Structure and Interaction-Based Design of Anti-SARS-CoV-2 Aptamers

Vladimir Mironov,*,[a, aa] Irina A. Shchugoreva,[b, c, d] Polina V. Artyschenko,[b, c, d] Dmitry Morozov,[e] Nicola Borbone,[f] Giorgia Olivier,[g] Tatiana N. Zamay,[d] Roman V. Moryachkov,[b, h] Olga S. Kolovskaya,[b, d] Kirill M. Lukyanenko,[b, d, i] Yanling Song,[j] Iuliia A. Merkuleva,[k] Vladimir N. Zabluda,[h] Georgy Peters,[l] Lyudmila S. Koroleva,[m] Dmitry V. Veprintsev,[b] Yury E. Glazyrin,[b, d] Ekaterina A. Volosnikova,[k] Svetlana V. Belenkaya,[k] Tatiana I. Esina,[k] Anastasiya A. Isaeva,[k] Valentina S. Nesmeyanova,[k] Daniil V. Shanshin,[k] Anna N. Berlina,[n] Nadezhda S. Komova,[n] Valery A. Svetlichnyi,[o] Vladimir N. Silnikov,[m] Dmitriy N. Shcherbakov,[k, m] Galina S. Zamay,[a] Sergey S. Zamay,[a] Tatjana Smolyarova,[b, k] Elena P. Tikhonova,[k] Kelvin H.-C. Chen,[k, n] U-Ser Jeng,[bb] Gerolama Condorelli,[cc, g] Vittorio de Franciscis,[u] Gerrit Groenhof,[e] Chaoyong Yang,[l, v] Alexander A. Moskovsky,[a] Dmitri G. Fedorov,[w] Felix N. Tomilin,[c, h] Weihong Tan,[v, x] Yuri Alexeev,[y, w] Maxim V. Berezovski,[y, z] and Anna S. Kichkailo,[b, d]

Abstract: Aptamer selection against novel infections is a complicated and time-consuming approach. Synergy can be achieved by using computational methods together with experimental procedures. This study aims to develop a reliable methodology for a rational aptamer in silico et vitro design. The new approach combines multiple steps: (1) Molecular design, based on screening in a DNA aptamer library and directed mutagenesis to fit the protein tertiary structure; (2) 3D molecular modeling of the target; (3) Molecular docking of an aptamer with the protein; (4) Molecular dynamics (MD) simulations of the complexes; (5) Quantum-mechanical (QM) evaluation of the interactions between aptamer and target with further analysis; (6) Experimental verification at each cycle for structure and binding affinity by using small-angle X-ray scattering, cytometry, and fluorescence polarization. By using a new iterative design procedure, structure- and interaction-based drug design (SIBDD), a highly specific aptamer to the receptor-binding domain of the SARS-CoV-2 spike protein, was developed and validated.

The SIBDD approach enhances speed of the high-affinity aptamers development from scratch, using a target protein structure. The method could be used to improve existing aptamers for stronger binding. This approach brings to an advanced level the development of novel affinity probes, functional nucleic acids. It offers a blueprint for the straightforward design of targeting molecules for new pathogen agents and emerging variants.

Introduction

The current COVID-19 pandemic has made it clear that virus outbreaks can still be an enormous threat to society despite advances in science and technology. While the challenges in dealing with the outbreak and its profound effect on our lives demand a coordinated multidisciplinary effort, the infection itself is a molecular process that can be understood by joint efforts of natural scientists. In particular, structural biology, organic synthesis, and computational chemistry can shed light on the molecular details of host recognition (the first step in the infection mechanism) and facilitate the development of efficient means to prevent that recognition and protect the population against COVID-19. COVID-19 is caused by a coronavirus, which is a positive-sense single-stranded RNA virus. Whereas coronaviruses mostly infect animals and cause only mild respiratory infections in humans, there have been notable exceptions, including the lethal severe acute respiratory syndrome (SARS)-CoV-1 endemic in 2003, the Middle East Respiratory Syndrome (MERS) endemic in 2012,[1] and now the COVID-19 pandemic.

The structural proteins of SARS viruses are the following: the spike (S) protein, the envelope (E) protein, the membrane (M), and the nucleocapsid (N) proteins.[2] The SARS-CoV-2 S-protein is important in virus infection because it mediates the viral entry into the host cells by binding to a host receptor via the receptor-binding domain (RBD). Therefore, the S-protein is considered the primary target for neutralizing antibodies, antiviral attachment, entry inhibitors, and vaccine development.[3]

For the SARS-CoV-2 S-protein pseudovirus system, it is known that: (I) human angiotensin-converting enzyme 2 (hACE2) is the receptor for SARS-CoV-2 binding,[20] (II) SARS-CoV-2 enters 293/hACE2 cells mainly through endocytosis, (III)
PIKfyve, TPC2, and cathepsin L are critical for the entry, and (iv) the S-protein in SARS-CoV-2 is less stable than in SARS-CoV-1.\[46\] The second co-receptor promoting SARS-CoV-2 entry into host cells is the tyrosine-protein kinase receptor UFO (AXL), which specifically interacts with the N-terminal domain of SARS-CoV-2 spike.\[50\]

Monoclonal and polyclonal antibodies, developed previously against the RBD in SARS-CoV, demonstrate only a limited binding to SARS-CoV-2.\[3,4,46\] Poor bio-availability, high cost, and a long time needed to develop and manufacture monoclonal antibodies, raises the necessity of a new methodology for rapid design of drugs, capable of binding selectively to a desirable epitope of a protein target with a known sequence.

On the other hand, oligonucleotides can also form multiple three-dimensional (3D) structures, some of which can fit a given protein target. Aptamers are artificial single-stranded RNA or DNA molecules consisting of 15–100 nucleotides. They can be designed to bind with high affinity and specificity to target proteins.\[7,8\] The mechanism of their interaction with the target is similar to that of antibodies. It can be characterized as hydrophobic, electrostatic, hydrogen bonding, and van der Waals interactions, as well as base stacking and shape complementarity.\[9\]

Aptamers have a high potential in diagnostics, therapeutics, and drug delivery systems.\[10–14\] However, they also could be used as antiviral drugs urgently needed during outbreaks of infections. Thus, efficient methods for rapidly screening sequences capable of blocking viruses binding to host cells should be developed beforehand.

