# *In silico* design and validation of high-affinity RNA aptamers for SARS-CoV-2 comparable to neutralizing antibodies

Yanqing Yang[1,4], Lulu Qiao[2,4], Yangwei Jiang[1], Zhiye Wang[2, 3,*], Dong Zhang[1,*], Damiano Buratto[1], Liquan Huang[1], and Ruhong Zhou[1,3,*]

[1]Institute of Quantitative Biology, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China.

[2]State Key Laboratory of Plant Environmental Resilience, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China.

[3]The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang 310058, China.

[4]These authors contributed equally: Yanqing Yang and Lulu Qiao.

* To whom correspondence should be addressed: rhzhou@zju.edu.cn (R.Z.); zhangd_iqb@zju.edu.cn (D.Z.); wangzhiye1@zju.edu.cn (Z.W.).

**Abstract**

Nucleic acid aptamers hold promise for clinical applications, yet understanding their molecular binding mechanisms to target proteins and efficiently optimizing their binding affinities remain challenging. Here, we present CAAMO (Computer-Aided Aptamer Modeling and Optimization), which integrates *in silico* aptamer design with experimental validation to accelerate the development of aptamer-based RNA therapeutics. Starting from the sequence information of a reported RNA aptamer, Ta, for the SARS-CoV-2 spike protein, our CAAMO method first determines its binding mode with the spike protein's receptor binding domain (RBD) through a multi-strategy computational approach. We then optimize its binding affinity via structure-based rational design. Among the six designed candidates, five were experimentally verified and exhibited enhanced binding affinities compared to the original Ta sequence. Furthermore, we directly compared the binding properties of the RNA aptamers to neutralizing antibodies, and found that the designed aptamer $Ta^{G34C}$ demonstrated a comparable binding affinity to the RBD compared to all tested neutralizing antibodies. This highlights its potential as an alternative to existing COVID-19 antibodies. Our work provides a robust approach for the efficient design of a relatively large number of high-affinity aptamers with complicated topologies. This approach paves the way for the development of aptamer-based RNA diagnostics and therapeutics.

**Keywords:** RNA aptamer, rational design, molecular dynamics, electrophoretic mobility shift assay, SARS-CoV-2

## 1. Introduction

Aptamers are short, single-stranded RNA or DNA oligonucleotides that bind tightly and specifically to target molecules [1,2], earning the designation of "chemical antibodies" [3,4]. They combine the advantageous properties of small organic compounds and antibodies, exhibiting strong affinity and specificity comparable to monoclonal antibodies, while remaining mostly non-immunogenic and highly capable of tissue penetration similar to those of small molecules [5]. These versatile attributes have led to their widespread applications in biomedical fields, including drug discoveries [6], biomarker identifications [7], therapeutics [8], diagnostics [9], and biosensors [10]. The U.S. Food and Drug Administration's (FDA) approval of Macugen (pegaptanib sodium, Pfizer / Eyetech) in 2004, which targets vascular endothelial growth factor (VEGF) to treat wet age-related macular degeneration, was a landmark achievement for aptamer-based drugs [11]. This success has driven further exploration of aptamer-based therapies. To date, a variety of aptamers have been proposed to neutralize a range of disease-related proteins, such as human epidermal growth factor receptor 2 (HER2) [12], epidermal growth factor receptor (EGFR) [13], and prostate-specific membrane antigen (PSMA) [14]. Recently, the significant contributions of mRNA vaccines [15], aptamers [16] and other RNA biotherapeutics [17,18] have highlighted the potential of RNA-based interventions during the COVID-19 pandemic. This global health crisis has underscored the importance of RNA therapeutics, including medical aptamers.

The systematic evolution of ligands by exponential enrichment (SELEX) technique, the most commonly used method for aptamer screening, still faces several challenges [19]. Firstly, due to its random selection, the candidate aptamers are identified regardless of any atomic mechanisms underlying the interactions between the nucleic acids and target proteins; and secondly, because of limited sizes of oligonucleotide libraries, the identified candidates may not be the best aptamers, which, instead, can be used as lead sequences for further optimization [20]. Therefore, developing *in silico* methods that can complement SELEX screening and provide atomic structure-based rational aptamer design is highly desirable. For instance, we successfully developed two optimized RNA aptamers targeting epithelial cellular adhesion molecule (EpCAM) using a structure-based *in silico* method. We further confirmed their enhanced affinities compare to a previously patented nanomolar aptamer [21].

These two EpCAM-targeted aptamers were relatively small (19 nucleotides) and structurally simple, with a 4-bp stem and a 10-nucleotide long hairpin loop. Their simplicity allowed for precise determination of the aptamer-protein complex structure and enabled structure-based design through conventional computational modelling and simulations. In comparison, RNA aptamers screened through SELEX technique are typically much larger, often comprising tens of nucleotides and featuring complex topologies, including tertiary structural motifs such as internal loops and G-quadruplex. These features not only contribute to increased difficulties in accurate structure modelling, but also increase the possibilities of conformational changes upon binding to the target molecules (i.e. folding upon binding) [22]. Furthermore, the limited number of experimentally determined structures for aptamer-protein complexes in the Protein Data Bank (PDB) present further challenges for an accurate structure modelling. These limitations highlight the need for more advanced computational frameworks to accurately predict the binding modes of RNA aptamers of varying sizes and topologies, thus enabling efficient *in silico* aptamer engineering.

A direct comparison between aptamers (also known as "chemical antibodies") and conventional biological antibodies in terms of binding mechanisms and affinities is of great interest. However, such comparative studies are currently limited. A prerequisite for such comparison is the availability of a target protein of functional importance for which both effective aptamers and antibodies exist. The receptor binding domain (RBD) of the spike (S) protein expressed by the SARS-CoV-2 is one such target (see Fig. 1A). It binds to human angiotensin-converting enzyme 2 (ACE2) with high affinity to facilitate viral entry into host cells [23,24]. Targeting RBD to block the interactions between SARS-CoV-2 spike protein and human ACE2 emerged as a promising therapeutic strategy [25] to fight COVID-19 pandemic. For example, RBD-targeted antibodies have shown substantial therapeutic potential due to their potent neutralizing effects [26,27]. Alternatively, a recent study reported an RNA aptamer generated through SELEX, termed RBD-PB6-Ta (hereafter referred to as "Ta"), that binds to the RBD of SARS-CoV-2 spike protein with high affinity and efficiently blocks viral infection at low concentrations (see Fig. 1B) [16]. Based on these observations, the RBD of SARS-CoV-2 spike protein represents an ideal target to directly compare the efficacy of aptamers and antibodies. Such comparisons could enhance our understanding of

the aptamer binding mechanism, guiding the rational design of aptamers that can possibly be comparable with or even superior to neutralizing antibodies, thereby complementing the existing means of treatment (see Fig. 1C).

In this work, we present the CAAMO (Computer-Aided Aptamer Modeling and Optimization) framework, which combines computational techniques and experimental validation to design high-affinity aptamers. Starting with an RNA sequence, we first predicted the most probable binding mode of the RNA aptamer with the target protein using a combination of *in silico* methods, including RNA structure prediction, ensemble docking, molecular dynamics (MD) simulations, steered molecular dynamics (SMD) simulations, and binding free energy calculations. We then performed a comparative analysis between the aptamer Ta and several popular neutralizing antibodies, focusing on their binding sites (using computational methods) and binding capabilities (through both computational and experimental approaches). We identified potential mutation sites for affinity enhancement and developed six novel aptamers through *in silico* mutagenesis study with free energy perturbation (FEP) method. Among these, electrophoretic mobility shift assay (EMSA) experiments confirmed that five had improved binding affinities compared to the original aptamer Ta. Notably, the aptamer $Ta^{G34C}$ exhibited the highest binding affinity to the RBD, outperforming all tested neutralizing antibodies in competitive binding assays. These findings demonstrate the effectiveness of the CAAMO framework in developing high-affinity RNA aptamers targeting the RBD of SARS-CoV-2 spike protein, providing new therapeutic strategies against COVID-19. Moreover, the CAAMO framework can be extended to aptamers that are larger and with more complex topologies.

