COVID-19 POTENTIAL TREATMENT/PROPHYLAXIS METHOD

RESEARCH PROPOSAL

Solubilized oligonucleotide ssDNA aptamers designed against the RBD motif of Spike protein delivered in dose-dependent manner through inhalation

Steve Liebich[1,2]

ORCID: 0000-0002-5379-4532

[1] Clarkson University, Potsdam (NY), Biomolecular Science Department

[2] Babu™ (**BabuBio, LLC**), Founder & CEO

Phone: +1 (315) 262 7005

Email: steven@babusolutions.com

I. Proposal Overview

Oligonucleotide aptamers synthesized through SELEX (Systemic Evolution of Ligands Through Exponential Enrichment) are more efficient, sensitive, selective, and cheaper than the standardly utilized antibodies. The well-established antibody industry surpasses the potential of aptamers, even though aptamers outweigh antibodies in almost every aspect of biological activity and production. Proposed is a therapeutic method based on single-stranded DNA aptamers targeted against the specific sequence (the RBD motif) of Spike protein responsible for SARS-CoV2 entry into human epithelial cells. Solubilized and packed in extremely cheap, pressurized metered-dosed inhalers (pMDI), the selectively chosen aptamer(s) can be quickly and effectively inhaled in the upper respiratory tree in the form of aerosolized particles covering the epithelial surface and blocking the SARS-CoV2 entry into the cells, thus inhibiting its replication. Due to lack of toxicity, high reproducibility, and cost-effectiveness, the designed drug (BABU-ONE™) can be safely used by every person worldwide for almost no cost. BABU-ONE™ works also as a prophylactic drug, protecting an individual from COVID-19, hence securing the safety of people finding themselves in the greatest danger (medical staff, volunteers, essential employees, elders). This is the first proposed treatment method based on solubilized aptamers and first of a kind suggested to be used in the fight against COVID-19.

II. Background and Significance

COVID-19 Global Pandemic.

A novel coronavirus, named by the WHO as "2019-nCoV" but also known as SARS-CoV2, is responsible for the new acute respiratory distress syndrome (ARDS) that originated from Wuhan (Hubei Province) in China on December 12th, 2019 [1]. The disease was designated as COVID-19. On March 11, the WHO officially acknowledged that COVID-19 had become a pandemic. On March 18, there were 191,127 confirmed cases of COVID-19 and 7807 deaths in more than 190 different countries and territories. As of April 3, there are 1,081,952 confirmed coronavirus cases and 58,149 deaths in 207 countries and territories. The situation worsens every day with no promise of transmission deceleration and with no effective treatment.

Prevailing symptoms are fever (over 100 ºF), cough, and breathing difficulties [2, 3], but myalgia, nasal congestion, GI symptoms, anosmia, and dysgeusia are also reported [4, 5]. The incubation period for COVID-19 is approx. 5 – 6 days, with a range of 1 to 14 days [6, 7], and with the reproductive number R_0 of 3.25 to 3.4, which is extremely high when compared to SARS [8]. The definite clinical diagnosis must be confirmed by RT-PCR testing of respiratory tract

samples [9]. Even though the main route of transmission is human-to-human through respiratory and extrarespiratory secretions [10], the virus shedding can also be detected by testing anal swabs, blood, and serum samples [11, 12]. The SARS-CoV2 nucleic acid shedding pattern resembles that of patients infected with *Influenza* rather than with SARS-CoV, and it is suggested that the asymptomatic transmission is possible and that individuals infected with SARS-CoV2 can be contagious before the onset of first symptoms [13-15].

Current therapeutic methods can be divided into four main groups. In Group I, combinations of oral and IV anti-viral drugs serve as a supplemental treatment approach with little to no therapeutic effects [16-18]. Group II represents antibiotics and general antimicrobial drugs that show some therapeutic efficacy against SARS-CoV2 [19-21]. Potential drugs of Groups III and IV are currently under development; Group III represents novel therapeutic molecules targeting either the antigenic proteins on the surface of SARS-CoV2, the virus replication machinery enzymes, or the human ACE2 (angiotensin-converting enzyme 2) receptor which are directly involved in virus entry into the ciliated epithelial (mainly bronchial) cells and type II pneumocytes [22-25]. Group IV pertains to putative vaccines that still await their development and clinical trials [26]. In sum, no clinically relevant therapeutic approach has been developed that could stop the disease spreading and save the infected patients' lives.

Insights Into the SARS-CoV2 Morphology and Genomic Data.

SARS-CoV2 is a single-stranded, positive-RNA *betacoronavirus* belonging to the family *Coronaviridae* which encompass two other highly infectious coronavirus strains, SARS-CoV and MERS-CoV, responsible for acute viral pneumonia outbreaks in 2002 and 2012, respectively [27, 28]. Based on genomic and comparative analysis of all the three coronaviruses, the similarity identity between SARS-CoV2 and SARS-CoV is 79%, and only 50% when compared to MERS-CoV [29]. There is growing evidence that all three originate from bats as the primary reservoirs [29, 30, 31]; for SARS and MERS, civet cats and camels were the amplifier hosts, respectively, and SARS-CoV2 has also been transmitted from bats to potential secondary hosts until it generated the human-only transmission [29, 32].

The RNA genome of SARS-CoV2 is made of approx. 29,674 nucleotides, with 10 coding sequences (CDS) and 14 open reading frames (ORFs) encoding 27 proteins: 15 non-structural proteins (nsps), 8 accessory proteins, and 4 main structural proteins: Spike (S), Matrix (M), Envelope (E), and N protein [29, 32, 33]. It is the S glycoprotein that binds the human ACE2 receptor and enables the enveloped virus to fuse with the plasma membrane and enter the cell [34]. Homotrimeric Spike protein consists of two subunits: S1 and S2. S1 allows SARS-CoV2

bind to ACE2, especially through its region-binding domain (RBD), whose mutations have been shown to be responsible for the virus superinfectivity and high transmission [32-34]. S2 subunit mediates the membrane fusion process [34, 35]. The interaction between Spike and ACE2 is a critical step in the replication cycle of SARS-CoV2. For detailed genomic and morphological information of SARS-CoV2, see Supplementary Material.

SELEX and High-Affinity Oligonucleotide Aptamers

Aptamers are short, highly structured, single-stranded oligonucleotides; they bind to a target molecule with a limited number of specific attachment sites, which defines their high selectivity used commonly in many biomolecular techniques and therapies [36, 37]. SELEX (Systemic Evolution of Ligands Through Exponential Enrichment) is an effective and verified method of isolating the high-affinity aptamers, which can remain unmodified or conjugated to miscellaneous chemical compounds as well as they can be chemically changed to bind to various biological molecules [38, 39]. The first FDA-approved pharmaceutical aptamer is Macugen (*pegaptanib sodium*, **Roche**), which in 2004 was admitted for treatment of Age-Related Macular Degeneration (AMD) [40]. Modern clinical therapy notes an enormous promise in these small and effective oligonucleotides.

Aptamers have also found their pharmaceutical niche as anti-viral drugs. Examples range from HIV-1 [41], to HCV [42], to Ebola [43], and to *Influenza* with aptamers targeted against NTPase [44], full HA1 protein [45, 48], glycosylated HA1 [46], and truncated (recombinant) HA protein [47]. The first two parameters of the *Influenza* virus infectivity inhibition is inhibition of hemagglutination (of chicken red blood cells) and rescuing *in vitro* cell cultures from a detrimental viral load (decreasing cell mortality). **What must be noted as paramount**, the A22 aptamer from [47] showed an imminent prevention action against the used A/Port Chalmers/1/73 virus even after 30 min exposure of the MDCK cells to the virus. The [47] paper is exceptional in one more sense: the researchers tested the effectiveness of their aptamer (A22) *in vivo*, by inoculating the *Influenza* virus in mice; the viral load was 1.2 log lower than in the infected control group by treating the mice intranasally with 50 μL of 2.5 nmol/mL of the aptamer. **The mice treated with the aptamer before being introduced to the virus showed over a hundred-fold (2.4 log) reduction in their viral titer**. This showcases a putative prophylactic effect of aptamers for viral diseases.