Nowadays, aptamers are designed primarily through the SELEX procedure (Systematic Evolution of Ligands by Exponential Enrichment).\[7,8\] SELEX is used to generate target-specific aptamers,\[17\] whereas high-throughput sequencing technologies with the corresponding bioinformatics tools are used to deconvolute the results of SELEX pipelines.\[18\] Several SELEX variants have been suggested for aptamer optimization.\[19–23]\n
Recently, alternative approaches for designing aptamers that rely on \textit{in silico} analyses have been proposed.\[24–28\] A combination of docking and classical molecular dynamics (MD) simulations are often used to sample possible aptamer-target binding poses.\[29–33\] Several computational techniques were used to predict the binding energies and interactions in the aptamer-virus complexes: interaction analysis,\[34\] docking,\[35\] and MD-based free energy calculations.\[36\] These methods have been proven to be a crucial part of the most modern pipelines for drug design.

Here, we present a new methodology for aptamer’s rational design based on their structure and interaction with the target. This new methodology, called SIBDD (Structure and Interaction Based Drug Design), can be considered as an extension of SBDD (Structure-Based Drug Design),\[37\] enhanced with the feedback from classical molecular dynamics simulations combined with quantum-chemical interaction energies evaluation. SIBDD is a powerful combination of computational screening simulations and directed mutagenesis with experimental validation at each cycle to obtain oligonucleotides with a high affinity and selectivity.

SIBDD methodology is generally designed for cases with limited information about the target. For example, it can be used for rapid response to novel, dangerous infections when the access to the samples is minimal. While it can take several months to get the first experimental structure of the target, initial rounds of SIBDD can be performed based on computational folding results obtained from the target’s primary sequence. This initial set of aptamers could be further improved when more experimental data become available. Finally, the aptamers can be adjusted afterward for various pathogen mutations.

This work demonstrated how the SIBDD approach could be applied to produce an aptamer binding to the RBD of the SARS-CoV-2 spike protein. Several RBD-binding aptamers have been designed, and their affinity has been experimentally validated and confirmed.

The pipeline for aptamers selection

The preliminary step in the SIBDD methodology (Figure 1) is computational aptamer library design, which can be found in Github repository (https://github.com/MolecularDesign/Aptamer_library_16nt). The structures of aptamer candidates and the target (RBD of SARS-CoV-2 spike protein) are determined by molecular modeling. Next step is a selection of aptamers from the created library by molecular docking to target protein. The sequence with the best affinity predicted by docking is selected for the iterative improvement in the pipeline (Figure 1). The electrostatic potential of the protein is computed and used for the docking in each iteration of candidate structures to the RBD. Next, molecular dynamics (MD) simulations are performed on the aptamer-protein complexes constructed in docking. A clustering of the MD trajectories provides a set of the most populated molecular conformations in the dynamics. In these conformations, the aptamer-protein interactions are analyzed in detail with quantum mechanical (QM) calculations based on the fragmentation approach called fragment molecular orbitals (FMO) method. The fragmentation of proteins and aptamers into smaller parts, amino acid residues, and nucleotides, respectively, makes it possible to obtain a detailed map of nucleotide-residue interactions with QM calculations by using pair interaction electrostatic decomposition analysis (PIEDA). FMO structure optimization of the protein-aptamer complexes are performed with the DFTB method, while PIEDA calculations are performed at the DFTB and RI-MP2 levels of theory. The obtained interaction energies are not expected to be quantitatively precise but sufficiently indicate which nucleotides bind strongly and to which residues, making possible to use such information in the following steps of SIBDD. They can be viewed as a qualitative measure to guide the aptamer selection process.

After \textit{in silico} modeling, experimental validation of the most promising candidate structures binding affinities towards target are performed. The most effective aptamers will be advanced for the further selection rounds. As a result, the best binding aptamer is revealed by experimental and theoretical methods.
In this work, the synthesized aptamers were tested against recombinant RBD protein using flow cytometric analysis. Fluorescence polarization assay experiments demonstrated the binding kinetics of the aptamers to RBD in comparison with random sequences. In addition, small-angle X-ray scattering (SAXS) experiments verified whether an aptamer binds to the RBD with a 1:1 ratio and forms a stable complex.

After completing the whole procedure (Figure 1), a decision is made as to whether a desired level of binding is achieved. If it is not, then, based on the feedback from the stability of aptamer–protein complexes in MD and nucleotide-residue interactions, combined with the experimental validation, a new modified set of aptamers can be made and the whole procedure repeated.

The SIBDD protocol in this study was applied to a search for an aptamer with a high binding affinity towards the RBD of SARS-CoV-2 S-protein. Nevertheless, the proposed protocol is general and can be applied to a wider range of ligands and protein targets. It should be noted that the individual steps of the SIBDD approach are very well parallelizable, making it suitable for use within continuous high-performance workflows in the future.

Figure 1. Structure and Interaction Based Drug Design (SIBDD) is an iterative procedure consisting of the following steps: several rounds of aptamer molecular design (screening of a DNA library and directed mutagenesis for fitting with the protein tertiary structure); 3-D modeling of the protein target and calculating its electrostatic potential; molecular docking of aptamers to the target; molecular dynamics simulations of the complexes; quantum-mechanical analysis of nucleotide-residue interactions using methods such as DFTB, FMO, RI-MP2, and PIEDA; and finally experiment verification of binding.

Results and Discussion

Structure of the SARS-CoV-2 spike receptor-binding domain

RBD is a flexible region on the top of the coronavirus spike that permits virus anchoring to the human ACE2 receptor on the cell surface. The structure of the SARS-CoV-2 spike protein is shown in Figure 2a. This protein is a trimer, comprising three identical monomers (Figure 2b). Each monomer contains an RBD (Figure 2c) and can bind to human cells.
Aptamer design steps

Creating an aptamer library

In this study, aptamer molecular modeling is performed using a novel in silico selection approach without any preliminary experiments. In general, in silico aptamer modeling can be divided into two stages:\cite{25,42} (I) searching oligonucleotide sequences, which form the part of the aptamer responsible for selective binding, and (II) constructing a structure of the aptamer that maintains the optimal conformation of that binding part. Considering the binding interface of the RBD (Figure 2e.2) as the primary target for the aptamer, a hairpin was chosen as the initial shape of the aptamer for docking with the concave upper region of the RBD protein. To obtain the initial library of aptamer candidates, a 16mer sequence 5'-GGAATT NNNN AATTCC-3' was designed to have a central variable “NNNN” loop of four nucleotides and two constant self-complementary regions at the 5’ and 3’ ends (Figure 3a). All possible combinations of the DNA nucleotides for the variable region have been considered in this study, resulting in the initial aptamer library of 44 = 256 entries. The double-stranded stem part (5’-GGAATT- and -AATTCC-3’) improves the aptamer affinity to positively charged RBD parts and maintains the aptamer structure.

