DNA aptamers block the receptor binding domain at the spike protein of SARS-CoV-2

Fabrizio Cleri¹,*, Marc F. Lensink² and Ralf Blossey²

¹University of Lille, CNRS UMR8520 IEMN, Institut d’Electronique, Microélectronique et Nanotechnologie, and Department of Physics, F-59652 Villeneuve d’Ascq, France, ²University of Lille, CNRS UMR8576 UGSF, Unité de Glycobiologie Structurelle et Fonctionnelle, F-59655 Villeneuve d’Ascq, France

Abstract

DNA aptamers are versatile molecular species obtained by the folding of short single-stranded nucleotide sequences, with highly specific recognition capabilities against proteins. Here we test the ability of selected DNA aptamers in interacting with the spike (S-)protein of the SARS-CoV-2 viral capsid. The S-protein, a trimer made up of several subdomains, develops the crucial function of recognizing the ACE2 receptors on the surface of human cells, and subsequent fusioning of the virus membrane with the host cell membrane. In order to do this, the S1 domain of one protomer switches between a closed conformation, in which the binding site is inaccessible to the cell receptors, and an open conformation, in which ACE2 can bind, thereby initiating the entry process of the viral genetic material in the host cell. Here we show by means of state-of-the-art molecular simulations that small DNA aptamers can recognize the S-protein of SARS-CoV-2. Moreover, their interaction with different regions of the S-protein can effectively block, or at least considerably slow down the opening process of the S1 domain, thereby largely reducing the probability of virus-cell binding. We also provide evidence that binding of the human ACE2 receptor may be drastically affected under such conditions. Given the facility and low cost of fabrication of specific aptamers, the present findings could open the way to both an innovative viral screening technique with sub-nanomolar sensitivity, and to an effective and low impact curative strategy.

*To whom correspondence should be addressed. Tel: +33 320 197928; Fax: +33 320 197884; Email: fabrizio.cleri@univ-lille.fr
1 Introduction

At the end of 2019, a novel virus belonging to the coronavirus family has been identified, initially in the population of the Chinese city of Wuhan. Since then, the virus has spread covering practically the whole world, requiring drastic measures both for treatment of the patients and to avoid uncontrolled spreading of the disease among the human population. This virus has been designated SARS-CoV-2 by the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses. Coronaviruses are enveloped viruses, their protein capsid being decorated by club-shaped glycoprotein spikes (S-protein) that protrude from the surface, as it is the case of, e.g., SARS and MERS viruses. However, this novel coronavirus is still distinct from both SARS and MERS, with multiple mutations identified in different genomic regions. The surface-covering S-proteins allow the virus to bind to certain receptors on human cells, such as the widely distributed ACE2. Like other members of the same family, SARS-CoV-2 carries a positive-sense, single-stranded RNA genome belonging to the Coronaviridae family with about 70% similarity in genetic sequence to SARS. The characteristic structure of its S-protein is made up of three protomers, each including two key domains, S1 and S2. S1 with its receptor-binding domain (RBD) is required for host-cell receptor binding, and S2 is required for membrane fusion. Because of its steric prominence, the S-protein is one of the main targets for both molecular-based therapy and screening of the virus.

No effective therapies for this virus have been found yet. A significant number of potential drug targets have been identified and are actively being tested in large cohort studies, such as remdesivir and favipiravir. Other drugs are repurposed from previous diseases, such as the anti-malaria chloroquine and hydroxychloroquine. In the absence of a proven therapy, adjunctive therapies such as corticosteroids, anticytokine or immunomodulatory agents, and immunoglobulin therapy are also under consideration.

Current anti-viral screening methods mostly analyse throat and nose swab samples with RT-PCR, which uses nucleic acids as target, and IgM / IgG biomarkers. The diagnostic accuracy of RT-PCR highly depends on the "virus-specific diagnostic window" and the analytical sensitivity of this assay is potentially plagued by false SARS-CoV-2 negativity, attributable to the low viral loads especially in asymptomatic or mildly symptomatic patients. Despite the large acceptance of these assays in clinical settings, they are expensive and time consuming. On the other hand, serologic tests are based on recognition of antibodies; however, IgM have little specificity since they are active for about any kind of viral infection that may have attacked the organism, and the more specific IgG arise only several weeks after the infection thus being of little help for the early detection. Given the highly transmissible nature of this virus, and its relatively high fatality rate, there is urgent need for highly specific, early-stage and selective testing, massively available and at the lowest possible cost.

Aptamers are artificial oligonucleotide or peptide molecules that bind to a target molecule with high specificity. Aptamer-protein-based analytical methods have become popular in the last years. Just like antibodies, aptamers are capable of binding a target, and also of modulating or blocking its activity. Generated by an in-vitro selection process from pools of random sequence oligonucleotides (the SELEX technique), targeted aptamers have already been produced for hundreds of different protein targets. A typical aptamer is 10-30 kDa in size (about 30-60 nucleotides), it binds its target with sub-nanomolar affinity and, most importantly, can discriminate against closely related targets. Structural studies indicate that aptamers are capable of using the same types of binding interactions that drive affinity and specificity in antibody-antigen complexes. Aptamers of various type have been already identified and tested in the anti-viral domain in recent years. For example, Chen et al. found that 5 pg/µL of their ssDNA aptamer could effectively stop replication of H5N1 avian-influenza virus; Jang et al. demonstrated an efficient SARS-helicase activity inhibition by a RNA aptamer. Recently, Song et al. identified two candidate ssDNA aptamers that seem to bind efficiently to the RBD of the S-protein of SARS-CoV-2; in another recent study, DNA aptamers were shown to be able to efficiently recognize the SARS-CoV-2 nucleocapsid protein.