It is the cited *Influenza* papers that I have used as my guidelines in developing the following treatment/prophylactic proposal.

III. Objectives.

Goal:

My goal is to develop a set of single-stranded oligonucleotide DNA aptamers with high binding affinity for the RBD motif (of SARS-CoV2) to human ACE2 receptors which will disrupt SARS-CoV2 entry into human lung epithelial cells (and other ciliated epithelial cells of the upper and lower respiratory tract). This potential COVID-19 drug will slow the infection process (no viral replication) and help the host immune system respond against SARS-CoV2. The highly selective aptamers, designed through the SELEX method, will be deposited as aerosolized drug in the oropharyngeal region and upper airways, which also can serve as a protective and prophylactic drug device best used in the individuals who have not been exposed to SARS-CoV2 yet or are going to the areas with high infection rate.

Specific objectives:

1. ***Objective 1: Expression and purification of the recombinant SARS-CoV2 Spike glycoprotein***.

   As presented above and described more in detail in the Supplementary Materials section, a specific target for the aptamer must be chosen in order for the binding to occur. Different sizes of targeted proteins have been used from a few amino acids to complete long-chain polypeptides. However, it must be ascertained that the binding specificity between the produced aptamer through the proposed method and Spike protein is as selective and strong as possible. Therefore, it must be the **RBD motif** (**437 → 508 amino acids**) that will be expressed in an expression plasmid in *E. coli*. There is no glycosylated residue between 437 and 508 (accessed from [49]), which will facilitate the PCR amplification and sequence cloning steps. The recombinant target protein (truncated Spike S1 subdomain) must be multihistidine (6x)-tagged.

2. ***Objective 2*: *In vitro selection of ssDNA aptamers with high binding affinity for the recombinant SARS-CoV2 Spike glycoprotein***.

   a. ***Objective 2.1.: Selection and synthesis of the pool of ssDNA oligonucleotides with a 40-nt randomized region flanked by two well-defined regions.***

   Since we must be certain that we have selected the most effective aptamers in terms of their binding affinity, a large pool of randomly synthesized DNA oligonucleotides must be prepared. According to [44-46, 48], a 40-nt randomized sequence (which will serve the role of the

aptamer core binding element) must be flanked by two short (~15 to 20 nucleotides) sequences with the restriction sites, which are also present in the plasmid, used to further clone the obtained well-functioning aptamer (I suggest the pGEMT plasmid). Proper primers must be chosen for the amplification step performed at the end of each SELEX cycle; the primers must bind the flanking regions.

b. ***Objective 2.2.: SELEX procedure, measurement of affinity of aptamers, cloning and sequencing of specific ssDNA aptamers.***

Each laboratory chooses its own standard procedure for SELEX, selecting proper incubation time and conditions, immobilization techniques, binding and elution buffers, and preservation methods. Below (Methodology), I go step by step through each methodology chosen for all objectives and the recommend desired conditions and buffers for each step. However, the ultimate decision on how the SELEX part should be executed I leave to the designated research group. Based on an extensive literature search and available data, I recommend using ssDNA aptamers due to their potentially lesser toxicity and smaller doses needed to be applied to the cellular cultures *in vitro* and to the infected individuals *in vivo*. Moreover, it is more time-effective to use ssDNA rather than ssRNA oligonucleotides which must be first transcribed to DNA for the cloning purposes.

3. Objective 3: Testing the selected aptamers in an engineered VeroE6/TMPRSS2 cell line infected with an "L" strain of SARS-CoV2.

Papers [45-48] use MDKC (Madin-Darby canine kidney) cell lines as the standard cell culture for *Influenza* incubation and aptamers therapeutic action, but I recommend using the VeroE6/TMPRSS2 engineered cell line which has been shown to be largely susceptible to SARS-CoV2 infection and available from the Japanese Collection of Research Bioresources (JCRB) Cell Bank in Japan [50, 51].

The "L" strains have been shown to have a higher transmission and a major mutations rate which make them the most dangerous and rapidly evolving strains of SARS-CoV2 [57]. If we want to find a potential aptamer effective in the most endangered and infected areas, we must target it against these strains, the "L" strains, that are the main cause of the severity of present COVID-19 situation.

Note: It has been shown that SARS-CoV2 uses co-receptors on the human cells, including TMPRSS2, a transmembrane serine protease [25, 52]. It has also been shown that type II pneumocytes, enterocytes in the terminal ileum, and nasal goblet secretory cells are the human

cells that co-express both ACE2 and TMPRSS2 [53]. The demographic *TMPRSS2* expression among different population groups vary, and it has been noted that the severity of infection of COVID-19 is associated with lesser expression or complete loss of *TMPRSS2* [54, 55]. These findings could explain why specific demographic regions in the world suffer worse from COVID-19 and experience greater virus transmission. Recent research has focused on finding potent existing and novel antibodies and biologically active molecules targeting TMRPSS2 [56].

4. ***Objective 4: Formulation of a pressurized metered-dose inhaler (pMDI) for the solubilized and stabilized form of the drug***

SARS-CoV2 is transmitted through droplets, human to human, and its presence in the blood is dubious and usually correlated with systemic disease [58, 59]. Therefore, the cloned, sequenced, and amplified aptamer ready to be used against SARS-CoV2 must not be delivered through muscle injection or IV, but as aerosolized particles inhaled in the upper respiratory tree (oropharyngeal, tracheobronchial). pMDIs are used in COPD and asthma with a precise aerosol amount and their mass makes them perfect carry-on medical devices which can be used by any trained or non-trained individual.

IV. Proposed approach

Objective 1

- **Suggested Design.**

An "L" nCoV-19 strain prepared from embryonated chicken eggs. RT-PCR primers must be corresponding to the RBD motif region (319 – 541) after RNA extraction in order to synthesize the cDNA transcripts of the region of interest. Amplification of the cDNA should be done under standard protocol. The suggested plasmid for six-multihistidine tag is pQE80-L or pQE30 expressed in many commercially available *E, coli* strains. I suggest using the Ni-NTA resin columns (5 – 10 mg fusion protein/mL capacity) for affinity chromatography [45-47]; elution with imidazol and dialyzed against binding buffer (50 mM Tris-HCl, pH 7.4, 5 mM KCl, 100 mM NaCl, 1 mM $MgCl_2$). Lyophilize or use it immediately in the next step.

- **Expected results.**

At the end of this procedure, we should have the (6x)His-tagged Spike S1 RBD subdomain cloned, expressed, and purified.

- **Further suggestions.**

A different cell culture can be chosen, such as COS, Huh-7, or L20. Another plasmid can be used for the protein expression. Binding buffer may be changed; for example, addition of dithioreitol (DTT) and 1% (w/v) BSA.

Objectives 2.1 & 2.2

- **Suggested Design.**

There are a number of miscellaneous SELEX procedures that can be utilized in this protocol, with every researcher having their individual preferences. Presented below is only an outline for this objective. Details can be found in [44-48, 60-65]; for exhaustive review of SELEX, I refer to [63-65].

a. Selection of randomized oligonucleotide sequences:

Choose a chemically synthesized DNA oligonucleotide library first. I recommend using a big library pool of DNA, approx. 10^{15} randomized oligonucleotides. The randomized region must be 40-nt long. According to ***Objective 2.2,*** presented is an exemplary construct of the randomized ssDNA aptamer with its flanking regions and selected primers designed for this sequence (*in silico*).