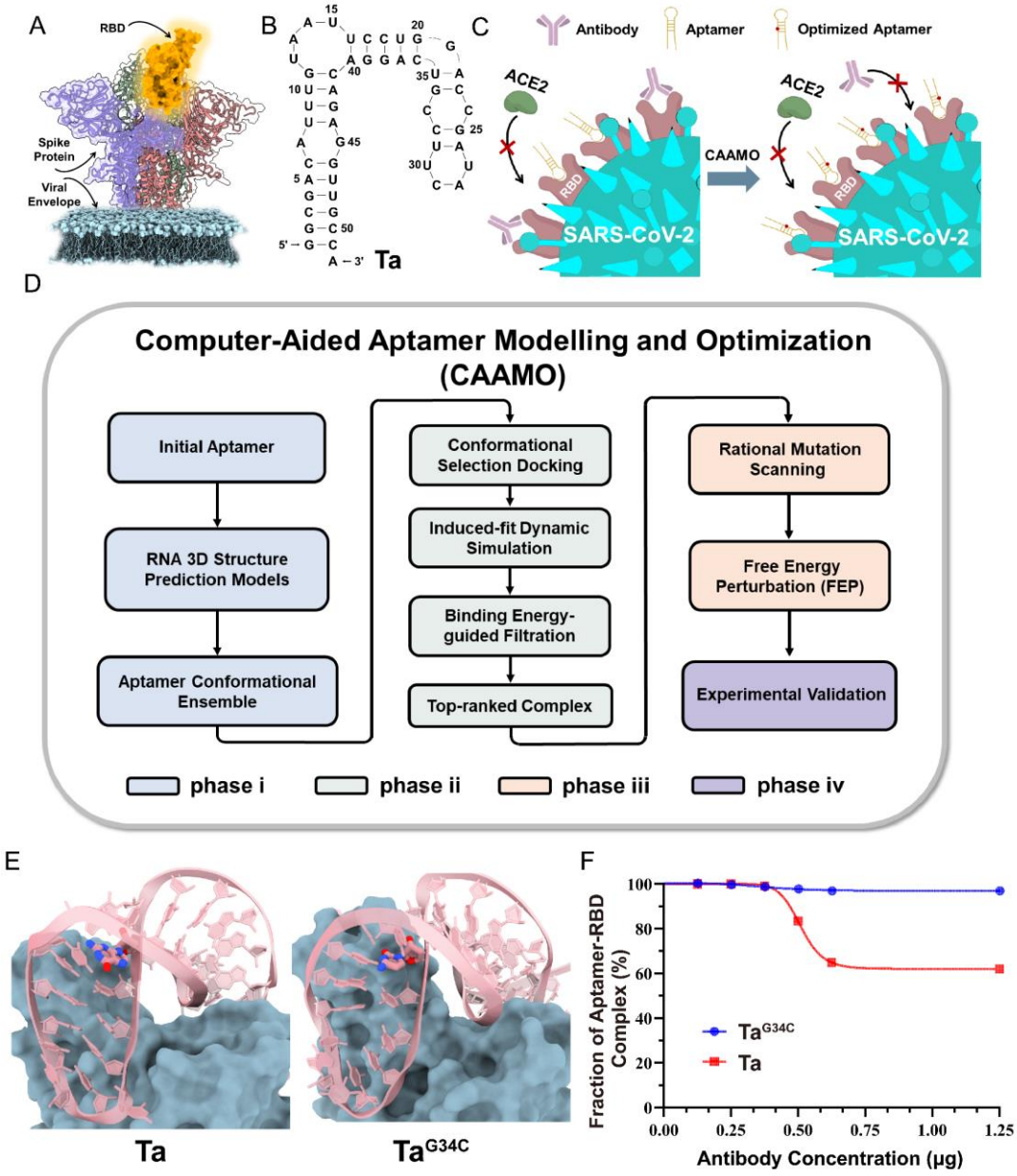


**Fig. 1.** The workflow of the CAAMO framework for designing high-affinity RNA aptamers and its application in the development of novel RNA aptamers targeting the RBD of the SARS-CoV-2 spike protein. (A) Schematic diagram of the spike protein (trimer) of SARS-CoV-2 virus with one of the three RBD regions highlighted in gold. (B) The sequence and Mfold-predicted secondary structure of a SELEX-derived RNA aptamer termed Ta targeting the RBD of SARS-CoV-2 spike protein, which was the initial input information of the CAAMO framework. (C) Schematic diagrams showing that high-affinity RNA aptamers optimized through the CAAMO framework can complement existing antibody-neutralizing treatments for COVID-19. Antibodies (purple) and RNA aptamers (yellow) that bind to the RBDs of the SARS-CoV-2 spike protein can neutralize viral infection by blocking its interaction with the human ACE2. The designed aptamers (marked by red dots) with binding affinities comparable to antibodies can further

strengthen the neutralizing treatment when antibody escape occurs. (D) Illustrative workflow of the CAAMO framework by integrating computational techniques with experimental validation. (E) A representative output after CAAMO framework optimization (the aptamer $Ta^{G34C}$) and its comparison with the original Ta sequence. RNA aptamer is colored by pink and RBD is shown in dark blue surface. The mutated nucleotides (G34 in wild type Ta and C34 in optimized $Ta^{G34C}$) are highlighted in pink sticks, atoms oxygen and nitrogen are colored by red and blue, respectively. (F) Competitive binding experiments to compare the binding capabilities of RNA aptamers Ta or $Ta^{G34C}$ and a commercial neutralizing antibody (SinoBiological, Cat: 40592-R001) to the RBD of SARS-CoV-2 spike protein. Gradient amount of the commercial neutralizing antibody was titrated into the buffer containing 5 μM aptamer (Ta or $Ta^{G34C}$) and 40 μM RBD. The intensity of aptamer-RBD bands were quantified with Image J and normalized to that of the mixture without antibody, which was set to 100%. Data were collected from the images of EMSA shown in Fig. 5D.

## 2. Results

### 2.1 Overview of the CAAMO framework for high-affinity RNA aptamer design

For RNA aptamers composed of tens of nucleotides with complicated topological structures, accurately determining their binding modes with target proteins is challenging due to the huge conformational space available, which hinders efforts in structure-based design of high-affinity aptamers. To address these challenges, we propose a promising framework named CAAMO, which integrates computational techniques with experimental validation. This CAAMO framework is designed not only to provide the structural insights into key nucleic acid-protein interactions but also to facilitate the efficient design of aptamers with enhanced affinity. The framework consists of four main phases (see Fig. 1D and Fig. S1): (i) constructing an aptamer conformational ensemble by employing several cutting-edge RNA three-dimensional (3D) structure prediction methods, (ii) identifying an appropriate aptamer binding mode through integrating conformational selection docking, induced-fit dynamic simulation [28], and binding energy-guided filtration, (iii) optimizing the aptamer sequence through *in silico* mutagenesis study using free energy perturbation calculations, and (iv) final experimental validation of the designed aptamer candidates. Further details on the CAAMO

framework are provided in the following subsections and in the "Materials and Methods" section.