In the present work, we investigate by means of state-of-the-art protein docking and large-scale molecular dynamics simulations, the interaction of some experimentally identified DNA aptamers with the S-protein of SARS-CoV-2. Our initial purpose was to measure the affinity of the aptamer for the binding domain of the S-protein, in support of the use of aptamers as fast and efficient anti-viral screening. However, an even more interesting question concerns the detailed molecular interaction between aptamers and the viral protein(s). Indeed, it could be possible that the same aptamers may block, or at least considerably slow down the transition of the S1 domain from the closed to the open conformation, thereby blocking the access of the cell surface receptors to the virus surface. In this work we focus on this key aspect, showing that the DNA aptamers, while binding very efficiently to the designated RBD on one protomer of the S-protein, also form and maintain stable bonds with the S2 domains of adjacent protomers. This extended bonding creates a sort of "bridge", which results in hampering the opening of the RBD to the cell receptors. Extensive MD simulations on two different experimentally identified aptamers, allowed to characterize in different docking configurations the nature and strength of the interactions, mainly hydrogen bonds, complemented by long-range Van der Waals interactions. Umbrella sampling of the configurations going from closed to open RBD, with and without the DNA aptamer attached, allowed to characterize the large variations in free-energy barriers, which permitted to set a size on a relative scale of the announced blocking effect. Simulation of force-driven docking of human ACE2 receptor to the S1-DNA complex demonstrate that the receptor binding site is strongly affected by the presence of the DNA...
aptamer, and leads to a drastic reduction of the binding efficiency. Once such predictions are experimentally validated, DNA aptamers could represent an alternative, low-cost and low-impact therapy, apt to reduce the virus efficacy in the host organism. Virtual screening of DNA aptamers by computer simulation could, moreover, quickly adapt to rapidly mutating viral targets, as well as to new Coronavirus-family strains that could appear in the future.

2 MATERIALS AND METHODS

2.1 Molecular structure of the S-protein and ACE2

The PDB entries 6VXX and 6VYB from the RCSB Data Bank\(^3\) were used as reference for the SARS-CoV-2 S-protein, in the closed and open forms, respectively (Figure 1). The S-protein is a homologous trimer, with each protomer being composed by the two domains S1 and S2, and a structural domain. The structures were passed through the pdb2gmx utility of the GROMACS package, to assign hydrogens to the residues. For the thermal equilibration simulations, the protonation state of histidines was automatically selected based on the closest possible hydrogen bonds; for the umbrella sampling simulations instead we had to impose a unique choice to all frames (see below), in order to maintain the same protein structure, therefore we arbitrarily imposed single protonation at the ND nitrogen.

The ACE2 human receptor molecular configuration was taken from the 6M0J entry of the RCSB\(^4\). Although ACE2 is observed to dimerize in-vitro\(^5\), however the interaction with the S-protein is likely to occur via only one monomer through one S-protein protomer, since microscopic observation of the whole S-protein never found instances with more than one protomer in the open configuration\(^4\), which is necessary for ACE2 binding. Therefore, the monomeric structure of ACE2 was retained for our study.

2.2 Molecular structure of candidate DNA aptamers

We took the sequences of the two candidate ssDNA aptamers from the recently published study cited above, by Song et al.\(^9\). These were extracted by a SELEX procedure of 12 rounds over a pool of several millions random sequences directed against the RBD fragment of the S-protein. After reduction of redundant fragments, the two best sequences were obtained (see Table S1). For each sequence, we firstly took the three directions, containing about 380,000 TIP3P water molecules, as target. A large number of docked structures with very close energies were produced by HADDOCK. The best five configurations for the apta1, labelled 8-1; and, for the sake of comparison, two-best configurations for the apta2, labelled 10-1 and 12-1 respectively (Figure 2), with very close binding affinity but rather different conformation.

It may be noted that the 3D conformations of the aptamers are deduced based on a two-step process, in which the secondary structure is firstly minimized on the basis of the simple nearest-neighbor interaction model, and then fed into a 3D model building program: as such, there is no guarantee that the lowest-energy structures selected in the first step would remain at the lowest energy also in the second step, which implies a substantial contribution of elastic energy, long-range and dihedral interactions. Secondly, the stereochemical docking of the aptamer to the protein domains is also subject to a considerable uncertainty, as different methods and codes are known to give somewhat different results. For both these issues, the substantial convergence of the results obtained on the two different configurations of apta2 (see below) constitutes a minimal proof of internal consistency.

2.3 Molecular dynamics simulations

The AMBER99 force field database\(^15,16\) and its recent BSC1 extension to improve the treatment of nucleic acids\(^17\) were used for the molecular bonded and non-bonded force parameters.