5'-GGG**CCATGG**CTT**CCGCGG**CT-(**N_{40}**)-**ATCTG**CAGC**ATGCAT**ATG-3'
NcoI SacII PstI NdeI

5'-GGGCCATGGCTTCCGCGGCT-3' *Forward*
5'-biotin-CATATGCATGCTGACGATT-3' *Reverse*

5'-biotin-ATGGCGTCTCAAGGCACCAACGATCTTATGAACAGATGG-3' [48] *Control*

Random DNA oligonucleotides can be synthesized in lab [66] or accessed through commercially available platforms (Agilent, Arbor Biosciences, ChemGenes Corp.). **This is one of the critical steps in SELEX**, make sure you have a large, well-synthesized DNA aptamers library with the flanking regions constructs. A control sequence must be synthesized (presented in the figure above). The reverse primer is biotinylated, so that future aptamers will be biotinylated as well so that during ssDNA creation (see below), the two strands of dsDNA can be separated.

b. **Partitioning of bound from non-bound aptamers:**

Usually, ssDNA oligonucleotides are denatured at 80ºC for 10 min and cooled on ice for another 10 min. 2-6 μg of the purified recombinant S1 RBD subdomain protein is mixed in an Eppendorf tube with ~ 20 nmoles of the ssDNA oligonucleotides, 100 – 500 μL Ni-NTA Superflow, and with 200 – 500 μL binding buffer; the mixture should be incubated for around 30 min in 37ºC with occasional shaking. *Radioactive markers can be used to assess the quantity of the bound and unbound aptamers*. **Negative selection** is favored by some, where the Ni-NTA Superflow is mixed an initial ssDNA pool for ~30 min in order to eliminate sequences for immobilization matrix components (the proposed control sequence in the figure above). Apart from centrifuging the eluted target-bound aptamers, denaturing methods like EDTA addition or heat treatment can be used to separate the aptamer-target complexes. *It is debatable whether to use a high or low nucleic acid:protein ratio, however, I recommend to use the high ratio in order to increase the number of high-affinity aptamers*. More stringent conditions (buffers, mixing time, separation methods, temp.) are usually used during the first 3-4 cycles, after which this stringency gradually decreases.

c. **Amplification of the initial oligonucleotide aptamers.**

Because the initial ssDNA pool is usually poor, 20-25 rounds of PCR amplification is required. Additional chemical groups can be added during this process.

d. **Turning dsDNA into ssDNA:**

Different methods can be used, but I prefer to use the simple, cheap, and quick method of streptavidin-coated paramagnetic beads which capture the biotinylated amplified dsDNA, which are further subjected to chemical or thermal denaturation, and the non-biotinylated DNA strands are eluted.

e. **Repeating the cycles:**

Steps a → c are repeated until the high-affinity binding ssDNA aptamers are generated. This usually takes 11-20 cycles, depending on the size of a target molecule and used conditions. In this protocol, I assume to obtain the *right* aptamers after ~15 cycles.

f. **Measurement of affinity of the aptamers:**

ELISA (Enzyme-linked Immunosorbent Assay) can be used to evaluate the binding affinity of the obtained aptamers (reverse screen assay) with monoclonal antibodies (mAb)

designed against the multihistidine tag [62]. A steady increase in the optical density (OD), indicates the ssDNA oligonucleotides binding affinity efficacy.

g. **Cloning and characterization:**

The selected aptamer(s) must be cloned into the pGEM-T vector; I used [67] to create the proposed aptamer flanking regions. Because aptamers very often are characterized by AG-rich regions and stem-loop structures, programs such as MFOLD can be used to predict the minimum free energy of the chosen aptamer(s) [68], so that "the best" aptamer can be used for the cloning step. The chosen aptamer is sequenced and expressed using a different vector, and after RNA extraction and cDNA generation, **it can be amplified for an indefinite time**.

- **Expected results.**

At the end of steps a → f, aptamer(s) with the highest binding affinity should be produced**. Step "g" helps select the most optimal ssDNA molecules and offers an infinite production line for as many patients as needed**.

- **Further suggestions.**

Flanking regions of the randomized ssDNA sequences, chemical modification of primers, separation method of the bound vs unbound ssDNA particles, conditions and buffers used, number of cycles, amount of the recombinant purified protein used – these can be modified to any laboratories expectations and protocols.

SELEX aptamer generation can be done by a commercial vendor (Illumina, for instance) or can be (half)-automated [69, 70]. However, it must be noticed that it is always better to perform SELEX in your own laboratory setting without automation, so that the entire process and each step can be tightly controlled and manipulated if needed.

Objective 3

- **Suggested Design.**

The antiviral activity of selected aptamers must be determined 72 hours after VeroE6/TMPRSS2 cell culture infection using the FFA (Focus Forming Assay). Because SARS-CoV2 does not have hemagglutinin (HA) protein, the generic HA (hemagglutination) assay cannot be performed. The FFA assay is based on antibodies targeted against a viral antigen (in this protocol, I suggest to use this newly designed immunoassay kit [71]). Afterwards, the MTT assay must be performed to measure host-cell viability. Usually, aptamers show dose-dependent

effectivity/toxicity changes. The MTT can also be used to evaluate the anti-viral activity of an aptamer (especially useful if the FFA test cannot be performed).

- **Expected results.**

  We expect to achieve the most optimal anti-viral activity of a selected ssDNA aptamer with the smallest dose possible (approx. 25-100 pmol). The FFA and MTT assays must be done with mock groups and non-infected cells to effectively determine that the viral titers decrease with the increase of the aptamer concentration.

- **Further suggestions.**

  Because SARS-CoV2 is transmissible only among people, infection of mice with the coronavirus is impossible. However, mice can still be used to evaluate the toxicity and side-effects of the selected S1 BRD aptamer(s) before clinical trials begin.

Objective 4

- **Suggested Design.**

  If a pMDI is chosen for the pulmonary route of the aptamer delivery, there are many laboratory ways of doing it, but the great advantage is that production of simple pMDIs is its extremely low cost; moreover, simple and fully effective inhalers can be 3D printed from safe-to-use materials. pMDI drug formulation depends on a technique incorporated, but the most feasible way is to suspend the chosen aptamer(s) in a propellant or in a mixture of propellants, such as ethanol and surfactants (long-chained with hydrophilic head groups coupled to enhance the ion-dipole interactions); hydrofluoroalkane (HFA) is the major propellant used in all pMDIs [72-74]. A list of novel, cheap, easy-to-use, and more advanced pMDIs can be found in [75].

- **Expected results.**

  Solubilized aptamer(s) targeting the Spike S1 RBD motif incorporated in a modern inhaler, which is cost- and time-effective, and very safe to use.

- **Further suggestions**

  I am uncertain as to how long the aptamer(s) inhaled into the upper airways remain on the epithelial surface of the cells. This must be tested in animal and further in human clinical studies. Based on a brief literature search, I assume 12 to 24 hours of effectivity with one single dose inhaled.