The molecular docking for all 256 models in the 16mer aptamer library (Figure 3a) was performed following the in silico selection scheme in Figure 1. For every aptamer, ten binding sites were predicted. The top side of the RBD is mostly exposed to the receptor (Figure 2e.2), and is presumed to be responsible for recognizing the ACE2 protein by the virus. The molecule variant that was predicted to have the most binding sites located in the upper part of the RBD (Figure 3c, Figure S1a) was selected. That aptamer, having the AGTC sequence in the

Figure 2. Structure and properties of SARS-CoV-2 spike protein RBD. (a) SARS-CoV-2 spike protein trimer, monomers are colored in blue, cyan, and green; (b) receptor-binding domain (RBD) position within the spike protein monomer; (c) active site of an RBD; and (d) an atomic model of an RBD, where C atoms are green, N atoms are blue, O atoms are red, S atoms are yellow, and H atoms are grey. Electrostatic potential of RBD is computed with APBS and plotted on a surface surrounding the protein: (e.1) inner view (i.e., facing the virus side), (e.2) top view, and (e.3) outer view (i.e., facing a host cell). Blue means positive potential value; red is negative. The total charge of RBD is +2; thus, the positive potential prevails, especially its side which is exposed to solution.
variable region (Figure 3b), is denoted as Apt16. It was used as the starting sequence for further modifications to improve the binding affinity to the key amino acid residues of the RBD reported by Lu et al.[40]

Simulations of Apt16-RBD complex

As mentioned above, Apt16 was selected as an initial sequence, and its binding sites with the protein were predicted by docking with RBD (Figure 3c, Figure S1a). An MD simulation of the RBD/Apt16 complex was performed to refine the binding pose of this complex further. These MD simulations were done with the GROMACS software package for at least 200 ns. After MD simulation, the conformers clustering showed that Apt16 mostly binds to the top part of RBD (Figures 3d and e).

From all clusters of conformations, the two with the highest populations in the MD simulations were chosen for further analysis. It was found that Apt16(1) (Figure 3d) mainly binds RBD with its 3'-ending nucleotides. For Apt16(2) conformation (Figure 3e), the nucleotides located both in the loop and in the stem part participate in the binding.

Next, a more detailed semi-quantitative analysis of residue-nucleotide interaction energies was performed. The MD structures of RBD/aptamer complexes were optimized at the semi-empirical density functional-based tight-binding (DFTB)[43] level. The FMO fragmentation approach was used for the optimized structures to get the interaction energies, which were evaluated with the non-empirical RI-MP2/6-31G(d,p) method. In all these simulations, the solvent environment was described by the polarizable continuum model (PCM). The procedure described above for Apt16 was also applied to the other aptamers.

The total interaction enthalpies, \( E_{\text{total}} \), between RBD and aptamer (Table 1) were obtained as a sum of residue-nucleotide pair interaction energies (PIE) at both DFTB and RI-MP2 levels of theory. PIE values allow us to compare the partial contribution to the binding energy from a single nucleotide in the aptamer and estimate the effect of point mutations. Thus, they can be used to systematically improve the aptamer sequence for better binding to the target molecule. It should be noted that because we did not attempt to account for the entropy contributions into PIE, \( E_{\text{total}} \) only includes binding enthalpies. However, entropy may play a significant role in the binding, thus making experimental checks important after each SIBDD cycle.

For example, for two different complexes RBD/Apt16(1) and RBD/Apt16(2), the values of \( E_{\text{total}} \) are \(-153.0 \text{ kcal mol}^{-1}\) and \(-91.6 \text{ kcal mol}^{-1}\), respectively, which indicates that RBD/Apt16(1) is more stable compared to RBD/Apt16 (2).

The color maps in Figure 3 d, e illustrate the pairs of nucleotides and amino acids whose contribution to the binding energy is the largest. The binding in the RBD/Apt16(1) complex is largely driven by G2, C15, T14, and C16 stem nucleotides, whereas in RBD/Apt16(2) the binding is mostly due to A7, G8, and T9 loop nucleotides. Participation of the duplex base pairs in the binding is not desirable because this region can be used to enhance an aptamer later with functional groups such as fluorescent labels. Thus, the strategy to improve the binding
affinity of Apt16 to RBD was to modify the aptamer to increase interactions involving the loop nucleotides rather than nucleotides of the duplex stem.

Iterative design of aptamers for binding with RBD

Experimental flow cytometry results (Figure 6a) showed for Apt16 a weaker binding than the random sequence of 40 nucleotides. Thus, to improve the aptamer binding affinity and specificity to the SARS-Cov-2 RBD domain, the initial sequence of Apt16 was modified. Based on the results from the MD simulations and quantum chemistry calculations discussed above, both the duplex and loop regions of the Apt16 aptamer were extended. First, to strengthen the complementary duplex structure, the stem regions were elongated from six to ten base pairs on both the 3' and 5' ends, resulting in the new hairpin-structure, the stem regions were elongated from six to ten base pairs on both the 3' and 5' ends, resulting in the new hairpin-forming template 5'-CCTAGGAATT-loop-AATTCCTAGG-3'.

Secondly, in RBD/Apt16 complexes, the interactions with the protein were mainly due to guanines and cytosines and, to a much lesser extent, to adenines. The reports\cite{44,45} indicate that the asparagine and the glutamate, exposed on the RBD surface, preferentially bind thymine to the other bases. We replaced the asparagine and the glutamate, exposed on the RBD surface, preferably to the sides of the RBD (Figures 4a.3 and a.5 and Figures 4b.3 and b.5, but Apt25 and Apt31 bind primarily to the top part of the RBD, as shown in Figures 3d and e and Figures 4b.3 and b.5, but Apt25 and Apt31 bind preferably to the sides of the RBD (Figures 4a.3 and a.5 and Figures 4c.3 and c.5) driven by the electrostatic interactions between the phosphate groups of the duplex part of the aptamer and the positively charged residues of RBD.