For the molecular dynamics (MD) simulations we used the GROMACS 2020 computer code package\(^18,19\). For the thermal stability study, the ensemble of the complete S-protein and DNA aptamers were solvated in a water box of size \(23 \times 23 \times 23\) nm\(^3\) with periodic boundary conditions in the three directions, containing about 380,000 TIP3P water molecules, plus Na\(^+\), Cl\(^-\) and Mg\(^{2+}\) ions to ensure neutralization of the phosphate backbone charge, and a physiological concentration of 0.1 M NaCl and 0.005 M MgCl\(_2\). Similar conditions were used for the umbrella sampling study, but with a smaller water box of \(14 \times 14 \times 18\) nm\(^3\). All the production MD runs were carried out at the temperature of 310
Figure 1. Ribbon model of the S-protein in the closed (a) and open (b) configurations. The S1 domain of each monomer is depicted in yellow, and the S2 in orange. In (b) also the ACE2 human receptor is represented (blue), interacting with the one S1 domain in open configuration (see red arrow; atomic structure obtained by aligning the pdb 6VXX, with the co-crystallized S-protein RBD and ACE2 structure, pdb 6M0J).

K and pressure of 1 atm. The low-mass, N-bonded glycans present in the experimental protein structures were removed for the MD simulations, since they were only hand-built into the cryo-EM density map structures, and adjusted by computer post-processing. Such a choice is further justified, since these sugars appear to be close to the docked aptamers but not directly interacting. Given the wide variability of in vivo glycans, for which the available information is currently evolving, we judged appropriate to focus this first study on clean protein conformations; in further studies, realistic glycan configurations could be added.

Coulomb forces were summed with particle-mesh Ewald sum, using a real-space cutoff of 1.2 nm (equal to the cut-off radius of shifted Van der Waals potentials). We used rigid bonds for the water molecules, with a time step of 2 fs for the thermal equilibration phases and 1 fs for production and force-pulling runs. For the thermal stability study, preparatory runs at constant-\(\{NPT\}\) and temperatures increasing in steps of 100K from \(T=10\) to \(T=310K\) lasted 20 ns, and were followed by thermal stability simulations at constant-\(\{NVT\}\), which extended to 100-200 ns for each configuration. Statistics were accumulated over the last part (50-100 ns) of each trajectory.

For the umbrella sampling and potential of mean force (PMF) study, we could not use one of the many free-energy sampling methods to obtain the lowest-energy path, because even the most refined sampling methods available are ineffective for a molecular system of this size and complexity. Therefore, we reconstructed a putative opening path from the closed (6VXX) to the open (6VYB) conformations of the S1-S2 domains of the S-protein, by using the morph utility of the Chimera package. 50 intermediate frames were obtained along the shortest geometric path, at distances of 0.25 Å along this fictitious reaction coordinate, and the corresponding configurations were reconstructed (note that such a spacing is one order of magnitude smaller than usually assumed in PMF calculations, to avoid potential problems of sampling homogeneity from the artificial reconstruction procedure). For the sake of working with homogeneous configurations, it was necessary to fill in the structural gaps between the two experimental sequences, since some residues were missing on either side, and several residues in the 6VYB were incomplete. Then, the 50 conformations of the S1-S2 domains were geometrically realigned on the reference 6VXX structure, by the TMalign utility program, thereby obtaining
50 complete configurations of the S-protein, each with one single monomer transitioning from closed to open. After this "cold" reconstruction process, the 50 configurations were run through the pdb2gmx GROMACS utility and solvated in ionized TIP3P water (see above), in such a way to obtain strictly the same atom-ordered structures, with the same number of water molecules and ions, in order to represent the putative result of a MD trajectory along the closed-to-open transition. These 50 configurations were firstly equilibrated, from 10K to 310K in steps of 5 ns, and subsequently used in the umbrella sampling, with short (10ns) force-constrained runs, to extract the potential of mean force along the putative opening pathway. The final extraction of the free-energy profiles by weighted-histogram analysis was done with the GROMACS wham utility. The same protocol was repeated for all the three aptamer docking configurations, by aligning on the reference 6VXX main structure the ensemble of the S1 and S2 subdomains carrying the docked DNA. For each new set of 50 frames, the whole procedure of thermal equilibration and force-constrained runs was repeated, and the potential of mean force was obtained.

Overall, the study used about 0.5 million hours of CPU time, on 960-1280 Intel CascadeLake cores + 96 NVIDIA V100 GPUs of the IDRIS Jean-Zay supercomputer in Orsay, and on 504-1008 Intel Broadwell cores of the OCCIGEN supercomputer in Montpellier, with typical running times of about 5 ns/hour of wall-clock time. About 0.3 Terabyte of raw data were accumulated for subsequent post-processing. All-atom microsecond trajectories for the complex (DNA+proteins) stored in the compact GROMACS-xtc format at frame intervals of 50 ps, are freely available upon request to the authors.