An RNA aptamer sequence termed "Ta" (containing 52 nucleotides; see Fig. 1B), previously shown to bind the RBD of SARS-CoV-2 spike protein with high affinity [16], was chosen to illustrate the CAAMO framework. The initial input to the CAAMO is the nucleotide sequence of the aptamer Ta, and the output is several *computationally* designed and subsequently *experimentally* validated aptamers with improved binding affinities compared to the initial sequence. A representative output (the aptamer $Ta^{G34C}$), that exhibits a ~3.3-fold stronger binding affinity than the original aptamer Ta, is shown in Fig. 1E. Notably, the success rate of the *in silico* design via the CAAMO framework is very promising, with five out of the six (~83%) computationally designed candidate aptamers experimentally confirmed to have improved binding affinities. Furthermore, competitive binding experiments with popular neutralizing antibodies revealed that the designed RNA aptamer $Ta^{G34C}$ has a higher binding affinity to the RBD of SARS-CoV-2 spike protein than the neutralizing antibodies (see Fig. 1F). Thus, the newly identified aptamer $Ta^{G34C}$ shows great potential as a complement to existing antibody-based neutralizing treatments for COVID-19, especially when antibody escape occurs in emerging SARS-CoV-2 variants [29]. Overall, these results not only indicate the proposed putative binding conformation of the aptamer Ta to the RBD as a reasonable representation of the binding complex, but also demonstrate the great potential of our CAAMO framework in aptamer design and optimization.

### 2.2 Determination of the binding model of the aptamer Ta to the RBD

Accurately predicting the binding complex of the RNA aptamer with the target protein is a critical step in structure-based *in silico* aptamer design, especially with aptamers of multiple nucleotides and complex topologies. To address this challenge, we constructed a multi-strategy approach that includes RNA 3D structure prediction, ensemble docking/clustering, and binding capacity assessment to identify the most probable binding conformation of the aptamer Ta with the RBD of SARS-CoV-2 spike protein (see Fig. 2A). The RBD conformation was extracted from the crystal structure of the RBD-ACE2 complex (PDB id: 6LZG) and then refined using MD simulation (see Fig. S2). For the aptamer Ta, we first predicted its secondary structure using Mfold [30], forming a stem-loop structure

containing five stems, two internal loops, two bulges and one hairpin loop (see Fig. 1B). Then, we built a 3D conformational ensemble using the state-of-the-art RNA 3D structure prediction models because of the inherent flexibility of unpaired nucleotides in loop regions and potential conformational changes upon binding to the target protein. A total of 25 representative aptamer Ta structures with different degrees of bending (see Fig. S3A) were selected for subsequent molecular docking. It should be noted that, since most popular structure prediction models, such as FARFAR2 [31], IsRNA2 [32] and SimRNA [33], adopt a conformation clustering strategy to obtain the top predictions, each of these 25 predicted structures represents a collection of similar 3D conformations.

From the ensemble docking complex pool (containing 1,000 aptamer-RBD binding poses; see Fig. S3B), we generated six major clusters, each representing distinct binding conformations. For each cluster, we selected the structure with the lowest binding energy, estimated using the molecular mechanics generalized Born surface area (MM/GBSA) approach based on single-frame conformation, as the representative binding conformation (see Fig. 2B and conformations 01-06 in Fig. S3C). MD simulations of these six representative conformations confirmed that they maintained stable binding modes over the course of the long-time simulations (see Fig. S4). We further analyzed the binding abilities between the aptamer Ta and RBD for these six MD refined conformations using MM/GBSA and steered molecular dynamics (SMD) simulations to determine the most likely binding conformation. As shown in Fig. 2C, the MM/GBSA binding energy of conformation 01 ($\Delta G$=-142.93±7.65 kcal/mol) is significantly lower than the other five candidates ($\Delta G$>-120 kcal/mol). Interestingly, we found that the binding energy of each conformation correlated with the sizes of its corresponding clusters (see Fig. 2C and Fig. S3C), with a larger cluster associated with lower binding energy. Additionally, we computed the rupture works required to separate the bound aptamer from the RBD using SMD simulations. Among four tested cases (conformations 01-04), the conformation 01 rupture work is the largest (254.2±47.9 kcal/mol) and significantly different from that of the other conformations (see Fig. 2D). Despite the limitations of MM/GBSA and SMD methods (more rigorous FEP method will be used later), both approaches consistently showed that conformation 01 had the strongest binding ability between the aptamer Ta and the RBD. Based on these results, we selected

conformation 01 as the putative binding model for further structure-based rational design of high-affinity aptamers.

In the putative binding complex, the aptamer Ta adopts a saddle-like shape to bind the RBD (see Fig. 2E). We divided the aptamer into three parts: the apical loop (contains two base pairs; nucleotides C24-C33), the bulge region (nucleotides G11-C23 and G34-C41), and the end stem part (contains one mismatch base pair; nucleotides G1-U10 and A42-A52). The apical loop and end stem parts bind to opposite sides of the RBD , respectively, while the bulge acts as a linker, adjusting the bending angle of the saddle-like shape. The binding between the aptamer Ta and RBD is stabilized by electrostatic, hydrogen bonding, and van der Waals interactions. In details, some important interactions include the electrostatic interactions between the phosphate groups of nucleotides C29 and U30 and the basic amino acid ARG408, and between the phosphate groups of C3, A42 and G43 and the LYS444; the hydrogen bonds between GLN409@Nε2-Hε21…C29@O2', LYS417@Nζ-Hζ3…U31@O2, U42@O2'-HO2'…ASN448@Oδ1, ASN450@N-H…U42@ O2', A14@N6-H61…GLY482@O, A14@O2'-HO2'…GLU484@Oε2, ASN487@ Nδ2-Hδ22…C24@O2' and TYR489@OH-HH…G25@O2' and the van der Waals packing interactions between nucleotides of the apical loop part and some aromatic residues of the RBD (such as TYR421, TYR453, PHE456, TYR473, PHE486, and PHE489) (see Fig. S5). These detailed interactions provide important insights for future *in silico* aptamer design.

In addition to the aptamer Ta, we also analyzed a weaker binding aptamer sequence, RBD-PB6-Tc (Tc), as a negative control [16]. We constructed a binding model for the Tc sequence with the RBD using the same approach as for the aptamer Ta (see Fig. S6). SMD simulations confirmed that the binding strength of Tc to the RBD was significantly weaker than that of the aptamer Ta (see Fig. S6D), which proves the reliability of our approach to determine the binding model of an RNA aptamer with a target protein. Additionally, our EMSA experiments also confirmed that the binding affinity of the Ta-RBD complex is stronger than that of the Tc-RBD complex (see Figs. 2F and 2G), in good agreement with the previous study [16]. The measured dissociation constant for the Ta-RBD complex is $K_d$=110.7 μM, while the sequence Tc is unable to bind to the RBD under all conditions tested.

These results motivate the design of high-affinity RNA aptamers targeting the RBD of the SARS-CoV-2 spike protein to enhance the treatment of COVID-19.