RBD/aptamer complexes: A comparative analysis

The binding of aptamers to RBD can be studied in terms of structure and interactions. For each RBD/aptamer complex, the aptamer remained bound to the RBD throughout MD simulations, adopting multiple binding sites. Apt16 and Apt27 bind primarily to the top part of the RBD, as shown in Figures 3d and e and Figures 4b.3 and b.5, but Apt25 and Apt31 bind preferably to the sides of the RBD (Figures 4a.3 and a.5 and Figures 4c.3 and c.5) driven by the electrostatic interactions between the phosphate groups of the duplex part of the aptamer and the positively charged residues of RBD.

Table 1. Total interaction energy $E_{\text{total}}$ (kcal mol$^{-1}$) and the number of hydrogen bonds $N_{\text{HB}}$ between aptamers and RBD. Energies were computed using FMO2/RI-MP2 method. FMO2/DFTB3 energies are provided for comparison.

| Aptamer$^{[a]}$ | $N_{\text{HB}}$ | $E_{\text{total}}$ (FMO2/RI-MP2) | $E_{\text{total}}$ (FMO2/DFTB3) |
|-----------------|----------------|-------------------------------|-------------------------------|
| Apt16(1)        | 12             | $-153.0$                       | $-100.9$                      |
| Apt16 (2)       | 9              | $-91.6$                        | $-65.1$                       |
| Apt25 (1)       | 11             | $-159.2$                       | $-107.8$                      |
| Apt25 (2)       | 12             | $-162.7$                       | $-112.9$                      |
| Apt27 (1)       | 14             | $-200.1$                       | $-132.8$                      |
| Apt27 (2)       | 17             | $-217.4$                       | $-157.9$                      |
| Apt31 (1)       | 13             | $-221.6$                       | $-167.9$                      |
| Apt31 (2)       | 9              | $-152.2$                       | $-108.3$                      |

$[a]$ Different conformer are labeled in parentheses.

Compared to RBD/Apt16, the number of strongly interacting residue-nucleotide pairs in RBD/Apt25 is larger, including interactions with thymines of Apt25 (the maps are shown in Figure 4a.2). The two most representative structures of RBD/Apt25 in MD simulations are shown in Figure 4 (a.3, a.5).

The total number of hydrogen bonds is almost the same as in the RBD/Apt16 complexes, but in the RBD/Apt25 clusters, they are formed mainly by T9, T11, T13, T15 thymines located both in the loop and in the stem parts. A detailed summary of the hydrogen bonds is given in SI (Table S1).

As mentioned above, the duplex function is to maintain a rigid hairpin conformation and provide electrostatic interactions with the protein. To further improve that, one of the A–T base pairs was replaced with a C–G pair to strengthen the duplex further, considering that C–G pairs are bound by three hydrogen bonds while A–T pairs only by two.\cite{46} Because the interaction data indicated that in Apt25 the T9 nucleotide of the duplex binds to RBD (Figure 4a.2, a.4), T9 was replaced with A9, and also the order of pairs in the central region of the duplex was changed to ensure that the specificity of the binding was due to the loop nucleotides rather than the stem. The new stem sequence was 5’-CGGATGGAATT – loop – ATTCCTACCG-3’.

On the other hand, the aptamer loop size was extended further to maximize its interaction surface with RBD and compensate for changes in the stem part. The loop was extended with an additional pair (in Apt27) and with two pairs (in Apt31) of the nucleotides. As a result, two new aptamers were created having the same duplex stem but a loop of different sizes: Apt27 (5’-stem – TTG TCTG – stem-3’) and Apt31 (5’-stem – ACTT TGTC TGTC – stem-3’). The structures of all discussed aptamers are shown in Figure 5. Aptamers Apt27 and Apt31 underwent the simulation rounds according to Figure 1. Comparative results of simulations for all aptamers, Apt16, Apt25, Apt27, and Apt31, are described in the following section.
tide interaction energies. The results are shown in Table 1, where more negative $E_{\text{total}}$ values correspond to a stronger attraction. Another important feature is the number of hydrogen bonds in the RBD/aptamer complexes (Table 1). The larger the number of hydrogen bonds, the higher is aptamer binding affinity to RBD. Detailed information regarding the hydrogen bonds is presented in Supporting Information (Table S1).

One can see from Table 1 that as a result of the aptamer design, the total interaction $E_{\text{total}}$ became stronger from $-91.6$ kcal mol$^{-1}$ for Apt16(2) to $-221.6$ kcal mol$^{-1}$ for Apt31(1). According to Table 1, the strongest interactions were observed for Apt27(2), showing a strong interaction of $-217.4$ kcal mol$^{-1}$. There are 17 hydrogen bonds between RBD and aptamer fragments in RBD/Apt27(2) and 13 hydrogen bonds in RBD/Apt31(1) complexes. The heat maps of interactions in Figures 4c.2 and 4b.2 show that in both the RBD/Apt31(1) and RBD/Apt27(2) complexes, some of the strongest interactions are due to thymines located in the loop. In the RBD/Apt27(2) complex, the largest contribution to the binding is due to pairs T12-Tyr505, T14-Lys417, and G17-Pro491, while in the RBD/Apt31(1) complex, the strongest interactions are observed for G14-Tyr369, G14-Ser373, and T15-Lys378 fragment pairs. In the next section, experimental studies on these aptamers are reported.

Experimental analyses of aptamer affinity to RBD

The experimental confirmation of aptamer binding to the spike RBD was performed by three different methods: flow cytometry, fluorescence polarization, and small-angle X-ray scattering.

![Figure 4](image-url)
Flow cytometry experiments

_In silico_ modeling predicts different binding affinities for the aptamers with RBD. To confirm this result, flow cytometry experiments of the binding of aptamer candidates with the recombinant purified RBD peptide were performed.

Ni beads were conjugated with recombinant RBD (RBD beads) and incubated with a designed aptamer or with a randomly controlled sequence. The fluorescence intensity of RBD-coated beads bound to the oligonucleotides was then measured on a flow cytometer. RBD of SARS-CoV-2 spike protein bound with Apt31 shows a stronger fluorescence signal against RBD than the initial library and the other three candidate sequences (Figures 6 a.1 and a.2).