3 RESULTS

3.1 Binding of DNA aptamers to the S-protein

The results of 100-200 ns long MD trajectories for the three aptamer configurations interacting with the S-protein trimer demonstrate a very stable bonding of each aptamer to the S1 domain of one single monomer of the whole protein. We extracted representative structures from the MD trajectories by the clustering algorithm of GROMACS. By looking at the centroid structures that collect most of the statistics (between 30 and 40% of the total trajectory), we observe that the smaller aptamer apta1 makes 10 hydrogen bonds with the S1 domain, whereas the longer apta2 makes about 11-12 strong hydrogen bonds in both configurations, plus a number of lighter and fluctuating bonds. Figure 3 (a–c) shows the H-bonding configurations for each of the 3 complexes, by representing with atomic spheres the interacting residues from the protein (cyan) and the DNA aptamers (red). The two different bonding configurations of apta2 share a subset of 6 H-bonds with the protein, mostly arginine residues R346, R357 and R466 plus K356 and N450, while the other 5 or 6 interactions are different in either case. The 10-1 configuration has the DNA contacting the RBD in a more central region, while the 12-1 configuration is somewhat shifted to contact with one half of the structure, the other half remaining "above" the RBD.

Importantly, however, we also find that the DNA aptamer docked at the S1 domain of one of the monomers of the CoV-2 spike protein, also starts interacting with the S2 domain of adjacent monomers. While for the smaller apta1 such an interaction is limited only to exchange of long range forces (VdW and electrostatic) with a few flanking residues from a nearby protein monomer, the longer apta2 is able to make new hydrogen bonds with the S2 domain of a different monomer, adjacent to the one to which it is primarily attached. We observe (see Fig. 3 (d–e), grey and red atomic spheres for...
Figure 3. Schematic of the hydrogen bonds formed by the DNA aptamers (red) interacting with the S-protein trimer in the closed conformation. (a–c) Binding of the apta1 (configuration 8-1) and apta2 (in the two different configurations, 10-1 and 12-1) to the S1 domain of one (cyan) monomer. H-bonded residues are depicted with atomic spheres, cyan for the protein and red for the DNA. (d–e) Extra hydrogen bonds formed by the apta2 with the S2 domain of a different (grey) monomer, in the two configurations.

Molecular contact surfaces were calculated with the PDB2PISA web utility, by using the standard rolling-sphere method with 1.4 Å probe radius. The apta2-S complex has 15.8 nm² of contact surface with the S1 domain in the first configuration (10-1), and 18.6 nm² in the second one (12-1); the complex with the shorter apta1 has a correspondingly smaller contact surface of 10.7 nm². PISA also computes the solvation free energy, by taking the difference between the isolated and interfaced atomic structures of the different fragments; such a value can be taken as a first estimate of the interfacial adhesion between the aptamer and the protein, however noting that the as-calculated value does not include the H-bonds energy. We thus have a $\Delta G = -24 \pm 1$ kcal/mol for both the apta2 configurations, and $\Delta G = -14 \pm 1$ kcal/mol for the apta1. Furthermore, the extra H-bonding interaction...
of apta2 with the S2 monomers adds 3.2 and 17.5 nm² of contact surface, respectively for the 10-1 and 12-1, with a corresponding extra contribution to the solvation energy of $\Delta G = -8.9$ and $-13.4$ kcal/mol.

By comparison, the interaction of the ACE2 human cell receptor with the same S1 domain (whose binding mode is quite similar to that of the SARS-CoV RBD, which also targets ACE2) is localized at a few residues that make up the RBD. In the case of ACE2, interaction occurs via N- and C-terminals of the $\alpha$-helix plus a small area of the $\alpha$2-helix, over a contact (buried) surface of 8.4 nm², the experimental dissociation constant obtained by surface-plasmon resonance is about 4.7 nM (values ranging from 1.2 to 14.7 nm, using different techniques were also reported), which translates into a binding free energy $\Delta G = -11.5$ kcal/mol at room temperature. It appears therefore that DNA aptamers could bond more strongly than ACE2 to the viral S-protein, and directly interfere in the receptor binding process.

It is worth noting that both the DNA aptamers used in the present study appear to contact the S1 domain in regions adjacent to the ACE2 small binding area, and likely could interfere with the ACE2-RBD interaction. The strong bonding interaction of DNA aptamers with the S1 (and S2) domain of the S-protein leads to severe mechanical deformations of the latter: many elements of the protein are destructured from helix and sheet to a disordered coil, and lead to a much more loose contact at the RBD region (see below).

### 3.2 Free-energy of opening of the S1 domain

The umbrella sampling study allowed us to obtain the free-energies and the kinetic barriers for the S1 domain going from the closed to the open configuration. As detailed in the Methods section above, we defined a putative reaction coordinate $\zeta$ along the shortest path connecting the two extreme experimental configurations, and traced the potential of mean force (PMF). The reaction coordinate is normalized in [0, 1], corresponding to a physical motion of about 1.2 nm of the center of mass of domain S1 with respect to the center of mass of S2. The free energy difference between the open and closed conformations is in all cases estimated by taking the difference between the minimum and the maximum of the PMF all along the $\zeta$ coordinate. Figure 4 reports the plot of the PMF for the free S-protein, and for the three docked aptamer configurations. It is observed that for the free protein (full curve), the transition from close to open is practically barrierless with an energy difference $\Delta G$ of just about $4.5 k_BT$, and proceeds at constant energy from $\zeta \approx 0.35$ to 1. Such findings confirm the experimental observation that the S1 domain can rather freely fluctuate between the two conformations, at physiological temperatures. However, the detailed reason why only one monomer at each time fluctuates in the open position, leaving the other two in the closed position, is yet to be understood.