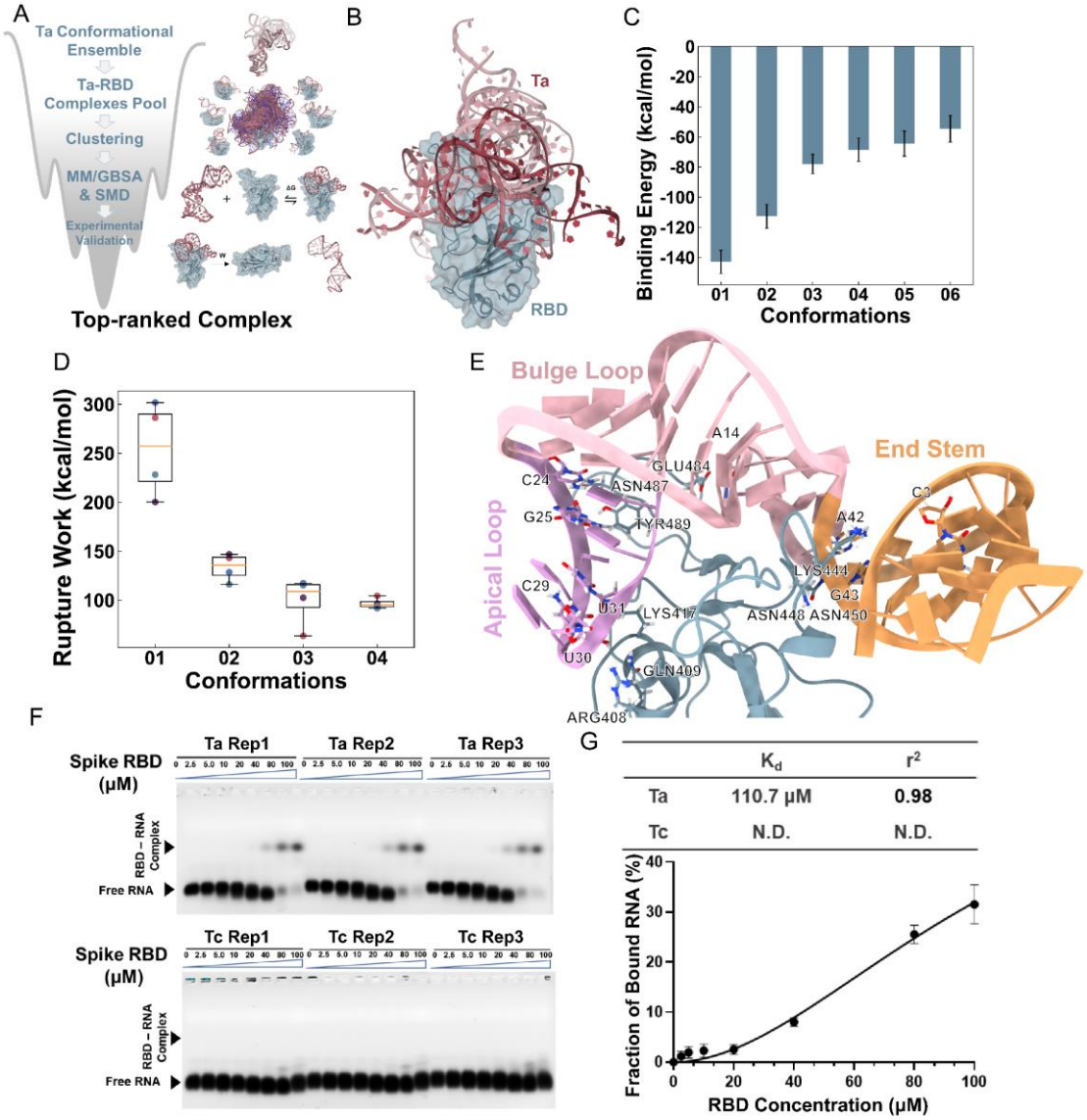


**Fig. 2.** Determination of the binding model of the aptamer Ta and the RBD of the SARS-CoV-2 spike protein *via* a multi-strategy approach. (A) Flowchart illustrating the combination of RNA 3D structure prediction, ensemble docking and clustering, and binding capacity assessment by MM/GBSA and steered MD to determine the most probable binding conformation of the aptamer Ta to the RBD. (B) Six main binding modes , i.e., conformations 01-06, with clearly distinct binding conformations after ensemble docking and clustering for further binding ability assessment. The RBD and the aptamer Ta are shown in dark blue and dark red, respectively. For clarity, the surface of RBD is also displayed. For the aptamer Ta, conformations 01-06 are named in a descending order of their respective cluster sizes and their colors are

lightened gradually. (C) The binding energies between the aptamer Ta and RBD estimated by MM/GBSA method for six binding candidates shown in plane (B). Data represent mean±s.d. collected from three independent calculations. (D) The rupture works required to separate the bound aptamer Ta from the RBD for different binding conformation candidates. Data were collected from four independent steered MD simulations. Limited by the available computing resources, only the first four binding conformations were assessed. The orange line is the median, boxes extended form lower to upper quartiles, whiskers showing the range of nonoutlier data. (E) Overview of the most probable binding model (conformation 01) of the aptamer Ta and RBD. The RBD is shown in cyan cartoon representation, the apical loop, bulge part, and end stem of the aptamer Ta are colored plum, pink, and sandy brown, respectively. The key residues of the RBD and nucleotides of the aptamer Ta for binding interactions are shown in sticks, all nitrogen atoms colored by blue, and all oxygen atoms by red. (F) EMSA results of the aptamer Ta (upper panel) and the weakly binding aptamer Tc (negative control, lower panel) bound to the RBD of the SARS-CoV-2 spike protein and (G) the resultant binding curve for the aptamer Ta. The dissociation constant ($K_d$) was calculated from the EMSA image quantification from three independent experiments.

## 2.3 Comparison of binding properties between the aptamer Ta and neutralizing antibodies

In recent years, researchers have developed numerous antibodies to neutralize SARS-CoV-2 infection by blocking the RBD of the spike protein from binding to ACE2 [34]. Their binding properties with the RBD have been extensively studied. Since we determined the most probable binding model of the aptamer Ta to the RBD, comparing the binding properties of the aptamer Ta with those of the neutralizing antibodies to the RBD is both feasible and meaningful. A preliminary comparison of the respective binding modes of ACE2, the aptamer Ta, and neutralizing antibodies to the RBD (see Fig. S7A) indicates that the aptamer Ta can precisely occupy the binding site of ACE2, similar to many neutralizing antibodies, thereby blocking SARS-CoV-2 invasion. To further explore this, we analyzed the contact ratios of residues on the RBD bound to ACE2 (derived from MD simulations), to the aptamer Ta (derived from MD simulations), or to the neutralizing antibodies (from all available crystal structures deposited in PDB). And the results indicate that the contact regions on the RBD bound by the aptamer Ta, ACE2 or the neutralizing antibodies were

largely similar (Fig. 3A). Additionally, we analyzed the electrostatic potential distribution on the RBD surface to assess how the highly negatively charged RNA aptamer interacts with the RBD (see Fig. 3A). The positively charged regions on both sides of the RBD promote the aptamer binding. To quantitatively analyze the binding complexes, we defined a key residue as the one with a contact ratio greater than 0.5. We found that ACE2 binds to the RBD using 12 key residues, such as LYS417, LEU455, PHE456, ALA475 (see Fig. 3B) while the neutralizing antibodies engage 19 key residues on the RBD, including all of ACE2's key residues except for THR500. This higher number of key residues corresponds with the stronger binding ability of neutralizing antibodies compare to ACE2, enhancing their ability to block SARS-CoV-2 invasion. For the aptamer Ta, we identified 27 key residues, most of which overlap with those of ACE2 and neutralizing antibodies. For instance, nine key residues on the RBD are shared by ACE2, neutralizing antibodies, and the aptamer Ta (see Fig. 3B), suggesting that the aptamer Ta may have comparable or even better binding ability to the RBD compared to the neutralizing antibodies.