V. Broader Impact and Potential In Decelerating COVID-19 Spread

Antibodies have been exhaustively used in the modern biology and pharmacy world becoming the almost-sole target of interest for most biotech companies [76]. Since the first time of SELEX employment by Tuerk and Gold in 1990, only Macugen has been approved by FDA in treating AMD [77]. Until now, approx. 20 aptamers have successfully entered clinical trials for a plethora of medical conditions including ovular vascular disease and various cancer types (AS1411 aptamer targeting a few cancer lines, and anemia) [78, 79]. The antibody-oriented trend seems to be even more peculiar if to consider the advantages that aptamers have over antibodies: 1) high flexibility, 2) much easier and cheaper production, and 3) unlimited amplification of once produced aptamer [80]. **It is the simplicity and reduced costs of production of aptamers that wakes skepticism in private firms and biotech giants which slows their proliferation**. 1564 patents associated with aptamers, thousands of publications, multiple initiated clinical trials, and only one commercially available aptamer – **this proposal can bring one more to the money-fat market and make a miracle**.

Aptamers are characterized by low to no immunogenicity when administered in preclinical doses 1000-fold greater than doses used in animal and human therapeutic applications [81, 82]. **No aptamers** designed for clinical trials so far have shown any toxicity or immunogenicity in animals and humans [reviewed in 36, 81, 82]. The only exception was Phase 2b of clinical trials for REG1 (novel anti-coagulant) which exposed some patients to severe allergic reactions, but those were associated with pre-existing anti-PEG (polyethylene glycol) antibodies found in the patients' blood [83]. There are two major concerns related to aptamers toxicology, and these are antibodies produced against PEG-conjugated aptamers (which is not the case in this proposal) and some toxicity of antisense therapeutics (which is also not the case in this proposal) [84-86], both concerns related to RNA aptamers (whereas we will use DNA aptamers). To cite [78]: "**We believe the bottleneck of commercialized aptamers for diagnostic and therapeutic applications is not due to any deficiencies in aptamer technology (…), but rather the general public ignorance about aptamer technology and the entrenched nature of antibody use (…). Another main reason (…) is the huge biotech industry's financial investment in the well-established antibody and humanized monoclonal antibody research**."

Aptamer technology will revolutionize therapeutics and diagnostics within several years, because aptamers are easy to produce, accessible to production modifications and regulation, cheap, infinitely reproducible, non-toxic and non-immunogenic (especially that our aptamer(s)

will not be delivered to the blood), highly selective, and immensely effective. Generation of an aptamer targeting the subdomain of SARS-CoV2's Spike protein will inhibit the virus attachment abilities and thus its replication in human host cells. Delivered in another cheap and simple device, pMDI, this novel COVID-19 drug (BABU-ONE™) has a great chance to rescue patients with mild to severe acute respiratory distress, help them recover from COVID-19 and alleviate most symptoms, and to act in prophylaxis, securing medical staff exposed to the virus every day and other essential employees as well as people living in the most endangered areas. This virus mutates fast, but not fast enough to undergo 71 successful punctual mutations resulting in 71 codon substitutions (an aptamer will target the 71-amino acid RBD subdomain). This will buy us time before a successful vaccine is developed.

VI. References

1. Gralinski, L. E., & Menachery, V. D. (2020). Return of the Coronavirus: 2019-nCoV. *Viruses*, *12*(2), 135.
2. Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, 102433.
3. Ren, L. L., Wang, Y. M., Wu, Z. Q., Xiang, Z. C., Guo, L., Xu, T., ... & Fan, G. H. (2020). Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese medical journal*.
4. Brann, D., Tsukahara, T., Weinreb, C., Logan, D. W., & Datta, S. R. (2020). Non-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients. *bioRxiv*.
5. Gu, J., Han, B., & Wang, J. (2020). COVID-19: Gastrointestinal manifestations and potential fecal-oral transmission. *Gastroenterology*.
6. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H., ... & Lessler, J. (2020). The incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and application. *medRxiv*.
7. Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet*, *395*(10228), 931-934.
8. Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine*.
9. Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *Jama*.
10. Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., ... & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. *Military Medical Research*, *7*(1), 1-10.
11. Zhang, W., Du, R. H., Li, B., Zheng, X. S., Yang, X. L., Hu, B., ... & Zhou, P. (2020). Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerging microbes & infections*, *9*(1), 386-389.
12. Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in different types of clinical specimens. *Jama*.

13. Kimball, A. (2020). Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility—King County, Washington, March 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*.

14. Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., ... & Guo, Q. (2020). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine*, *382*(12), 1177-1179.

15. Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. *Jama*.

16. Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... & Yu, T. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, *395*(10223), 507-513.

17. Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., ... & Li, X. (2020). A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. *New England Journal of Medicine*.

18. Dong, L., Hu, S., & Gao, J. (2020). Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discoveries & Therapeutics*, *14*(1), 58-60.

19. Poklepovic, A., & Gewirtz, D. A. (2014). Outcome of early clinical trials of the combination of hydroxychloroquine with chemotherapy in cancer. *Autophagy*, *10*(8), 1478-1480.

20. Gautret, P., Lagier, J. C., Parola, P., Meddeb, L., Mailhe, M., Doudier, B., ... & Honoré, S. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents*, 105949

21. Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., ... & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell research, 30(3), 269-271.

22. Bhavesh, N. S., & Patra, A. (2020). Virtual screening and molecular dynamics simulation suggest Valproic acid Co-A could bind to SARS-CoV2 RNA depended RNA polymerase.

23. Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., ... & Qin, C. (2020). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Research*, 1-13.

24. Wang, C., Li, W., Drabek, D., Okba, N. M., van Haperen, R., Osterhaus, A. D., ... & Bosch, B. J. (2020). A human monoclonal 1 antibody blocking SARS-CoV-2 infection. *Biorxiv*.

25. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Müller, M. A. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.

26. Milken Institute: https://milkeninstitute.org/sites/default/files/2020-03/Covid19-Tracker-3-36-20-FINAL.pdf

27. Peiris, J. S. M., Chu, C. M., Cheng, V. C. C., Chan, K. S., Hung, I. F. N., Poon, L. L., ... & Chan, K. H. (2003). Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *The Lancet*, *361*(9371), 1767-1772.

28. Zumla, A., Hui, D. S., & Perlman, S. (2015). Middle East respiratory syndrome. *The Lancet*, *386*(9997), 995-1007.

29. Dhama, K., Sharun, K., Tiwari, R., Sircar, S., Bhat, S., Malik, Y. S., ... & Rodriguez-Morales, A. J. (2020). Coronavirus disease 2019–COVID-19.

30. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., ... & Bi, Y. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565-574.

31. Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews Microbiology*, *17*(3), 181-192.

32. Shu, C., Huang, X., Brosius, J., & Deng, C. (2020). Exploring Potential Super Infection in SARS-CoV2 by Genome-Wide Analysis and Receptor–Ligand Docking.

33. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., ... & Sheng, J. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host & microbe*.

34. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260-1263.

35. Uddin, M., Mustafa, F., Rizvi, T. A., Loney, T., Al Suwaidi, H., Al Marzouqi, A., ... & Nowotny, N. (2020). SARS-CoV-2/COVID-19: Viral Genomics, Epidemiology, Vaccines, and Therapeutic Interventions.

36. Keefe, A. D., Pai, S., & Ellington, A. (2010). Aptamers as therapeutics. *Nature reviews Drug discovery*, *9*(7), 537-550.

