure with the non-conserved amino acid positions highlighted in orange. Right Panel: The sticks representation of the non-conserved residues among SARS-CoV-2 samples.

Conclusion:

From a structural standpoint the catalytic site of RdRp is a good target for broad-spectrum inhibitors against coronaviruses considering that only 17% of the residues lining catalytic site binding pockets were not conserved across 27 Alpha- and Betacoronavirus entries. It is worth mentioning that clinically, the inhibition of RdRp which is achieved by remdesivir has not been fully confirmed to be an effective therapy for COVID-19.

According to the SARS-CoV-2 variants reported in table 1, the properties of most of the non-synonymous variants at the catalytic site were conserved, suggesting high chances of a light effect on ligand binding.

It would be crucial to investigate the structure of RdRp combined with remdesivir in the state right before it gets incorporated into the RNA primer. This would reveal the main interactions that remdesivir makes with RdRp sidechains and allow to better predict the effect of mutations both across coronaviruses broadly and more specifically across samples from COVID-19 patients on remdesivir binding.

References:

Yin, W., Mao, C., Luan, X., Shen, D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y., Tian, G., Jiang, H., Tao, S., Shen, J., Jiang, Y., Jiang, H., Xu, Y., & Xu, H. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, *368*(6498), 1499-1504. https://doi.org/10.1126/science.abc1560

Goldman, J. D., Lye, D. C.B., Hui, D. S., Marks, K. M., Bruno, R., Montejano, R., Spinner, C. D., Galli, M., Ahn, M., Nahass, R. G., Chen, Y., SenGupta, D., Hyland, R. H., Osinusi, A. O., Cao, H., Blair, C., Wei, X., Gaggar, A., Brainard, D. M., & Subramanian, A. (2020). Remdesivir for 5 or 10 Days in Patients with Severe Covid-19. *The New England Journal of Medicine*. https://doi.org/10.1056/NEJMoa2015301

Sheahan, T. P., Sims, A. C., Zhou, S., Graham, R. L., Hill, C. S., Leist, S. R., Schäfer, A., Dinnon, K. H., Montgomery, S. A., Agostini, M. L., Pruijssers, A. J., Chapell, J. D., Brown, A. J., Bluemling, G. R., Natchus, M. G., Saindane, M., Kolykhalov, A. A., Painter, G., Harcourt, J., … Baric, R. S. (2020). An orally bioavailable broad-spectrum antiviral inhibits SARS-Cov-2 and multiple endemic, epidemic and bat coronavirus. https://doi.org/10.1101/2020.03.19.997890