When a DNA aptamer is docked to the S1 domain, some important energetic changes indeed arise. The shorter apta1 that, as described above, has only a relatively weak interaction with neighboring domains of the S-protein; however, it appears to affect significantly the opening kinetics (dash-dotted curve): the $\Delta G$ is unchanged with respect to the free protein, but a substantial modification of the PMF profile appears. The opening occurs at nearly constant energy up to $\zeta = 0.4$, followed by two successive plateaux of about $2k_BT$ each to arrive at the fully opened conformation.

The situation changes with the longer apta2 docked to the S-protein. In the 10-1 configuration, the $\Delta G$ is increased to 6.7$k_BT$; the opening follows a nearly linear growth of free energy, a shallow secondary minimum appears around $\zeta = 0.65$, followed by a second nearly linear ramp to the final conformation. Even more drastically, in the 12-1 configuration the $\Delta G$ jumps to 10.8 $k_BT$, thus signifying a relative reduction of the opening probability by about a factor $10^{-3}$ (ratio of the $\Delta G$ exponentials); the opening trajectory follows a nearly steady linear ramp, with a mild change of slope around $\zeta \sim 0.2–0.6$; a sharp minimum appears right before the final opening (however, such a feature could also be due to the numerical noise that affects the extremes at $\zeta \sim 0$ and $\sim 1$ of all PMF plots, because of the reduced overlap of the sampling windows).

Despite some known limitations in interpreting PMF results, a steady slope in the PMF vs. $\zeta$ plot gives an indication of the force needed to move from one conformation to another of the system. The slopes of about 5 and 10 $k_BT$/nm observed for the two docking configurations of apta2, indicate an extra resistance to spontaneous switching of the S1

![Figure 4. Plot of the potential of mean force extracted along the opening pathway $\zeta$, for the free S-protein (full line), and the S-protein with one DNA aptamer docked (dashed lines). The free energy is in all cases estimated by the difference between the maximum and minimum value of the PMF along the reaction coordinate.](image)
domain from closed to open (with corresponding forces in the range 20-40 pN) once the DNA aptamer is docked. It is also worth noting that such force-displacement curves could readily be subject to direct experimental testing by means of single-molecule force spectroscopy methods.

### 3.3 On the binding of the ACE2 receptor to S1 domain

It is now interesting to look at the possible interaction of the ACE2 receptor with the S1 domain, in such a partly-open conformation modified by the presence of the DNA aptamer.

In its native conformation, ACE2 makes a large number of H-bonds at the RBD residues 498-501 with the $\alpha_1$-helix, plus bonds at 417, 453 and 474, according to the study by Yan et al.\(^5\); similarly, H-bonds at residues 455, 487, 493 and 501 are reported by Lan et al.\(^6\); further, weaker interactions (salt bridges, VdW) are also observed at some other residues in the range 440 to 505 of S1. A recent, detailed theoretical study indicated the key role of hydrophobic interfaces and charge complementarity in establishing the interaction of ACE2 with the RBD\(^29\). Note that the whole S1 subdomain covers residues about 330 to 515 (see below, Figure 6(a)).

After binding the DNA apta2 (lowest energy 10-1), the adhesion capability of ACE2 to the open conformation of the S-protein is largely reduced. In a first attempt, we contacted the ACE2 receptor to the RBD of the S-protein with apta2, taken in the final stage of the opening pathway, by rigidly shifting the coordinates of ACE2 according to the experimental superposition. Due to the deformation of the RBD, the contact surface area decreases from 8.4 to 6.2 nm\(^2\); bonding is also much affected, the number of H-bonds being reduced from 10 to 5, after losing contact between the S1 loop and the C-terminal of the $\alpha_1$-helix; the solvation free energy estimated by the PDBePISA method goes from $-4.5$ kcal/mol to 0. However, the most notable information that comes from this rigid-shift superposition is that the ACE2 sterically conflicts with the DNA aptamer over a large region, therefore the structure of this ACE2-S-DNA complex must necessarily be modified upon the mutual interaction.

Thus, in a second step, we performed force-driven docking: the ACE2 structure was rigidly shifted about 1 nm away from the contact, and pushed back into the S-protein complete trimer under a constant force directed along the vector joining the centers of mass of ACE2 and the S1 subdomain. In this way, the S1-apt2 complex had to accommodate for the presence of the incoming ligand, which in turn was deformed as well by the forced contact. By following the behaviour of the contact surface and solvation free energy $\Delta G$, over a

---

**Table 1.** Hydrogen bonds formed at the ACE2-S1 interface in the crystallographic experimental configuration (RCSB entry 6MJ0\(^6\)), and in the "best binding" configuration from molecular dynamics simulation with the apta2 DNA aptamer docked to S1. Donor/acceptor species are labelled according to the AMBER99 atom codes\(^15\); molecular structure data analysed by the PDBePISA utility\(^25\).