37. Tombelli, S., Minunni, M., & Mascini, M. (2005). Analytical applications of aptamers. *Biosensors and Bioelectronics*, *20*(12), 2424-2434.
38. Nutiu, R., & Li, Y. (2003). Structure-switching signaling aptamers. *Journal of the American Chemical Society*, *125*(16), 4771-4778.
39. Krieg, A. M. (2008). Toll-like receptor 9 (TLR9) agonists in the treatment of cancer. *Oncogene*, *27*(2), 161-167.
40. Ng, E. W., Shima, D. T., Calias, P., Cunningham, E. T., Guyer, D. R., & Adamis, A. P. (2006). Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature reviews drug discovery*, *5*(2), 123-132.
41. Ramalingam, D., Duclair, S., Datta, S. A., Ellington, A., Rein, A., & Prasad, V. R. (2011). RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production. *Journal of virology*, *85*(1), 305-314.
42. Fukuda, K., Umehara, T., Sekiya, S., Kunio, K., Hasegawa, T., & Nishikawa, S. (2004). An RNA ligand inhibits hepatitis C virus NS3 protease and helicase activities. *Biochemical and biophysical research communications*, *325*(3), 670-675.
43. Binning, J. M., Wang, T., Luthra, P., Shabman, R. S., Borek, D. M., Liu, G., ... & Amarasinghe, G. K. (2013). Development of RNA aptamers targeting Ebola virus VP35. *Biochemistry*, *52*(47), 8406-8419.
44. Jang, K. J., Lee, N. R., Yeo, W. S., Jeong, Y. J., & Kim, D. E. (2008). Isolation of inhibitory RNA aptamers against severe acute respiratory syndrome (SARS) coronavirus NTPase/Helicase. *Biochemical and biophysical research communications*, *366*(3), 738-744.
45. Choi, S. K., Lee, C., Lee, K. S., Choe, S. Y., Mo, I. P., Seong, R. H., ... & Jeon, S. H. (2011). DNA aptamers against the receptor binding region of hemagglutinin prevent avian influenza viral infection. *Molecules and cells*, *32*(6), 527-533.
46. Kwon, H. M., Lee, K. H., Han, B. W., Han, M. R., Kim, D. H., & Kim, D. E. (2014). An RNA aptamer that specifically binds to the glycosylated hemagglutinin of avian influenza virus and suppresses viral infection in cells. *PloS one*, *9*(5).
47. Jeon, S. H., Kayhan, B., Ben-Yedidia, T., & Arnon, R. (2004). A DNA aptamer prevents influenza infection by blocking the receptor binding region of the viral hemagglutinin. *Journal of Biological Chemistry*, *279*(46), 48410-48419.

48. Cheng, C., Dong, J., Yao, L., Chen, A., Jia, R., Huan, L., ... & Zhang, Z. (2008). Potent inhibition of human influenza H5N1 virus by oligonucleotides derived by SELEX. *Biochemical and biophysical research communications*, *366*(3), 670-674.
49. https://viralzone.expasy.org/resources/Coronav/P0DTC2.txt
50. Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., ... & Sakata, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proceedings of the National Academy of Sciences*, *117*(13), 7001-7003.
51. https://cellbank.nibiohn.go.jp/english/
52. Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., ... & Niemeyer, D. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Journal of virology*, *85*(9), 4122-4134.
53. Ziegler, C., Allon, S. J., Nyquist, S. K., Mbano, I., Miao, V. N., Cao, Y., ... & Muus, C. SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues.
54. Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., & Nagata, N. (2019). TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. *Journal of virology*, *93*(6), e01815-18.
55. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015).
56. Wang, X., Dhindsa, R., Povysil, G., Zoghbi, A., Motelow, J., Hostyk, J., & Goldstein, D. (2020). Transcriptional Inhibition of Host Viral Entry Proteins as a Therapeutic Strategy for SARS-CoV-2.
57. Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., ... & Cui, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*.
58. Baig, A. M., Khaleeq, A., Ali, U., & Syeda, H. (2020). Evidence of the COVID-19 Virus Targeting the CNS: Tissue Distribution, Host–Virus Interaction, and Proposed Neurotropic Mechanisms. *ACS chemical neuroscience*.
59. Young, B. E., Ong, S. W. X., Kalimuddin, S., Low, J. G., Tan, S. Y., Loh, J., ... & Lau, S. K. (2020). Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *Jama*.
60. Shangguan, D., Li, Y., Tang, Z., Cao, Z. C., Chen, H. W., Mallikaratchy, P., ... & Tan, W. (2006). Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences*, *103*(32), 11838-11843.

61. Surugiu-Wärnmark, I., Wärnmark, A., Toresson, G., Gustafsson, J. Å., & Bülow, L. (2005). Selection of DNA aptamers against rat liver X receptors. *Biochemical and biophysical research communications*, *332*(2), 512-517.
62. Daniels, D. A., Chen, H., Hicke, B. J., Swiderek, K. M., & Gold, L. (2003). A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. *Proceedings of the National Academy of Sciences*, *100*(26), 15416-15421.
63. Svobodová, M., Pinto, A., Nadal, P. O. S. C. K., & O'Sullivan, C. K. (2012). Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Analytical and bioanalytical chemistry*, *404*(3), 835-842.
64. Gopinath, S. C. B. (2007). Methods developed for SELEX. *Analytical and bioanalytical chemistry*, *387*(1), 171-182.
65. Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2007). SELEX—a (r) evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular engineering*, *24*(4), 381-403.
66. Clark, M. A., Acharya, R. A., Arico-Muendel, C. C., Belyanskaya, S. L., Benjamin, D. R., Carlson, N. R., ... & Davie, C. P. (2009). Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature chemical biology*, *5*(9), 647-654.
67. https://www.addgene.org/vector-database/2853/
68. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406-3415.
69. Hünniger, T., Wessels, H., Fischer, C., Paschke-Kratzin, A., & Fischer, M. (2014). Just in time-selection: A rapid semiautomated SELEX of DNA aptamers using magnetic separation and BEAMing. *Analytical chemistry*, *86*(21), 10940-10947.
70. Cox, J. C., & Ellington, A. D. (2001). Automated selection of anti-protein aptamers. *Bioorganic & medicinal chemistry*, *9*(10), 2525-2531.
71. https://www.creative-diagnostics.com/news-sars-cov-2-immunoassay-kits-83.htm
72. Newman, S. P. (2005). Principles of metered-dose inhaler design. *Respiratory care*, *50*(9), 1177-1190.
73. Vinarov, Z., Katev, V., Radeva, D., Tcholakova, S., & Denkov, N. D. (2018). Micellar solubilization of poorly water-soluble drugs: effect of surfactant and solubilizate molecular structure. *Drug development and industrial pharmacy*, *44*(4), 677-686.
74. Roche, N., & Dekhuijzen, P. R. (2016). The evolution of pressurized metered-dose inhalers from early to modern devices. *Journal of aerosol medicine and pulmonary drug delivery*, *29*(4), 311-327.