As shown in Figure 6a, the fluorescence intensity induced by all aptamers, except for Apt16, is higher than that caused by a random sequence. Moreover, the binding level of each consequent aptamer developed during the _in silico_ evolution is higher than that of all others preceding it, validating the simulation-based design.

In good agreement with the interaction energies from QM calculations, the flow cytometry data confirmed that Apt31 has the strongest affinity for the RBD relative to a random DNA sequence. Furthermore, its induced fluorescence signal was comparable to that induced by RBD beads incubation with the _in vitro_ selected DNA aptamer CoV2- RBD-1C63 (Figure 6).

Experimental analyses of aptamer binding to recombinant RBD measured using fluorescence polarization assays

Flow cytometry experiments confirmed binding of Apt27 and Apt31 with the recombinant RBD of the spike protein fixed on the Ni beads. To ensure that the two best candidates can bind the non-fixed RBD, an additional method, fluorescence polarization assay was used. It is a highly sensitive method capable of detecting polarization changes of fluorescent molecules due to their binding with other molecules floating in the solution.

This experiment used synthetic FAM-labeled Apt27 and Apt31 and two random control sequences of 20 and 40 nucleotides in length to estimate their binding to the recombi-
nant RBD peptide. RBD was produced based on the gene sequence encoding the 308 V–542 N region of the SARS-CoV-2 spike protein (GenBank MN908947). A Chinese hamster ovary (CHO) cell expression system was used to harvest this peptide. The RBD has post-translational modifications similar to the native viral RBD protein; therefore, it is a useful model for testing the binding ability of in silico generated candidate aptamers.

Binding kinetics measurements were performed on a Clariostar Plus microplate reader (BMG LABTECH, Germany); 10 nM of FAM-labeled aptamers Apt27, Apt31, or 10 nM control random sequences R20 (20-nucleotide long) and R40 (40-
nucleotide long) were titrated with increasing concentrations of RBD peptide.

As shown in Figure 6b, the fluorescence polarization of aptamer Apt31 increased from 200 to 250 mP after the first RBD injection, to 290 mP after the second injection, and to 330 mP after the third injection. Further injections did not increase fluorescence polarization, thus indicating a binding saturation (Figure 6b). The fluorescence polarization of aptamer Apt27 slightly increased after adding the RBD peptide. Conversely, the R20 and R40 control sequences did not significantly bind to RBD, thus confirming the enhancement of the aptamer candidate’s binding abilities during the in silico selection.

Experimental analyses of aptamer binding to recombinant RBD by small-angle X-ray scattering

Using the SAXS intensity plots against the scattering angle, one can estimate the molecule sizes and form factor in the solution. The SAXS curve slope shows features of molecular shape (Figure 6c). The inverse Fourier transformation of SAXS curves reveals the pair-distance distribution in the real space of a molecule in solution.

According to the SAXS data, the RBD protein has a more folded structure and larger size than the aptamer (Dmax = 8.3 nm; the radius of gyration is 2.48 nm). Apt31 has a prolate shape with a 2 nm thickness, 6.0 ± 0.2 nm length, and radius of gyration Rg = 1.77 ± 0.07 nm. The volume of the particle is estimated as Vp = 33.91 nm³, which corresponds to the molecular weight MWSAXS = 8.95 (7.95–9.95) kDa. It has good accordance with the expected molecular weight of the aptamer MWexp = 9.52 kDa. The bead model was used for the molecular design of the aptamer model, and finally, the theoretical SAXS data were compared to the experimental data (Figure 5a, b). The obtained discrepancy between molecular model and SAXS data χ² = 1.298 shows a good coincidence in the spatial structure restoration. SAXS spectrum of an equimolar mixture of the aptamer and the RBD protein was obtained. We observed the additional peaks with respect to individual RBD or aptamer SAXS spectra corresponding to a complex with an approximate size of 15 nm (Figure 6c). The overall electron density volume corresponds to the molecular weight in the range of 40–52 kDa. The masses of aptamer and protein were estimated to be in ranges of 10.8–13.1 and 20.8–33.5 kDa, respectively (with 95% credibility). The SAXS dataset for Apt31 has deposited at the SASDBDB[4] (https://www.sasbdb.org/draft-preview/3576/1oi-q9akb5r/), accession code: SASDMU7.

Conclusions

DNA aptamers are promising ligands for the detection of SARS-CoV-2[48,49] and blocking its viral activity[50] because of their high specificity, ease of modification, and use. In this work, a novel methodology, structure- and interaction-based drug design (SIBDD), has been presented and applied to design a new 31mer DNA aptamer targeting with high affinity the RBD domain of the SARS-CoV-2 spike protein.

The selection procedure in SIBDD combines (1) of in silico library of aptamer structures with (2) electrostatic potential calculations to guide the docking of aptamers to RBD, (3) molecular dynamics simulations and (4) quantum chemistry-based analysis of the binding energies. The qualitative insights into the residue-nucleotide interactions obtained in this procedure are used to guide the rational modification process of the aptamers in vitro. The validity of the SIBDD approach has been validated with (1) flow cytometry, (2) fluorescence polarization, and (3) small-angle X-ray scattering measurements. The experimental results indicated the aptamer Apt31 as the most effective, fully supporting the SIBDD in silico prediction.

To the best of our knowledge, this is the first report of a systematic and rational in silico design of a DNA aptamer. The binding efficacy of Apt31 to the RBD has been proven using three different experimental methods. Further experimental studies are underway to assess the antiviral properties of Apt31.

Two routes can be envisaged to take advantage of the new aptamer: (1) therapeutic applications to prevent the virus from binding to human cells, and (2) diagnostic usage to detect the virus presence. Preliminary results demonstrate that the developed aptamers are promising candidates for detecting and blocking the SARS-CoV-2 virus. Furthermore, the SIBDD protocol described in this work can facilitate in silico the development of diagnostic and therapeutic ligands for SARS family viruses.

Given the high versatility of aptamers, the computational design of RBD-binding molecules offers a promising blueprint for the development of diagnostic and therapeutic tools for other transmissible diseases.