We performed binding energy estimations using MM/GBSA calculation and competitive binding assays with EMSA experiments to further compare the binding abilities of the aptamer Ta and neutralizing antibodies to the RBD. We selected three potent antibodies that can effectively neutralize SARS-CoV-2 infection, including antibody P2C-1F11 [35] (PDB id: 7CDI), 2H2 [36] (PDB id: 7DK4), and S2E12 [34] (PDB id: 7K4N), and calculated their binding energy to the RBD, along with that of ACE2. As shown in Fig. 3C, antibodies P2C-1F11 (-119.24±7.86 kcal/mol) and 2H2 (-107.18±8.89 kcal/mol) have lower binding energies (indicating stronger binding abilities) to RBD than ACE2 (-94.94±11.95 kcal/mol), while antibody S2E12 (-85.77±8.55 kcal/mol) exhibits a similar binding energy to ACE2. These findings confirm the reliability of our binding energy estimations using MM/GBSA calculation. The aptamer Ta, with a binding energy of -142.93±7.65 kcal/mol, has a significant lower binding energy than both ACE2 and the three antibodies, indicating a stronger binding ability of the aptamer Ta to the RBD. We further tested the binding ability of the aptamer Ta with a commercial neutralizing antibody (SinoBiological, Cat: 40592-R001, termed 40592-R001) that targets the RBD of the SARS-CoV-2 spike protein. The antibody

40592-R001 demonstrated high affinity for the RBD in our native gel protein-protein binding assays (Fig. 3D). We then added the antibody 40592-R001 into a mixture of the aptamer Ta and RBD, monitoring the changes in the formation of aptamer-RBD complex via fluorescence intensity in the ESMA experiments. In principle, if the aptamer Ta binds more strongly to RBD than the antibody 40592-R001, the presence of the antibody should not disrupt the formation of the aptamer-RBD complex; otherwise, the amount of the aptamer-RBD complex would be reduced. Notably, the Ta-RBD complex formation remained unchanged after adding the antibody (Fig. 3E), suggesting that the aptamer Ta indeed has a comparable binding compared to 40592-R001. These results also confirmed that the aptamer Ta binds to the same site on the RBD as the 40592-R001 antibody. In conclusion, the aptamer Ta presents a promising alternative, or at least a complementary approach, to conventional antibody-based neutralizing therapies against COVID-19. These findings prompt us to further optimize the aptamer Ta to enhance its binding ability to the RBD of the SARS-CoV-2 spike protein.

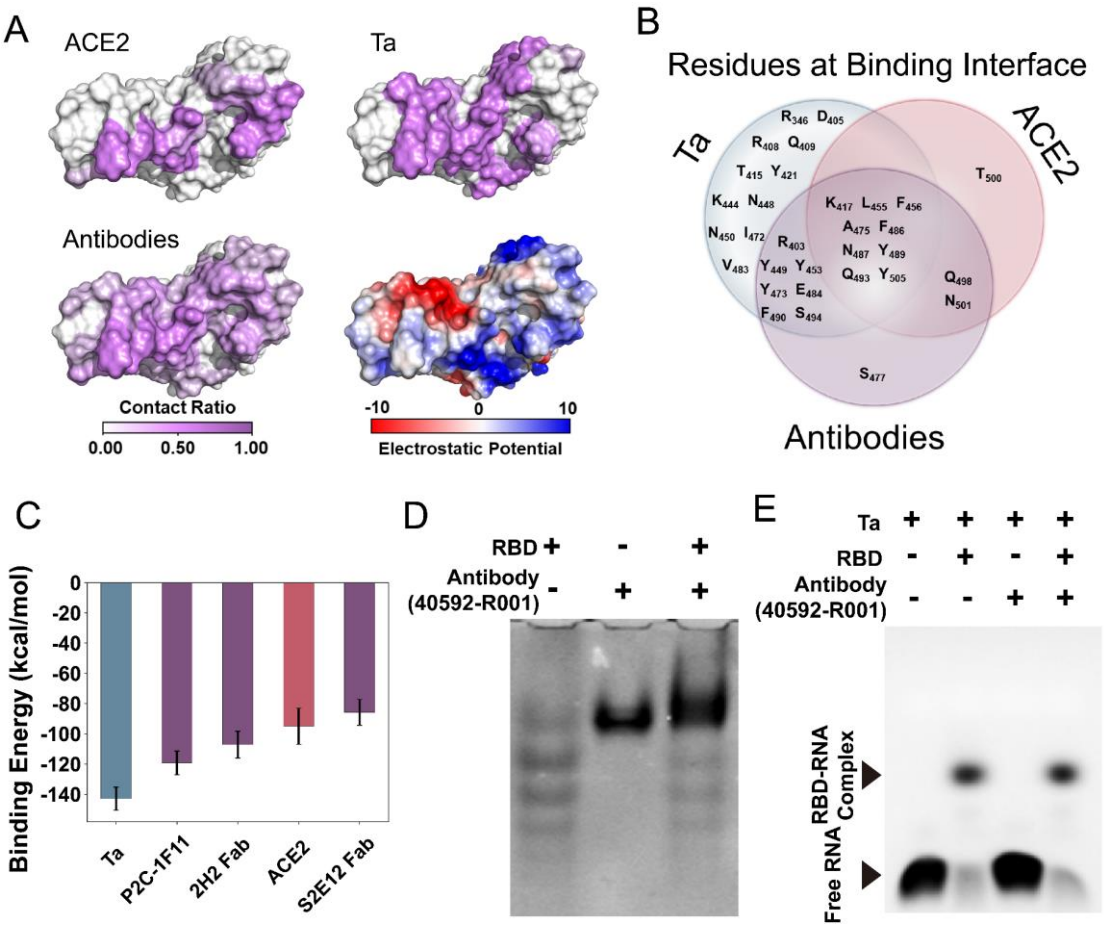

**Fig. 3.** The aptamer Ta exhibits a comparable binding capability to RBD compared to neutralizing antibodies. (A) Contact ratios of residues on the RBD by ACE2 (derived from MD simulations), the aptamer Ta (derived from MD simulations), and neutralizing antibodies (derived from all available crystal structures deposited in PDB). For reference, the electrostatic potential distribution on the RBD surface generated by PyMOL (version 2.3.5) program was also shown. (B) Key residues on RBD (with contact ratio larger than 0.5 in plane (A)) contacted by ACE2, the aptamer Ta, and neutralizing antibodies were displayed as a Venn diagram. (C) Binding energies estimated by MM/GBSA calculations for ACE2, the aptamer Ta, and three representative antibodies (P2C-1F11, 2H2, and S2E12) binding to RBD. (D) Binding ability of a commercial antibody (40592-R001) to RBD was examined by a protein binding assay. Three micrograms of RBD protein and 3 μg of the antibody individually or mixed together were run a native-PAGE gel, which then was stained with Coomassie brilliant blue. (E) The RBD binding abilities of the aptamer Ta and commercial antibody 40592-R001 were compared by EMSA competitive binding experiments. The aptamer-RBD complex bands were shown after running on an agarose gel following the incubation of 40 μM RBD protein, 5 μM aptamer Ta and the antibody 40592-R001.

### 2.4 Structure-based rational design of the aptamer Ta to improve binding affinity

Since the SELEX screening process explored only a limited sequence space, we anticipated that the SELEX-derived aptamer Ta could be further optimized *via* rational mutation scanning to improve its binding affinity to the RBD of the SARS-CoV-2 spike protein. In principle, nucleotide mutations in the aptamers can affect the binding affinity, which can be characterized by changes in binding free energy ($\Delta\Delta G$), and we accurately assessed these changes using FEP calculations. A negative change in binding free energy ($\Delta\Delta G<0$) indicates improved binding affinity, while a positive change ($\Delta\Delta G>0$) means a decrease in binding affinity. FEP is regarded as one of the most rigorous and reliable methods in estimating binding free energy changes, achieving high accuracy in identifying key residues and their mutational effects for many protein-protein, protein-ligand, and protein-RNA complexes, with results comparable with experiments [21,37]. However, unlike protein and peptide drugs, the structure of RNA molecules is very sensitive to its nucleotide composition and even a single mutation may cause significant changes in its secondary (2D) structure, potentially affecting its binding modes. To address this, we employed secondary

structure analysis (SSA) to examine the structural similarity before and after nucleotide mutation. Moreover, we used MFold [30] to predict the 2D structure with the lowest free energy for each mutated sequence. Structural similarity between the wild-type (WT) Ta sequence and its mutants was quantified by base-pair (BP) similarity, defined as the ratio of shared base pairs ($N_{shared}$) to the total base pairs in the WT sequence ($N_{WT}$). Overall, as summarized in Fig. 4A, we combined rational mutation scanning, SSA, and FEP to design Ta analogues with enhanced binding affinity.