75. Ibrahim, M., Verma, R., & Garcia-Contreras, L. (2015). Inhalation drug delivery devices: technology update. *Medical Devices (Auckland, NZ)*, *8*, 131.
76. G de la Torre, B., & Albericio, F. (2019). The pharmaceutical industry in 2018. An analysis of FDA drug approvals from the perspective of molecules. *Molecules*, *24*(4), 809.
77. Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *science*, *249*(4968), 505-510.
78. Kaur, H., Bruno, J. G., Kumar, A., & Sharma, T. K. (2018). Aptamers in the therapeutics and diagnostics pipelines. *Theranostics*, *8*(15), 4016.
79. Ponce, A. T., & Hong, K. L. (2019). A Mini-Review: Clinical Development and Potential of Aptamers for Thrombotic Events Treatment and Monitoring. *Biomedicines*, *7*(3), 55.
80. Rozenblum, G. T., Lopez, V. G., Vitullo, A. D., & Radrizzani, M. (2016). Aptamers: current challenges and future prospects. *Expert opinion on drug discovery*, *11*(2), 127-135.
81. Rusconi, C. P., Yeh, A., Lyerly, H. K., Lawson, J. H., & Sullenger, B. A. (2000). Blocking the initiation of coagulation by RNA aptamers to factor VIIa. *Thrombosis and haemostasis*, *84*(11), 841-848.
82. Eyetech Study Group. (2002). Preclinical and phase 1A clinical evaluation of an anti-VEGF pegylated aptamer (EYE001) for the treatment of exudative age-related macular degeneration. *Retina*, *22*(2), 143-152.
83. Ganson, N. J., Povsic, T. J., Sullenger, B. A., Alexander, J. H., Zelenkofske, S. L., Sailstad, J. M., ... & Hershfield, M. S. (2016). Pre-existing anti–polyethylene glycol antibody linked to first-exposure allergic reactions to pegnivacogin, a PEGylated RNA aptamer. *Journal of Allergy and Clinical Immunology*, *137*(5), 1610-1613.
84. Armstrong, J. K., Hempel, G., Koling, S., Chan, L. S., Fisher, T., Meiselman, H. J., & Garratty, G. (2007). Antibody against poly (ethylene glycol) adversely affects PEG-asparaginase therapy in acute lymphoblastic leukemia patients. *Cancer*, *110*(1), 103-111.
85. Farman, C. A., & Kornbrust, D. J. (2003). Oligodeoxynucleotide studies in primates: antisense and immune stimulatory indications. *Toxicologic pathology*, *31*(1_suppl), 119-122.
86. Marquis, J. K. & Grindel, J. M. Toxicological evaluation of oligonucleotide therapeutics. Curr. Opin. Mol. Ther. 2, 258–263 (2000).

VII. Supplementary Information.

Note: Citations not shown to minimalize the length of the proposal. Per request, all citations used to obtain the data below can be presented.

1. Genomic analysis of SARS-CoV, SARS-CoV2, and MERS.

SARS-CoV and SARS-CoV2 use the human ACE2 receptor in ciliated bronchial epithelial cells and type II pneumocytes. MERS-CoV enters cells via DPP4 (dipeptidyl peptidase 4) infecting unciliated bronchial epithelial cells and type II pneumocytes. Both SARS-CoV and MERS-CoV used amplifier hosts: market civets and dromedary camels, respectively; however, both viruses originated in bats.

The genomic difference between SARS-CoV and bat SARSr-CoV differ in three regions: ORF8, ORF3, and S. The S1 region is found to be less conserved and in bats it is produced in two clades: normal or deleted (residues 5, 12, and 13). Different SARS-CoV strains vary in binding affinity for human ACE2 and, consequently, in their infectivity, although residues 479 and 487 of the S1 RBD motif of SARS-CoV are conserved in all SARS-CoV strains and are responsible for animal-to-human transmission.

SARS-CoV2 shows greater similarity to SARS-CoV (79%) than to MERS-CoV (50%). Some of the major differences between SARS-CoV2 and SARS-CoV are the lack of 8a protein in the former, and truncated 8b and 3b proteins in the former. Surprisingly, SARS-CoV2 is closer to the SARS-like bat CoVs than to the human SARS-CoV (although, SARS-CoV2, SARS-CoV, MERS-CoV, and SARS-like bat CoVs are all *betacoronaviruses*. Genome similarity between SARS-CoV2 and the SARS-like CoVs detected in bats and Malayan pangolins is higher (0.85 to 0.96) than to SARS-CoV (0.80).

The advantage that SARS-CoV2 has is due to 138 non-synonymous substitutions (SNPs) found mostly in the ORF1ab (papain-like protease, helicase, RNA-dependent RNA polymerase) and Spike regions. Below are presented some of the identified non-synonymous substitutions in SARS-CoV2 (compared to SARS-like and SARS-CoVs):

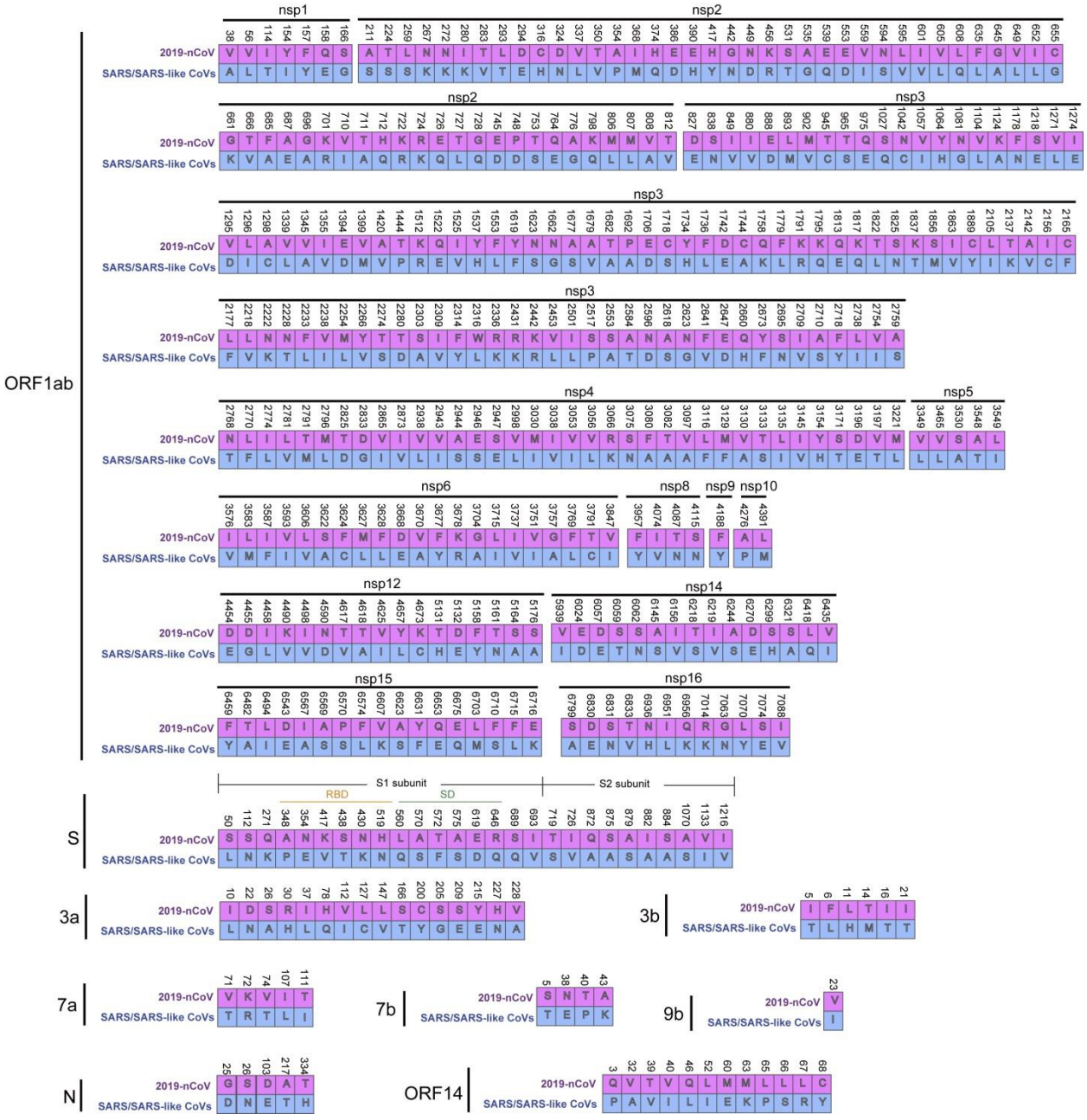

Figure 1. Genome substitutions in SARS-CoV2 (adapted from [33]).