Methods

Molecular modeling

Molecular dynamics simulations

The X-ray structure of the complex of the SARS-CoV-2 RBD protein with human enzyme ACE2 (PDB ID: 6M0J) was used as a reference.[51] The remaining sugar residues were removed from all glycosylation sites. Protonation states of histidines and hydrogen atoms positions were assigned using the GROMACS pdb2gmx module.

The protein model was then submerged into a water box 15 nm × 15 nm × 15 nm. The negative charge of the protein complex system was neutralized with Na+ ions. Then, Na+ and Cl− ions at the physiological concentration (0.15 M) were added to the system.

MD simulations were performed by using GROMACS 2019.8 software.[52] The Amber14sb[53] force field for proteins and the TIP3P model for water[54] were used. NPT MD trajectories with 200 ns length were generated at the 310 K temperature using a V-rescale thermostat with 1 ps time constant[55] and at 1 bar.
pressure using a Parrinello-Rahman barostat with 5 ps time constant and 4.5 × 10⁻³ bar⁻¹ compressibility.

The lengths of all bonds involving hydrogen atoms were constrained to their equilibrium values, taken from the force field parameters, with the LINCS⁵⁷ method. The SETTLE algorithm was used to constrain the internal degrees of water molecules.⁵⁸ This allowed using a 2 femtoseconds time step for trajectory integration. The smooth particle mesh Ewald method⁵⁹ was used to account for periodic electrostatics with a real-space cutoff of 1.2 nm. Lennard-Jones non-bonded interactions were calculated using a cutoff of 1.2 nm with a smooth switching starting at 1.0 nm.

Molecular electrostatic potential

The structures of SARS-CoV-2 spike RBD in water for computing the potential were obtained from MD simulations by taking snapshots every 5 ns, resulting in about 70 structures. The protein atoms were taken from these snapshots and aligned by minimizing RMSD between backbone heavy-atom positions. These structures were prepared for the molecular electrostatic potential calculation using the PDB2PQR program⁶⁰ and the PARSE⁶¹,⁶² continuum electrostatics force field.

The linearized Poisson-Boltzmann equations were solved by using a multigrid approach implemented in APBS.⁶³ The Na⁺ and Cl⁻ ionic concentration of 0.15 M was used to simulate physiological conditions. The computed molecular electrostatic potential (MEP) values on a 3-D grid were then averaged over using a multigrid approach implemented in APBS. The computed molecular electrostatic potential (MEP) values on a 3-D grid were then averaged over using a multigrid approach implemented in APBS.

Clustering of MD trajectories

The most representative structures of aptamer/RBD complexes were obtained by the clustering analysis of MD simulations. Before clustering, the trajectories were aligned by the positions of the Cα atoms of the RBD.

Clusters of structures and their centers were computed using the quality threshold algorithm⁶⁵ implemented in VMD. The RMSD of the phosphorus atoms was used as a metric function; the cutoff was 0.7 nm. At most, five clusters were extracted for each aptamer.

Structure preparation of aptamer-RBD complexes

The RBD model was taken from the MD simulations described above. The three-dimensional models of aptamers were built using the Avogadro program. Aptamer-RBD complexes were obtained by docking using HDOCK, a web server for protein-DNA/RNA docking. For each complex, the top 10 models according to the HDOCK score were collected for further analysis.

Quantum-mechanical calculations

For accelerating QM calculations and obtaining pairwise interactions, the fragment molecular orbital (FMO) method⁶⁶ was used. In the FMO calculations, each residue within the RBD protein was represented by a fragment (189 fragments total); the fragmentation was performed using the Facio⁶⁷ program. Apt16 was divided into 16 nucleotide fragments, and other aptamers were likewise fragmented into nucleotides. Fragment calculations were performed in the fully polarizable embedding, and by calculating pairs of fragments, inter fragment interactions were incorporated, such as charge transfer, exchange-repulsion, electrostatics, dispersion, and solvent screening.⁶⁸ FMO was validated in various protein-ligand binding studies, including complexes of SARS-CoV-2.⁶⁹–⁷²

The geometry optimization of both aptamers and RBD/aptamer complexes was carried out using third-order density-functional tight-binding (DFTB3) combined with the two-body FMO²¹⁹ using 3ob parameters,⁷⁰ D3(BJ) empirical dispersion,⁷¹ and the conductor-like polarizable continuum model of solvation (C-PCM).⁷² All calculations were performed with GAMESS(US) program.⁷³

Individual pair interaction energies (PIEs) can also be computed by using FMO. In the RBD-aptamer systems, they represent residue-nucleotide interaction energies. By summing the values of all pairs, the total value (E_{total}) was obtained. The PIE and total interactions can be computed at a higher level of theory than the level used for geometry optimizations. In this work, we used two methods for PIE energy calculation. First one was the DFTB3 method used for structure calculation. The second one is RI-MP2 method with the 6-31G(d,p) basis set together with the auxiliary rimp2-cc-pVdz basis set⁷⁴ for the resolution of identity. Both methods were combined with the C-PCM model for solvent. The choice of basis sets is a compromise between speed of calculation on the one hand, and accuracy of interaction energies on the other. We note that for the purpose of informing aptamer design, these interaction energies need to be qualitatively correct (i.e. trends), rather than quantitatively and small errors are tolerable. See Supporting Information Tables S2 and S3 for the estimates of BSSE errors associated with the chosen method.

Experimental details

Oligonucleotide synthesis

Oligonucleotide synthesis for fluorescence polarization assay was performed on an ASM-800 DNA/RNA synthesizer (Biosset, Russia) on a 0.1 μM scale. Oligonucleotides were synthesized using a standard phosphoramidite 0.05 M solution in CH₃CN (ChemGenes Inc., USA).

The fluorescent dye 5(6) carbonylfluorescein was attached at the 5'-end of the oligonucleotide, which was deprotected and removed from the solid support using concentrated NH₄(aq). According to standard manufacturer protocols, purification was accomplished by using a reverse-phase cartridge (ChemGenes Inc., USA).
Synthesis and purification of recombinant RBD

Plasmid construction was performed based on the gene sequence encoding the SARS-CoV-2 Spike protein (GenBank MN908947). The codon composition of the sequence was optimized by using GeneOptimizer. The resulting nucleotide was synthesized as part of the pGH vector (lltd. DNA-Synthesis, Russia).