Based on the binding complex between the aptamer Ta and RBD constructed above, we selected 16 vital nucleotides (A14, C23-G34, A40, and A42-G43) on the aptamer Ta that showed high contact frequency with RBD (see Fig. 4B) for exhaustive single-mutation scanning. These selected nucleotides play a crucial role in Ta-RBD binding interactions, and their mutagenesis studies may aid in designing analogues with improved binding affinity. After SSA based on BP similarity, we found that most single mutations preserved a WT-like 2D structure (Fig. 4C) with BP similarity greater than 0.9 (such as $Ta^{G34C}$ in Fig. S8), but several single mutations cause significant rearrangement of their 2D structure with a BP similarity less than 0.9 (for instance, $Ta^{U30G}$ in Fig. S8). Mutations with a BP similarity greater than 0.9 were subjected to FEP calculations to further evaluate their impact on binding free energy changes, with the results shown in Fig. 4D and Table S3.

For nucleotides in the apical loop of the aptamer Ta (see Figs. 1B and 2E), mutation to any other base types resulted in an increased binding free energy and weakened binding affinity, such as $\Delta\Delta G$=1.94±0.43 kcal/mol for G25C, 0.66±0.16 kcal/mol for U27G, and 4.66±0.76 kcal/mol for C29G. This suggests that G25, U27 and C29 are already optimal choices for RBD binding. The previous study [16] highlighted the importance of the apical loop in directing the aptamer Ta's binding to the RBD, and truncation of this apical loop was proven to significantly reduce its binding ability. Our FEP calculations align with this observation. Next, though the end stem part of the aptamer Ta also tightly binds to the RBD (see Figs. 2E and 4B), base mutations in this region had minimal effects on binding affinity, likely due to dominant electrostatic interactions between the negatively charged phosphate groups and the positively charged residues of RBD. For instance, A42G showed a $\Delta\Delta G$ *of* -0.22±0.56 kcal/mol and G43U a $\Delta\Delta G$ of 0.37±0.30 kcal/mol. Furthermore, since the bulge

part of the aptamer Ta serves as a linker to regulate the saddle-like binding shape with RBD (see Figs. 2E and 4B), nucleotide mutations in this region may also affect binding ability despite their low contact frequency with the RBD. Indeed, disruption of the C23-G34 base pair improved the binding affinity of the aptamer to RBD, with C23A, C23G, and C23U yielding $\Delta\Delta G$ values of -2.15±0.43, -2.49±0.49, and -1.96±0.58 kcal/mol, respectively, and G34A, G34C, and G34U showing $\Delta\Delta G$ values of -2.50±0.28, -3.05± 0.26, and -2.65±0.33 kcal/mol, respectively (see Fig. 4D and Table S3). Ultimately, we generated six candidate sequences ($Ta^{C23A}$, $Ta^{C23G}$, $Ta^{C23U}$, $Ta^{G34A}$, $Ta^{G34C}$, and $Ta^{G34U}$) predicted to have stronger binding affinity to RBD than WT aptamer Ta.

We conducted EMSA experiments to validate the binding ability of these designed aptamers. Preliminary binding results of the RBD-RNA complex bands showed that five of the six designed sequences ($Ta^{G34C}$, $Ta^{G34U}$, $Ta^{G34A}$, $Ta^{C23A}$, and $Ta^{C23U}$) exhibited higher fluorescence intensity than WT aptamer Ta (see Fig. 4E). Further, the binding curves with different RBD concentrations and the resulting dissociation constant ($K_d$) measurements confirmed the superior binding capabilities of these five sequences compared to WT aptamer Ta (see Fig. S9 and Table S2). As shown in Fig. 4F, except for $Ta^{C23G}$, the binding free energy changes derived from EMSA experiments ($\Delta\Delta G_{exp}$) were consistent with FEP calculations ($\Delta\Delta G$) for the remaining five designed aptamers. Here, the experimental binding free energy change is $\Delta\Delta G_{exp} = -k_B T \ln(K_d^{wt}/K_d^{mut})$, where $K_d^{wt}$ and $K_d^{mut}$ are dissociation constants for WT aptamer Ta and the designed sequence, respectively, and $k_B$ is the Boltzmann constant and $T = 310\,K$. The magnitude of the binding free energy changes generated from FEP calculations tend to be greater, likely due to limitations in the force field parameters. Nonetheless, the high success rate (0.83, 5/6) achieved in this structure-based rational design process underscores the reliability of the RBD-aptamer Ta complex model proposed in Fig. 2E. These results highlight the potential of our CAAMO framework as an effective tool for optimizing aptamer binding affinity.