#### Experimental configuration (RCSB entry 6MJ0)

| ACE2 side residue | bond species | length (Å) | S1 side residue |
|-------------------|--------------|------------|----------------|
| GLN42             | NE2          | 3.24       | O GLY446       |
| GLN42             | NE2          | 2.79       | OH TYR449      |
| TYR83             | OH           | 2.79       | OD1 ASN487     |
| TYR83             | OH           | 3.54       | OH TYR489      |
| LYS353            | NZ           | 3.08       | O GLY496       |
| TYR41             | OH           | 2.71       | OG1 THR500     |
| ARG393            | NH2          | 3.73       | OH TYR505      |
| ASP30             | OD2          | 2.90       | NZ LYS417      |
| ASP38             | OD2          | 2.69       | OH TYR449      |
| GLN24             | OE1          | 2.69       | ND2 ASN487     |
| GLU35             | OE2          | 3.50       | NE2 GLN493     |
| TYR41             | OH           | 3.67       | N ASN501       |
| LYS353            | O            | 2.78       | N GLY502       |
| GLU37             | OE2          | 3.46       | OH TYR505      |

#### Best-binding configuration from force-driven docking

| ACE2 side residue | bond species | length (Å) | S1 side residue |
|-------------------|--------------|------------|----------------|
| TYR41             | OH           | 2.30       | O THR500       |
| LYS353            | NZ           | 2.79       | OD1 ASN501     |
| LYS353            | NZ           | 2.82       | OH TYR505      |
Figure 6. Contact regions between the ACE2 receptor (purple ribbons; light-red, or blue spheres for contacting residues), the S1 subdomain (cyan ribbons; green spheres), and the DNA aptamer (red ribbons; red spheres). The remaining of the whole S-protein trimer is not shown, but is included in the MD simulation with water and ions. View angle approximately from the central axis of the S-protein trimer. (a) The experimental configuration from Ref.[6] (RCSB entry 6MJ0). (b) The configuration for the start of the force-driven docking: ACE2 from (a) is rigidly shifted up by 1 nm; blue/white arrows indicate modifications of the receptor binding site induced by the DNA aptamer. (c) The "best-binding" configuration issued from the cluster analysis of the 500 ns trajectory at T=310 K; contact surface between ACE2 and S1 is indicated by the light-red and green spheres, and by red-blue spheres for ACE2-DNA contact.

sequence of snapshots taken at intervals of 20 ps, a "best binding" configuration (that is, minimising excess free energy and maximising contact surface and number of H-bonds, see Figure 5) could be identified. This corresponds to configurations around approach coordinate 0.475 (arbitrary units between [0, 1]). However, note that the solvation free energy $\Delta G$ is positive for the whole approaching trajectory: a positive value indicates hydrophilic interfaces, therefore absence of protein affinity.

This configuration was further equilibrated at T=310K for 100 ns, and cluster analysis of the resulting trajectory revealed the average binding configuration of ACE2 to the S-protein in the presence of ssDNA apta2. Figure 6 shows the contact region between the receptor (purple ribbons) and the S1 domain (cyan ribbons), respectively, in the experimental co-crystallized conformation (a), and in the MD simulation (b,c), also including the DNA apta2 (red ribbons); residues implicated in the interface are depicted by VdW spheres. By comparison with the experimental adhesion structure (Fig. 6(a), light red-green spheres), it can be seen that the presence of the aptamer has the double effect of: (i) deforming the binding site, in particular by pushing down the 476-486 loop (Fig. 6(b), blue arrow) and by disrupting some of the beta-sheets (white arrows); and (ii) shielding the RBD from the ACE2 receptor, which remains on the periphery of the binding surface with a much limited interaction. The contact surface between ACE2 and S1 is reduced to merely 1.2 nm$^2$, and the $\Delta G$ is positive at about 0.3 kcal/mole. A limited number of H-bonds is formed by ACE2 with the RBD: the receptor is able to find the main binding region, however the number of H-bonds is reduced from about 14 to just 3 (see Table 1), with bonds at TYR41 and LYS353 still observed but modified, and the salt bridge between ASP30 and LYS417 disappears. Notably, however, a much stronger binding is made with the protruding tip of the DNA aptamer, with a negative $\Delta G = -2$ kcal/mol, and two strong H-bonds between the polar residues LYS68-GLN60 and the O3' backbone oxygens of nucleotides A34-A35 (Fig. 6(c), red-blue spheres). Such conformations, with the receptor doubly bonded to the DNA aptamer and (to a weaker extent) to part of the S1 subdomain, are therefore likely to preclude furthering of the interaction between the receptor and the viral S-protein, and to hamper the very early stages of the membrane fusion process.

4 DISCUSSION

The spheroidal surface of the SARS-CoV-2 virus is decorated with a large density of copies of the transmembrane S-protein, its three protomers being composed of two major S1 and S2 catalytic domains, plus other structural regions. As it is becoming clear from the recent literature, coronavirus entry in the host cell requires a concerted action of the receptor binding at the S1/S$^B$ domain (typically, the receptor ACE2 present at the surface of most human cells), and the subsequent
proteolytic processing of the S1-S2 link (also susceptible to furin cleavage), to allow the fusion domain S2 to initiate the fusion process between the virus and cell membrane. The S1 domain is experimentally found in two conformations: a “closed” one, in which the receptor binding sites (RBD) are inaccessible to ACE2, and an “open” one, in which ACE2 can effectively bind one S-protein from the virus. Both cryo-microscopy and X-ray diffraction data have shown that the S-protein protomers fluctuate between these two conformations with about 50/50 occupation probability; moreover, always only one protomer is observed in the open conformation, but never 2 or 3 at the same time. In our study we analyzed the interaction of two experimentally identified DNA aptamers with the whole trimeric structure of the S-protein, instead of focusing just on the very small binding regions as is typically done both in experimental and molecular docking studies. This more conservative and extensive choice allowed to reach some important conclusions, as detailed in the following.