A few more mutations in Spike S1 have been identified (not listed in the figure above), such as H36Y, N341D, D351Y, and V354F. The SARS-CoV2 mutant with G251V (in ORF3) and V354F (in Spike S1) has been shown to manifest a super infectivity. Because of all these non-synonymous substitutions in the ORF1ab and S regions, SARS-CoV2 underwent adaptive evolution **after** infecting

humans. We still lack data to determine the pace of genomic variations for SARS-CoV2, however, it is not as fast as some presume, but still the mutations that have already happened make the virus a "super" virus with a high R_0 (3.250 to 3.40).

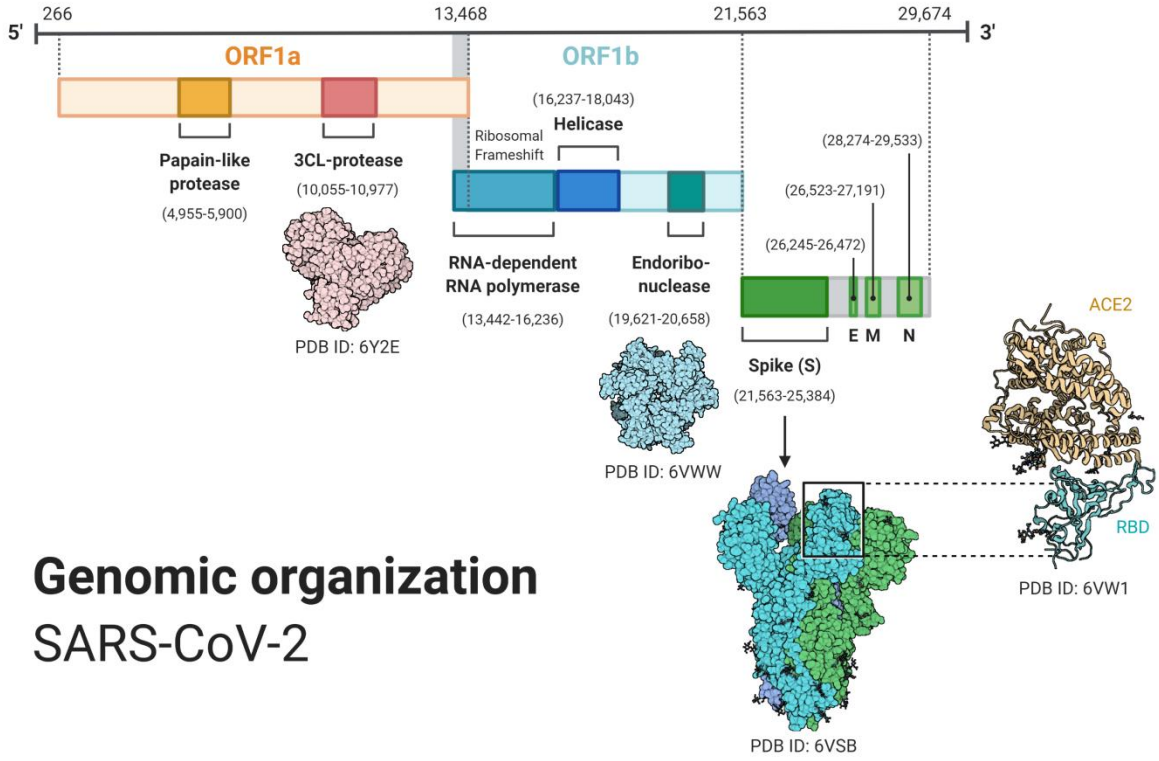

Figure 2. Genomic representation of SARS-CoV2.

2. Genomic analysis of SARS-CoV2.

SARS-CoV2 has 14 ORFs encoding 27 proteins. At the 5'-terminus the pp1ab and pp1a genes are located, comprising 15 non-structural proteins (nsps) (nsp1 to 10, 12 to 16). The 3'-terminus encodes 4 structural genes (S, E, M, N) and 8 accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14). Depending on a paper or database, SARS-CoV2 is built of ~ 29,674 nt having 10 CDS.

SARS-CoV	**V445**	**Y449**	**Y473**	**Q474**	**A475**	**E484**	**G485**	**F486**	**N487**
SARS-CoV2	R426	S432	T433	Y436	P462	D463	S472	N473	
Difference	19	17	40	38	13	21	13	13	

Two main amino acid residues in SARS-CoV's S protein are associated with its ACE2 recognition: K479 and S487, however, punctual mutations have occurred at these residues in SARS-CoV2 causing a shift of the two crucial residues: Q493 and N501. When comparing SARS-CoV2 and SARS-CoV, there is also a shift in the residues forming the capping loops.

In sum, punctual mutations in the ACE2 receptor-binding motif (changing the tropism of SARS-CoV2) and the changes in residues of the capping loops determine the binding specificity and infectivity of SARS-CoV2 in humans. These changes not only improve the interaction abilities with the human ACE2, but also improve the receptor recognition.

By March 18, referring to several papers, 27 amino acid substitutions were identified in SARS-CoV2: 6 substitutions in the RBD (319-541), 6 substitutions in the underpinning subdomain (SD, 569 – 655), 4 substitutions in the C-terminal of S1 subdomain (650, 570, 572, 575). When compared to SARS-CoV and SARS-CoV-like viruses, SARS-CoV2 has gained 380 amino acid substitutions; however, proteins nsp7, nsp13, E, and M, accessory proteins p6 and 8b have remained unchanged.

Of 101 SARS-CoV2 strains analyzed, 72 strains display a CT haplotype caused by an SNP in *ORF1ab* (T8517C, position in the genome: 8,782) known as the "L" haplotype. 29 strains exhibit a TC haplotype caused by an SNP in *ORF8* (S84L, position in the genome: 28,144) known as the "S" haplotype. Data analysis shows that the "S" type is ancestral, even though its contribution to the total pool of analyzed strains (~30%) is significantly lower than the "L" type (~70%), which was found to represent a higher transmission rate and more aggressive infection (worse clinical representation) because of accumulating evolutionarily advantageous mutations.

When analyzing the S protein (https://viralzone.expasy.org/resources/Coronav/P0DTC2.txt), I found that the S1 subdomain encompasses residues 13-685, the S2 subdomain encompasses 686-1273 residues. The RBD region is found in 319-541 (notice that it contains the residues 493 and 501) with the RBD ACE2 receptor-binding motif between 437 and 508. The cleavage sites are at 685-686 (cleavage site 1) and at 815-816 (cleavage site 2). 22 glycosylation sites have been found, but none of the 71 residues in the RBD ACE2 receptor-binding motif are glycosylated.