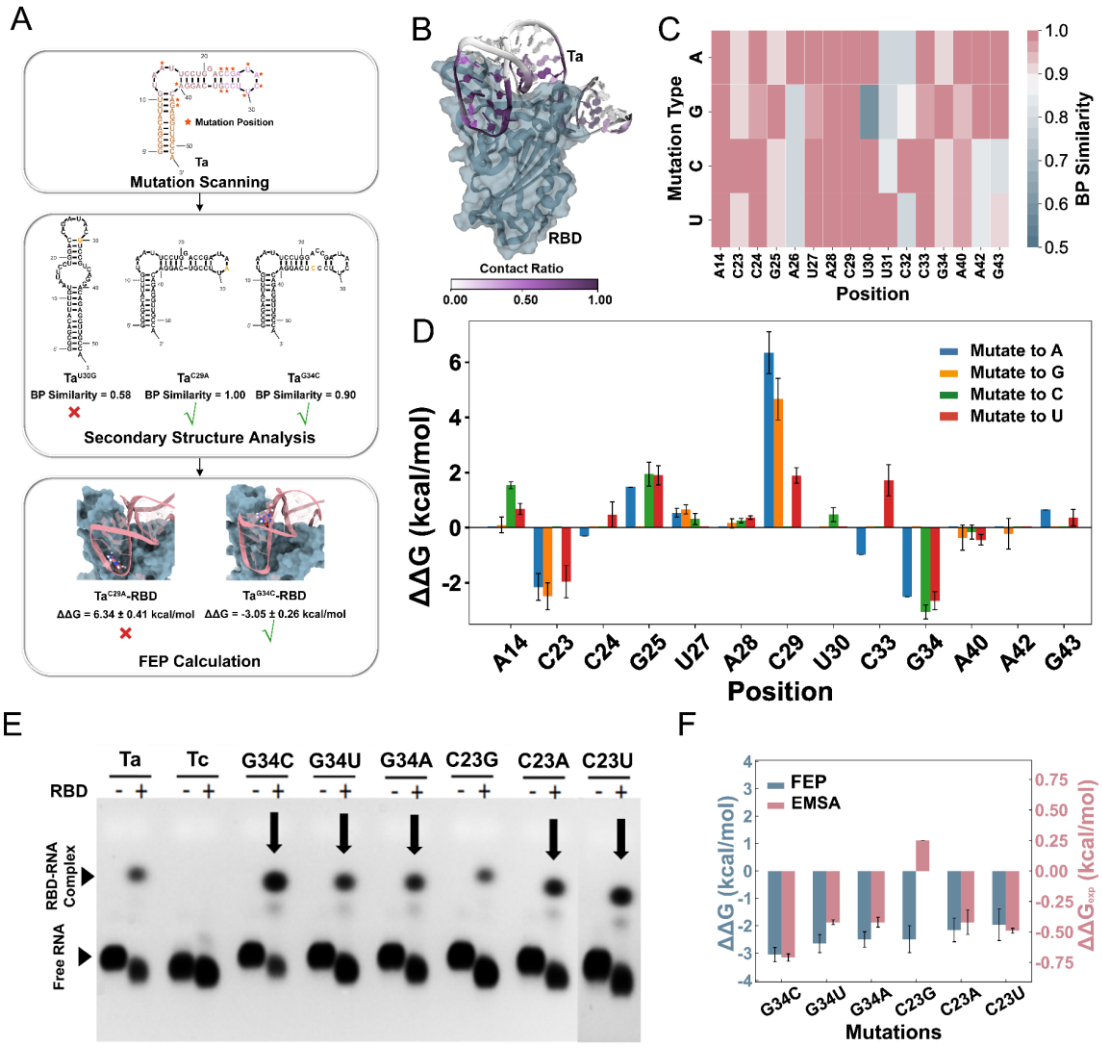


**Fig. 4.** Structure-based rational design of Ta analogues with improved binding affinities and their experimental validation. (A) Flowchart illustrating the combination of rational mutation scanning, secondary structure analysis (SSA), and free energy perturbation (FEP) to optimize aptamer binding affinities. (B) Contact ratios of nucleotides on the aptamer Ta bound to RBD (dark blue surface). Data were collected from three independent MD simulations. (C) SSA based on BP similarity for 16 selected nucleotides mutated to other three bases. Definition of BP similarity was given in the main text. Only mutations with BP similarity greater than 0.9 were subjected to further FEP calculations. (D) The binding free energy changes assessed by FEP calculations for selected single mutations. Data represent mean±s.d. collected from five independent simulations. (E) EMSA results of Ta, Tc, and 6 designed candidate sequences bound to the RBD of SARS-CoV-2 spike protein. The aptamer-RBD complex bands were detected by running an agarose gel after incubation of 40 μM of RBD protein with 5 μM indicated aptamer variant. (F) Comparison of the binding free energy changes derived from FEP calculations (*ΔΔG*, left scale)

and EMSA experiments ($\Delta\Delta G_{exp}$, right scale) for 6 designed candidate aptamers. Data represent mean±s.d. collected from 5 (FEP)/three (EMSA) independent replications.

### 2.5 Designed aptamer $Ta^{G34C}$ shows excellent binding ability to RBD

Among the five designed candidate sequences, both the FEP calculation and EMSA experiment confirmed that aptamer $Ta^{G34C}$ has the highest binding affinity to the RBD of SARS-CoV-2 spike protein. As shown in Fig. 5A, the dissociation constant derived from EMSA experiments for aptamer $Ta^{G34C}$ is $K_d$=33.5±1.6 μM, which is approximately 3.3-fold higher compared to WT Ta ($K_d$=110.7 μM). The remaining four designed aptamers exhibited approximately a 2-fold increase in binding affinity relative to WT Ta, with $K_d$=55.5±4.1 μM for $Ta^{G34A}$ and $K_d$=55.7±9.8 μM for $Ta^{C23A}$ (see Fig. S9 and Table S2).

We performed MD simulations on the $Ta^{G34C}$-RBD binding complex to explore the molecular mechanism underlying the improved binding affinity of aptamer $Ta^{G34C}$. As shown in the Fig. 5B, the C23-G34 base pair is located in the bulge region of WT Ta, which plays a critical role in regulating its saddle-like binding shape. We speculated that disrupting the C23-G34 base pair through mutation (e.g., G34C) could reduce the strain during aptamer binding to the RBD. The G34C substitution can alter the binding environment of adjacent nucleotides, such as U35, allowing them to form tighter contacts with the RBD loop region (from GLN474 to TYR489). For instance, our simulations showed that the distance between nucleotide U35 and PHE486 is shorter in the $Ta^{G34C}$-RBD complex than that in the WT Ta-RBD complex (see Fig. 5C). This reduced distance remains stable throughout the MD simulations, indicating that U35 and PHE486 form a stable π-π stacking interaction after the G34C substitution. Additionally, the distance between C34 and PHE486 in $Ta^{G34C}$-RBD is also closer compared to G34 and PHE486 in WT Ta-RBD. These findings support our hypothesis regarding the improved binding affinity in aptamer $Ta^{G34C}$ and provide a basis for further in silico design of new aptamers.

We also conducted competitive binding experiments to compare the binding capacities of the designed aptamer $Ta^{G34C}$ and a commercial monoclonal SARS-CoV-2 neutralizing antibody [38] against the RBD. As shown in Fig. 5D and Fig. 1F, when the monoclonal SARS-CoV-2 neutralizing antibody 40592-R001 was added to the WT Ta-RBD complex, the

fluorescence intensity of the complex band gradually decreased, indicating that the antibody at high concentrations can partially replace the aptamer Ta in the Ta-RBD binding complex and has comparable, though weaker, binding ability. However, when the same antibody was added to the designed $Ta^{G34C}$-RBD complex, the fluorescence intensity of the complex bands remained nearly unchanged at all tested antibody concentrations, indicating that the binding affinity of $Ta^{G34C}$ is significantly stronger than that of the monoclonal SARS-CoV-2 neutralizing antibody or WT aptamer Ta.

In summary, the combined results from FEP calculated and EMSA measured binding affinities, binding molecular mechanism analysis, and antibody competitive binding assays clearly demonstrate that the designed aptamer $Ta^{G34C}$ exhibits excellent binding ability to RBD. These findings highlight the importance of optimizing SELEX-derived aptamers through structure-based rational design to enhance their binding affinity.