One possible way in which aptamers could act as therapeutic devices would be to design their target nucleotide sequence so as to directly interfere with the receptor binding at the RBD. This was not entirely the case for the two experimentally identified aptamers used in this study. As we showed in the last Section, by means of a combination of docking and force-driven molecular dynamics simulations, their interaction with S1 occurs at a region nearby the RBD, close enough to strongly modify the interaction site, and partly hide it from contact with the human ACE2 receptor. However, in order to exploit a more direct blocking effect, more precisely targeted aptamers should be identified experimentally.

On the other hand, another possibility is that aptamers may bind in such a way to limit, or even block the opening of the S1 domain, which is indeed the critical step to elicit the interaction with the cell receptor. Our finding that DNA aptamers with strongly specific interaction with the S1 domain, can also interact with the S2 of another protomer, thereby making a kind of “bridge” between the two adjacent protomers, induces important consequences. Results of free energy calculations by the umbrella sampling method, clearly demonstrate the possibility that the DNA aptamer bridging between two S monomers can actively block, or at least slow down considerably the opening of S1, which is the critical step to elicit the interaction with the cell receptor, thereby suppressing, or strongly reducing the receptor binding probability. The relatively high free energies of binding of the aptamers to the S-protein point to a very high (even sub-nanomolar) sensitivity of the recognition mechanism.

In the recent literature, it has been advanced the hypothesis that Coronaviruses could also bind to membrane-bound sialic acids as a preliminary step, to facilitate ACE2 binding and subsequent membrane fusion, with affinities in the range of ∼50µM. Given the values of ΔG and extra force observed in our simulations, even a concerted protein-glycan working to prepare the opening of the S1 and initiate membrane fusion appears unlikely, once the DNA aptamer is docked to S1. However, such a possibility calls for a detailed study of both the specific and more general roles of glycans in the S1-ACE2-aptamer binding process, which is beyond the purpose of the present study.

5 CONCLUSION

In this work we investigated by means of state-of-the-art protein docking and large-scale molecular dynamics simulations, the interaction of some experimentally identified DNA aptamers with the S-protein of SARS-CoV-2. We showed that the DNA aptamers can bind efficiently to the designated receptor-binding domain (RBD) on one protomer of the S-protein, but also form and maintain stable bonds with the S2 domains of adjacent protomers. Such an extended bonding interaction, actually impossible to deduce from the experimental measurements of generic binding affinity, is found to strongly restrain the opening of the RBD to the cell receptors, and should lead to a drastic reduction of the virus/cell binding efficiency.

Overall, the present results constitute a qualitative, rather than quantitative, suggestion for a novel biochemical interaction process, which may have important impact on the molecular mechanisms underlying viral invasion of the host cell. The fact that DNA aptamers are extremely selective, with sub-nanomolar sensitivity, very cheap to produce in large quantities, and extremely biocompatible with practically no adverse effects (since they have very little affinity for targets different from the one against which they are designed), make such findings a potential lead for a novel therapeutic concept.

ACKNOWLEDGEMENTS

Computer resources provided by the CINES and IDRIS French Supercomputing Centres, under grants a2020/077225 and by generous extensions thereof.

Conflict of interest statement.
None declared.

References

1. Xu, J. et al. Systematic comparison of two animal-to-human transmitted human Coronaviruses: SARS-CoV-2 and SARS-CoV. Virus 12, 244 (2020).
2. Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 395, 565–574 (2020).
3. Walls, A. C. et al. Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 180, 281–292 (2020).
4. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263 (2020).
5. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).

6. J.Lan et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

7. Cheng, C. et al. Potent inhibition of human influenza H5N1 virus by oligonucleotides derived by SELEX. *Biochem. Biophys. Res. Comm.* **366**, 670–674 (2008).

8. Jang, K. J., Lee, N. R., Yeo, W. S., Jeong, Y. J. & Kim, D. E. Isolation of inhibitory RNA aptamers against severe acute respiratory syndrome (SARS) coronavirus NTPase/Helicase. *Biochem. Biophys. Res. Comm.* **366**, 738–744 (2008).

9. Song, Y. et al. Discovery of aptamers targeting receptor-binding domain of the SARS-CoV-2 spike glycoprotein. *Anal. Chem.* **92** (2020).

10. Chen, Z., Wu, Q., Chen, J., Ni, X. & Dai, J. A DNA aptamer based method for detection of SARS-CoV-2 nucleocapsid protein. *Virol. Sin.* **35**, 351–354 (2020).

11. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406–3415 (2003).

12. Zuker, M. & Jacobson, A. B. Using reliability information to annotate RNA secondary structures. *RNA* **4**, 669–679 (1998).

13. Popenda, M. et al. Automated 3D structure composition for large RNAs. *Nucl. Acids Res.* **40**, e112 (2012).

14. G.C.P van Zundert, J. P. R. *et al.* The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).

15. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Prot. Chem.* **66**, 27–85 (2003).