CDC of S protein (3,821 nt) accessed from GenBank (MT263395.1) is presented below:

```
ATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCAAT
TACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTTCAGATCCTCAGT
TTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTTCCAATGTTACTTGGTTCCATGCTATACATGTC
TCTGGGACCAATGGTACTAAGAGGTTTGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTT
CCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGACCCAGTCCCT
ACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTT
TTGGGTGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGA
ATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAA
AAATCTTAGGGAATTTGTGTTTAAGAATATTGATGGTTATTTTAAAATATATTCTAAGCACACGCCTATT
AATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAGATTTGCCAATAGGTATTA
ACATCACTAGGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGG
TTGGACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTAAAATATAAT
GAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTACGTTGA
AATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGT
TAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCATCTGTT
TATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCAT
TTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGC
AGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAGATTGCTGAT
TATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTA
AGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGA
TATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATGGTGTTGAAGGTTTTAATTGTTACTTT
CCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCATACAGAGTAGTAGTACTTT
CTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTGGTTAAAAACAA
ATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTG
CCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGA
TTCTTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAACAAATACTTCTAACCA
GGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACT
CCTACTTGGCGTGTTTATTCTACAGGTTCTAATGTTTTTCAAACACGTGCAGGCTGTTTAATAGGGGCTG
AACATGTCAACAACTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCA
GACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGT
GCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATACCCACAAATTTTACTATTAGTGTTACCA
CAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAAC
TGAATGCAGCAATCTTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAACCGTGCTTTAACTGGAATA
GCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTACAAAACACCACCAA
TTAAAGATTTTGGTGGTTTTAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATT
TATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGC
CTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCCTTACTGTTTTGCCACCTT
TGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGAC
CTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAATGGTATTGGA
GTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAA
TTCAAGACTCACTTTCTTCCACAGCAAGTGCACTTGGAAAACTTCAAGATGTGGTCAACCAAAATGCACA
AGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAATTTTGGTGCAATTTCAAGTGTTTTAAATGATATC
CTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATAGGTTGATCACAGGCAGACTTCAAAGTT
TGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAGCTTCTGCTAATCTTGCTGCTAC
```

```
TAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTATG
TCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACAAGAAAAGA
ACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTTC
AAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACA
TTTGTGTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTG
AATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACCAGATGTTGATTTAGG
TGACATCTCTGGCATTAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCC
AAGAATTTAAATGAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCAT
GGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATAGTAATGGTGACAATTATGCTTTGCTGTAT
GACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAAATTTGATGAACACGAC
TCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAA
```

The entire S protein sequence extracted from Expasy.org database and modified by the author is presented below:

```
 60    MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS
120    NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV
180    NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE
240    GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT
300    LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK
360    CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN
420    CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD
480    YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST EIYQAGSTPC
540    NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN
600    FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP
660    GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY
720    ECDIPIGAGI CASYQTQTNS PRRARSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI
780    SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE
840    VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC
900    LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM
960    QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN
1020   TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA
1080   SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA
1140   ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP
1200   LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL
1260   QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCCMTSC CSCLKGCCSC GSCCKFDEDD
1273   SEPVLKGVKL HYT
```

The red-marked sequence is the RBD region of the S1 subdomain. The blue-colored sequence is the RBD ACE2 receptor-binding motif, which contains the two key residues: Q and N (yellow-colored). With the help of the translation database and ORFs finder (https://web.expasy.org/translate/), it was manageable to trace the exact RNA sequence of the RBD region containing 695 nucleotides which can be used for vector cloning and recombinant S1 RBD expression:

```
ggaatctatcaaacttctaactttagagtccaaccaacagaatctattgttagatttcctaatattacaaacttgtg
ccctttttggtgaagtttttaacgccaccagatttgcatctgtttatgcttggaacaggaagagaatcagcaactgtg
ttgctgattattctgtcctatataattccgcatcattttccacttttaagtgttatggagtgtctcctactaaatta
aatgatctctgctttactaatgtctatgcagattcatttgtaattagaggtgatgaagtcagacaaatcgctccagg
gcaaactggaaagattgctgattataattataaattaccagatgattttacaggctgcgttatagcttggaattcta
acaatcttgattctaaggttggtggtaattataattacctgtatagattgtttaggaagtctaatctcaaacctttt
gagagagatatttcaactgaaatctatcaggccggtagcacaccttgtaatggtgttgaaggttttaattgttactt
tcctttacaatcatatggtttccaacccactaatggtgttggttaccaaccatacagagtagtagtactttcttttg
aacttctacatgcaccagcaactgtttgtggacctaaaaagtctactaatttggttaaaaacaaatgtgtcaatttc
aa
```

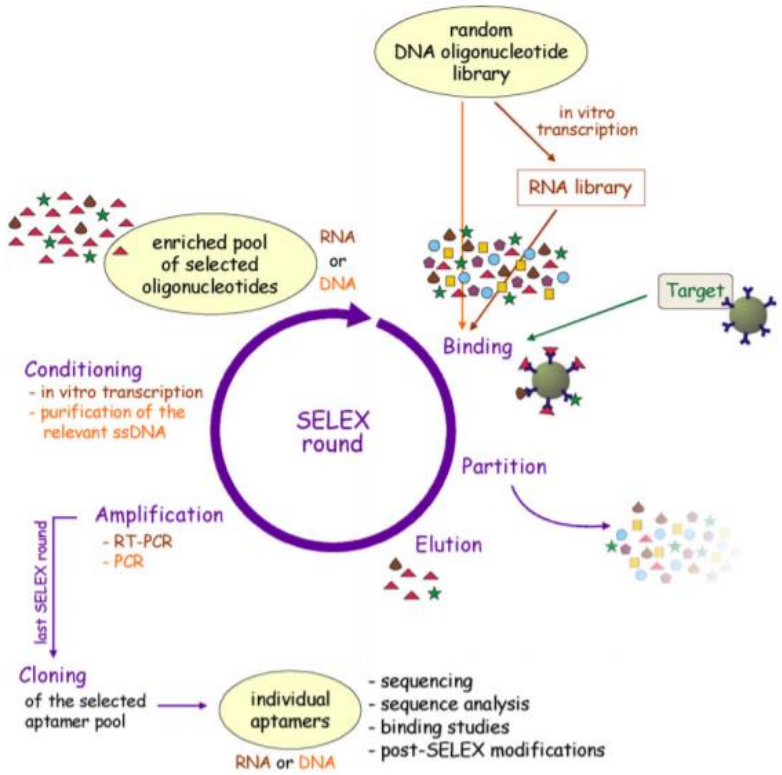

Figure 3. Genomic relatedness and host transmission of *betacoronaviruses* , including the COVID-19 SARS-CoV2 (adapted from [29]).

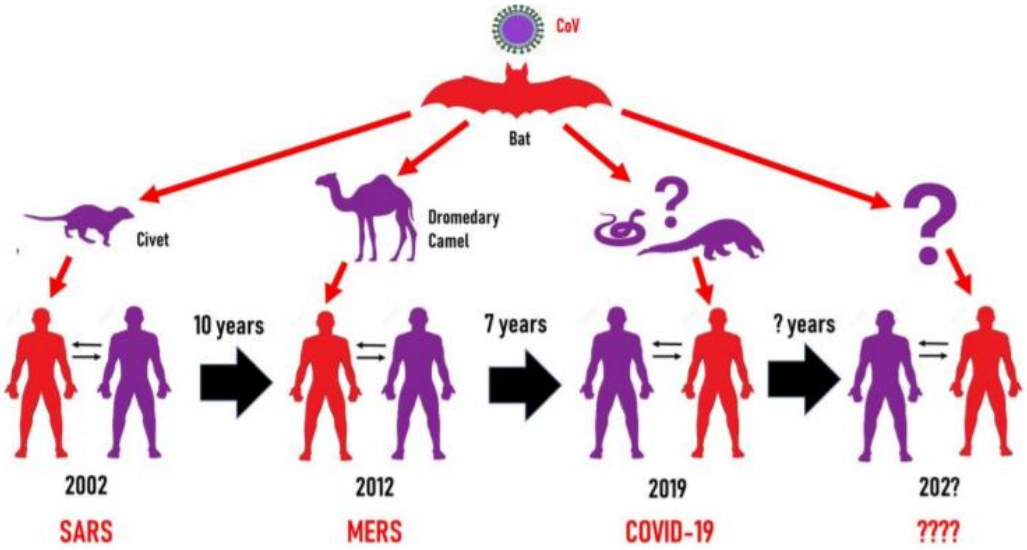

Figure 4. SELEX method (adapted from [64]).