The sequence corresponding to the region 308 V-542 N was used to obtain RBD. The signal sequence of tissue plasminogen activator was inserted by PCR using the following primers: Tp-RBD- F5’-GACGCCTATGGTGCAATG-3’ and Tp-RS-5’-CAGAGCCACA CAGCA- GAGCC CTCTCTTAT CGATCCATG GTGCCCGCG GGC TACGCTA TAGTGAG-3’ on the pLPP vector matrix with the use of primers Tp-RBD- F5’-TGCTGTGGTG TGCTGCTGTG TGGAAGCTGTC TTCGTTTTCC GCGTGGAAAA GGGCAGTCTA CAGC-3’ and RBD4-RS- AAAAAAGTCC ACGAGGCTGA TCAGCCGGTT AAAC-3’ on the pGH-S matrix. PCR products were annealed and inserted into the vector pVEX2 AhlII and SalI to allow 6 × His to be included in the protein. The CHO cell expression system was used to develop the SARS-CoV-2 Spike protein trimer and RBD. Using Lipofectamine 3000 (ThermoFisher, USA) per the manufacturer’s protocol, the CHO-K1 cell line was transfected with pVEX2-RBD or pVEX2-RS vectors and plasmid pCMV (CAT) T7-SB100. The latter encodes transposase SB100, providing integration of expression cassettes into the mammalian host genome.

After three days, the antibiotic puromycin (InvivoGen, USA) was added to the culture medium at a final concentration of 10 μg/mL to select resistant cell clones that had received the resistance gene to this antibiotic from the pVEAL2 vector. The selection was carried out for three days; then, the polyclonal cell culture was plated into a 96-well plate at the final concentration of one cell per well and incubated for two weeks under selective pressure. The wells were examined for the presence of colonies, and cultural fluids were screened for productivity. Clones with the highest productivity were used for protein isolation. Protein purity was determined by SDS-PAGE and visualized by Coomassie blue staining and Western blotting using rabbit anti-RBD antibodies (Sino Biological, China).

According to the manufacturer’s protocol, the recombinant RBD was purified by metal chelate chromatography on a Ni NTA column (Qiagen, Germany). Additional purification was carried out by chromatography on columns connected in series with cation exchange (SP-Sepharose) and anion exchange (Q-Sepharose) sorbents, equilibrated with 20 mM Tris-HCl, pH 8.2. After the protein was applied, the columns were washed with 20 mM Tris-HCl, pH 8.2. The columns with the SP-sepharose sorbent, on which RBD is not adsorbed under these conditions (pH 8.2), was disconnected; and the proteins bound to the Q-sepharose sorbent were dialyzed against PBS and subjected to sterilizing filtration through filters with a pore size of 0.22 μm. The resulting protein was purified by anion exchange column (Q-Sepharose) and size exclusion column (Sephacryl S-200) with a flow rate of 1 mL/min, fractions were collected, and purity was determined by SDS-PAGE. The reduced RBD was verified by mass spectrometry and amino acid analysis.

Fluorescence polarization analyses of aptamer binding to RBD

The ClarioStar Plus microplate reader (BMG LABTECH, Germany) was used for fluorescence polarization measurements. The optical system had the following settings: excitation (482 ± 16 nm), dichroic filter LP 504 nm, emission (530 ± 40 nm), and target polarization (P) value 200 mP (mP° = 1000). Signal time integration was 3 s for each measurement. The measurements were performed at 37 °C.

Aptamers were titrated with the protein as follows: 150 μL of aptamers Apt27 and Apt31 or control random sequences R20 (20 nucleotides long) and R40 10 nM (40 nucleotides long) at a concentration of 10 nM in phosphate-buffered saline were placed into wells of flat-bottomed black 96 well plates (Grenier, Germany) in 14 replicates. The RBD protein was injected into the wells and vortexed automatically by the instrument ClarioStar Plus microplate reader. The final concentration of the protein after each ejection was 7 μM, 14 μM, 17 μM, 20 μM, and 24 μM. The kinetics of the fluorescence polarization changes was measured. Aptamer and RBD synthesis were described above.

RBD/Apt31 complex preparation for SAXS

First, 117 μL of binding buffer (PBS, pH 7.4, including 136.8 mM NaCl, 10.1 mM NaHPO4, 2.7 mM KCl, 1.8 mM K2HPO4, 0.55 mM MgCl2) was added to 58 μL of RBD (5 mg/mL) stock solution to prepare a 5 mg/mL protein concentration. Then 1 μL of aptamer Apt31 solution (745 nM) was mixed with 48 μL binding buffer, pH 7.4 (stock solution of Apt31). The ClarioStar Plus microplate reader was used for fluorescence polarization. The optical system had the following settings: excitation (482 ± 16 nm), dichroic filter LP 504 nm, emission (530 ± 40 nm), and target polarization (P) value 200 mP (mP° = 1000). Signal time integration was 3 s for each measurement. The measurements were performed at 37 °C.

Aptamers were titrated with the protein as follows: 150 μL of aptamers Apt27 and Apt31 or control random sequences R20 (20 nucleotides long) and R40 10 nM (40 nucleotides long) at a concentration of 10 nM in phosphate-buffered saline were placed into wells of flat-bottomed black 96 well plates (Grenier, Germany) in 14 replicates. The RBD protein was injected into the wells and vortexed automatically by the instrument ClarioStar Plus microplate reader. The final concentration of the protein after each ejection was 7 μM, 14 μM, 17 μM, 20 μM, and 24 μM. The kinetics of the fluorescence polarization changes was measured. Aptamer and RBD synthesis were described above.

SAXS measurements and data processing

SAXS data from RBD, Apt31, and RBD/Apt31 complex samples were obtained at the BioMUR beamline at the NRC “Kurchatov Institute” [79] The wavelength on the beamline was 0.14 nm, which corresponds to energy 8.58 keV. The sample to detector distance was 700 mm. The SAXS data were recorded by a Dectris Pilatus 1 M detector during the X-ray 300s exposures of each image. The temperature of each sample during the exposition was set to 20 °C. These images were radially averaged relatively to the center of the beam. The aptamer and the protein were measured at the concentrations of 8, 4 and 2 mg/mL. For the RBD/Apt31 complex the concentrations were 5, 2.5 and 1.25 mg/mL. Dilution of the samples allowed us to define the most appropriate concentration
which gives sufficient SAXS intensity with a good signal-to-noise ratio and does not already show multiple scattering effects.