16. III, T. E. C. & Case, D. A. Twenty-five years of nucleic acid simulations. *Biopolymers* **99**, 969–977 (2013).

17. Pérez, A., Luque, F. J. & Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **45**, 196–205 (2012).

18. Berendsen, H., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm* **91**, 43–56 (1995).

19. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **7**, 306–317 (2001).

20. Woo, H. *et al.* Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *J. Phys. Chem. B* **124** (2020).

21. Casalino, L. *et al.* Shielding and beyond: The roles of glycans in SARS-CoV-2 spike protein. 11 June 2020, biorXiv pre-print: not peer-reviewed. URL https://doi.org/10.1101/2020.06.11.146522.
SUPPLEMENTARY MATERIALS

Supplementary figure S1. 2D structures of the DNA aptamers apta1 ((a), with 51 nucleotides) and apta2 ((b), with 67 nucleotides) issued by mfold. Color codes refer to the ss-count, indicating the propensity for base-pairing, increasing from red (0), to green (0.33), to blue (0.67), to black(1). The complete color code and further explanations of the computational procedure can be found in Ref.[12], and at the url: https://unafold.rna.albany.edu (see section "Structure display and free energy determination"). Both (a) and (b) closely resemble the 2D-foldings from Ref.[9]. For apta2 we also show in (c) an alternate folding, which had a very close free energy; however, only folding (b) was used to generate the 3D structure. The following tables provide the individual contributions of each structural element to the overall folding free-energy $\Delta G$.

### Supplementary Table 1. Aptamer apta1 2D-folding (a).

| Structural element | $\Delta G$ | Information |
|--------------------|------------|-------------|
| External loop      | -1.70      | 1 ss bases + 2 closing helices |
| Stack              | -2.20      | External closing pair G31-C50 |
| Stack              | -2.40      | External closing pair T32-A49 |
| Stack              | -3.30      | External closing pair C33-G48 |
| Stack              | -2.10      | External closing pair C34-G47 |
| Helix              | -10.00     | 5 base pairs |
| Hairpin loop       | +6.20      | Closing pair A35-T46 |
| Stack              | -2.10      | External closing pair C1-G30 |
| Stack              | -2.10      | External closing pair A2-T29 |
| Stack              | -3.40      | External closing pair G3-C28 |
| Stack              | -2.10      | External closing pair C4-G27 |
| Stack              | -2.20      | External closing pair A5-T26 |
| Helix              | -11.90     | 6 base pairs |
| Interior loop      | +1.20      | External closing pair C6-G25 |
| Stack              | -3.30      | External closing pair C10-G22 |
| Helix              | -3.30      | 2 base pairs |
| Hairpin loop       | +4.40      | Closing pair C11-G21 |
**Supplementary Table 2. Aptamer apta2 2D-folding (b).**

| Structural element | \( \Delta G \) | Information                              |
|--------------------|----------------|------------------------------------------|
| External loop      | -0.40          | 3 ss bases + 2 closing helices            |
| Stack              | -2.40          | External closing pair C53-G66            |
| Helix              | -2.40          | 2 base pairs                             |
| Hairpin loop       | +4.10          | Closing pair G54-C65                     |
| Stack              | -1.10          | External closing pair A1-T50             |
| Stack              | -2.40          | External closing pair T2-A49             |
| Stack              | -3.30          | External closing pair C3-G48             |
| Helix              | -6.80          | 4 base pairs                             |
| Interior loop      | +0.40          | External closing pair C4-G47             |
| Stack              | -2.40          | External closing pair G6-C45             |
| Helix              | -2.40          | 2 base pairs                             |
| Bulge loop         | +4.50          | External closing pair A7-T44             |
| Stack              | -3.40          | External closing pair G13-C43            |
| Stack              | -2.10          | External closing pair C14-G42            |
| Stack              | -2.10          | External closing pair A15-T41            |
| Stack              | -3.40          | External closing pair G16-C40            |
| Helix              | -11.00         | 5 base pairs                             |
| Hairpin loop       | +5.90          | Closing pair C17-G39                     |

**Supplementary Table 3. Aptamer apta2 2D-folding (c).**

| Structural element | \( \Delta G \) | Information                              |
|--------------------|----------------|------------------------------------------|
| External loop      | -1.10          | 9 ss bases + 2 closing helices            |
| Stack              | -1.40          | External closing pair C53-G60            |
| Helix              | -1.40          | 2 base pairs                             |
| Hairpin loop       | +4.70          | Closing pair G54-T59                     |
| Stack              | -1.10          | External closing pair A1-T50             |
| Stack              | -2.40          | External closing pair T2-A49             |
| Stack              | -3.30          | External closing pair C3-G48             |
| Helix              | -6.80          | 4 base pairs                             |
| Interior loop      | +0.40          | External closing pair C4-G47             |
| Stack              | -2.40          | External closing pair G6-C45             |
| Helix              | -2.40          | 2 base pairs                             |
| Bulge loop         | +4.50          | External closing pair A7-T44             |
| Stack              | -3.40          | External closing pair G13-C43            |
| Stack              | -2.10          | External closing pair C14-G42            |
| Stack              | -2.10          | External closing pair A15-T41            |
| Stack              | -3.40          | External closing pair G16-C40            |
| Helix              | -11.00         | 5 base pairs                             |
| Hairpin loop       | +5.90          | Closing pair C17-G39                     |
Supplementary Figure 1