The SAXS patterns were obtained in the range of scattering vectors from 0.1 to 4.5 nm\(^{-1}\) \((s = 4\sin(\theta)/\lambda\), where 2\(\theta\) is the scattering angle). The SAXS pattern from the buffer was also recorded and subtracted from the solution signal to get the SAXS data from the sample molecules only.

Since the signal-to-noise ratio of the recorded SAXS curves on wide angles and the set of concentrations used for the experiment were not optimal, only the scattering vector range 0.2 < s < 1.4 nm\(^{-1}\) was used. Data obtained from SAXS measurements at the concentrations 4 mg/ml for the aptamer and protein separately and 2.5 mg/ml for the complex yielded the monodispersity of the solution for each sample, which allows the calculation of the structure parameters for the aptamer, protein, and their complex.

To reconstruct the spatial structure of the Apt31 aptamer, additional size-exclusion-chromatography (SEC) SAXS measurements were performed at the TPS 13A BioSAXS beamline at NSRRC, Taiwan.\(^{(1)}\) The sample was prepared in Tris buffer solution and exposed by X-rays of 15 keV at 22°C with an online HPLC system. The initial concentration was 13.4 mg/ml; HPLC column provided a high dilution of the sample during the measurements. The X-Ray beam size at the sample position was 320 × 260 μm². Sample to detector distance was 1210.06 mm. Total 128 data frames were recorded by the detector Eiger X 9 M, each 2 s exposure time per frame. All of the frames recorded with a constant Φg value over the HPLC (high-performance liquid chromatography) sample flow were radially averaged into the one-dimensional scattering intensities, two of them were merged into resulting SAXS curve. SAXS measurements, data acquisition, and processing were performed according to standard pipelines.\(^{(28-30)}\)

The reduction and analysis of experimental SAXS data (buffer subtraction, averaging, merging, gyration radius calculation, Guinier and Kratky analyses, pair-distance distribution function \(p(r)\) calculation, bead modeling, molecular model validation, MD and bead model alignment) were made using the program suite ATSAS.\(^{(8)}\) The bead model was constructed using the program DAMMIN\(^{(46)}\) according to the calculated \(p(r)\) function. Evaluation of the molecular model was performed by a simulated annealing process using the program CRYSOL.\(^{(40)}\)

**Author Contributions**

V.M., I.A.Shch., P.V.A., D.M., D.G.F., Y.A., F.N.T., A.A.M., G.G., N.B., G.O., G.C., V.deF., T.N.Z., and A.S.K. performed the computational design of the aptamers. L.S.K. and V.N.S. synthesized oligonucleotides. I.A.M., D.N.Sh., E.A.V., S.V.B., T.I.E., A.A.I., V.S.N., and D.V.Sh. synthesized and purified proteins. Y.A.’s work at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. The work of D.M. and G.G. has been done as part of the BioExcel CoE (www.bioexcel.eu), a project funded by the European Union contracts H2020-INFRAEDI-02-2018-823830 and H2020-EINFRA-2015-1-675728. D.M. and G.G. also thank the CSC-IT center in Espoo, Finland, as well as PRACE for awarding access to resource Curie-Rome based in France at GENCI. V.M. thanks Russian Foundation for Basic Research (project number 19-03-00043). A.B.’s and N.K.’s work was supported by the Ministry of Science and Higher Education of Russian Federation (state assignment of the Research Center of Biotechnology RAS). V. deF. G.C., N.B. and G.O. are grateful to FISR2020_00177 Shield, Italian Ministry of Education and Research, for funding. GC is grateful to the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement: CONCreTE 872391; PRISAR2 872860. Use of the 13 A BioSAXS beamtime at the Taiwan Photon Source is acknowledged. The work of M.V.B was funded by the Canadian Institutes of Health Research grant OVI1-170353. SAXS measurements and PIEDA analyses were funded by the Russian Science Foundation (project N° 21-73-20240 for A.S.K.).

**Acknowledgements**

The authors are grateful to JCSS Joint Super Computer Center of the Russian Academy of Sciences – Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences” for providing supercomputers for computer simulations. The authors thank the RSC Group (www.rscgroup.ru) and personally Mr. Oleg Gorbachev for the constant support and establishment of “The Good Hope Net Project” (www.thegoodhope.net) multifunctional non-profit anti-CoVID research project. The authors also thank the Helicon Company (www.helicon.ru) and personally Olesya Kucenko, Alexander Kolobov, Leonid Klimov for instrumental support and help with conducting fluorescence polarization assays, which were performed on a demo instrument Claristor Plus microplate reader (BMG LABTECH, Germany). We thank Dr. Yong-Zhen Zhang for providing the genome sequence of 2019-nCoV and Dr. Xinquan Wang for providing the crystal structure of the binding domain of the SARS-2 Spike protein. The authors are grateful to Aptamerlab LCC financial support (www.aptamerlab.com). Y.A.’s work at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. The work of D.M. and G.G. has been done as part of the BioExcel CoE (www.bioexcel.eu), a project funded by the European Union contracts H2020-INFRAEDI-02-2018-823830 and H2020-EINFRA-2015-1-675728. D.M. and G.G. also thank the CSC-IT center in Espoo, Finland, as well as PRACE for awarding access to resource Curie-Rome based in France at GENCI. V.M. thanks Russian Foundation for Basic Research (project number 19-03-00043). A.B.’s and N.K.’s work was supported by the Ministry of Science and Higher Education of Russian Federation (state assignment of the Research Center of Biotechnology RAS). V. deF. G.C., N.B. and G.O. are grateful to FISR2020_00177 Shield, Italian Ministry of Education and Research, for funding. GC is grateful to the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement: CONCreTE 872391; PRISAR2 872860. Use of the 13 A BioSAXS beamtime at the Taiwan Photon Source is acknowledged. The work of M.V.B was funded by the Canadian Institutes of Health Research grant OVI1-170353. SAXS measurements and PIEDA analyses were funded by the Russian Science Foundation (project N° 21-73-20240 for A.S.K.).

**Conflict of Interest**

The authors declare no conflict of interest.

**Data Availability Statement**

The data that support the findings of this study are openly available in github at https://github.com/MolecularDesign/Aptamer_library_16nt, reference number 37.
Keywords: aptamers · fragment molecular orbitals method · molecular dynamics · SARS-CoV-2 · SAXS