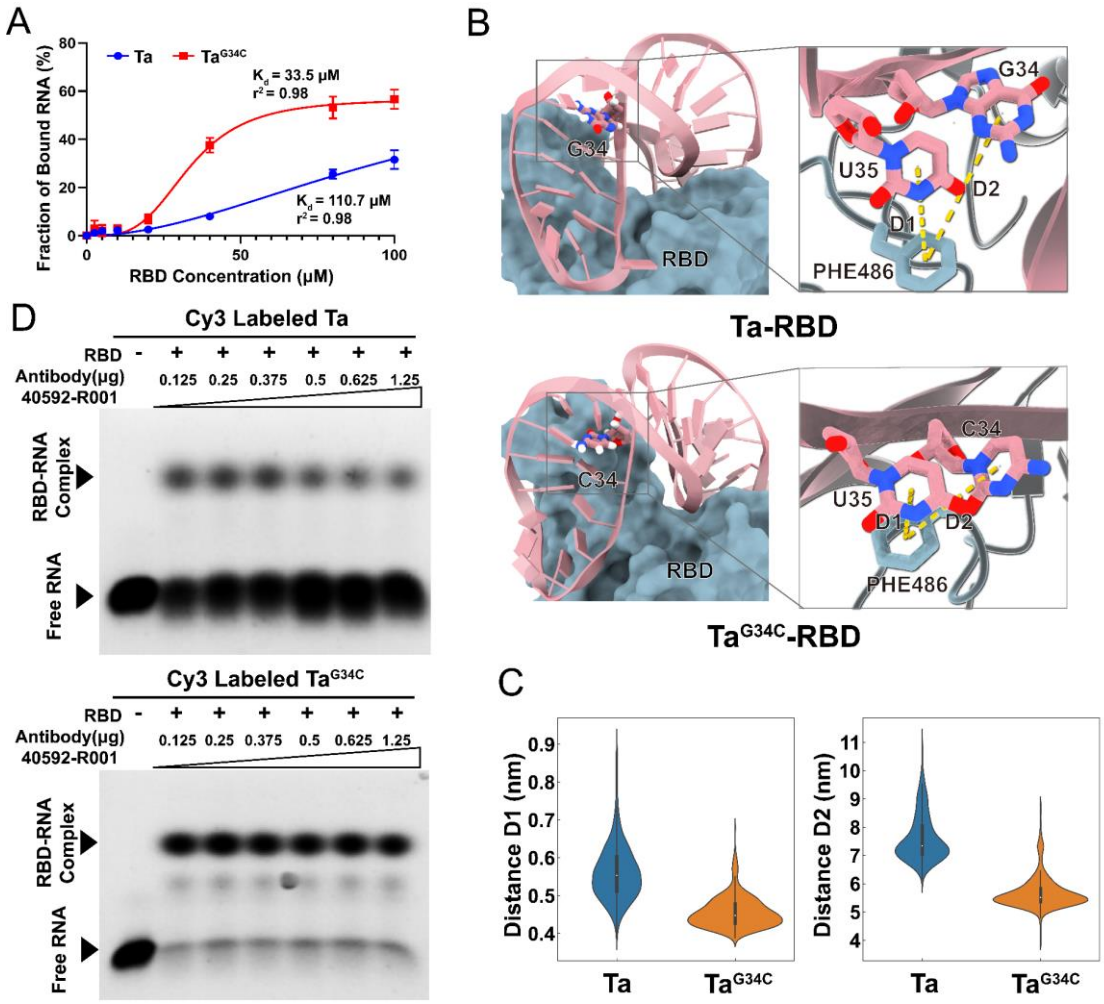

**Fig. 5.** Designed aptamer $Ta^{G34C}$ showed superior binding ability than WT Ta or antibody. (A) Binding curves and the resultant dissociation constants ($K_d$) for WT Ta and the designed $Ta^{G34C}$ with RBD proteins. The $K_d$ values were calculated from the EMSA image quantification with s.d. from three independent experiments. (B) Comparison of the Ta-RBD and $Ta^{G34C}$-RBD binding complexes. RNA is shown as a dark red ribbon while RBD is displayed as a dark blue surface. Zoom-in is the nucleotide G34 (WT Ta)/C34 ($Ta^{G34C}$) and its surrounding nucleotide U35 and residue PHE486 shown in sticks. Atoms oxygen and nitrogen are colored in red and blue, respectively, and carbon atoms are colored by their locations. Definitions of two related distances (D1 and D2) are also displayed. (C) Violin plots of distributions of two related distances (D1 and D2) between selected RNA nucleotide-protein residue in Ta-RBD and $Ta^{G34C}$-RBD binding complexes, respectively. Data were collected from three independent MD simulations. (D) EMSA images of competitive binding experiments to characterize the RBD binding abilities of RNA aptamers (WT Ta and $Ta^{G34C}$) and the commercial monoclonal SARS-CoV-2 neutralizing antibody 40592-R001. The aptamer-RBD complex bands were showed by running an agarose gel after incubation of 40 μM of RBD protein and indicated aptamer with varying concentrations of the antibody 40592-R001. Results showed that $Ta^{G34C}$, but not WT Ta, exhibited a higher binding affinity to the RBD proteins than that of the antibody.

## 3. Discussion and conclusion

Determining the 3D structure of RNA aptamer-target protein complexes is crucial for understanding the binding mechanism and optimizing binding efficacy for various applications. Despite the discovery of aptamers for over 1,100 target proteins [39], only a limited number of aptamer-protein complex 3D structures are available (only 119 deposited in the PDB). The scarcity of experimentally determined aptamer-protein structures is primarily due to the inherent flexibility of RNA molecules and the high cost of experimental procedures. However, with recent advancement in protein and RNA 3D structure prediction [32,40,41] and improvements in atomic force field parameters [42,43], as well as the increased availability of high-performance computing resources, computational modeling of RNA aptamer-protein complexes has become increasingly promising. In this work, we developed a multi-strategy computational approach to determine the most probable RNA aptamer-protein binding conformation. Our approach integrates RNA 3D structure prediction,

ensemble docking, clustering, and binding capacity assessment, with a critical focus on identifying the lowest-energy binding conformation using various energy assessment methods. We believe that the predicted binding conformation reliably represents the closest approximation among the predicted ensemble, although it may not perfectly mimic the native structure conformation in all systems.

We found that the aptamer Ta binds to the RBD of SARS-CoV-2 spike protein in a saddle-like shape, where the apical loop and end stem parts bind to two opposite sides of the RBD, while the bulge serves as a connector and regulates the bending angle. Theoretical analysis and experimental validation confirmed that this binding model is reliable. With this 3D structure, we were able to conduct structure-based rational optimization of the aptamer sequence to improve its binding affinity and then to perform a head-to-head comparison between the RNA aptamer and neutralizing antibodies.

The development of aptamer-based RNA therapeutics involves iterative rounds of design, testing, and optimization. A key step in this process is optimizing the original aptamer sequence to design a series of analogues with comparable or even improved binding affinities to the target protein. *In silico* structure-based aptamers design, which involves nucleotide mutation, insertion, deletion and their binding affinity assessment, is gaining prominence [44]. Our proposed framework, CAAMO, enables efficient aptamer lead optimization requiring only the aptamer's nucleotide sequence as the input information. Unlike peptide and protein counterparts, RNA structures are highly sensitive to nucleotide composition, and even a single mutation may cause rearrangement in its secondary structure. To address this issue, we employed secondary structure analysis to evaluate the effect of mutations on RNA folding and only submitted those maintaining a similar folding structure to subsequent FEP-guided binding free energy evaluations.

For the Ta-RBD binding complex, mutation scanning revealed that mutations of nucleotides in direct contact with the RBD showed negligible or increased binding free energy changes, whereas mutations in the middle portion, which has fewer contacts with RBD but regulates the bending angle of RNA aptamer, resulted in reduced binding free energy and improved binding affinity. These findings suggest that during structure-based optimization, attention must be paid not only to the nucleotides at the binding interface but

also to the regulatory nucleotides that affect the binding shape. Our CAAMO framework generated six optimized candidate sequences, and EMSA experiments confirmed that five of them exhibited significantly stronger binding affinity to RBD than WT Ta (about 2 to 3.3-fold improvement). This high success rate (~83%) validates the reliability of the putative 3D binding model and demonstrates the effectiveness of our computational framework. Furthermore, since the current study only considered single-nucleotide mutations, other design strategies such as double mutations, multiple mutations, nucleotide insertions and deletions should be explored in the future studies.

Aptamers, often referred to as "chemical antibody", offer a valuable comparison to real antibodies in terms of binding properties, yet such comparative analyses are currently limited. The RBD of SARS-CoV-2 spike protein is an ideal target for these comparisons, as binding modes for both the aptamer Ta and neutralizing antibody-RBD complexes are available. Our computational and experimental studies showed that the aptamer Ta has comparable binding abilities to the RBD compared to popular neutralizing antibodies. Analysis of the underlying binding details revealed that more critical RBD residues are involved in aptamer Ta-RBD interaction, while MM/GBSA calculations and competitive binding assays confirmed the stronger binding capacity of the aptamer Ta. Specifically, a newly designed aptamer ($Ta^{G34C}$) exhibited excellent RBD binding ability, surpassing both WT Ta and commercial antibodies. Given the concerns surrounding antibody-dependent enhancement (ADE), which could worsen COVID-19 outcomes by increasing virus infectivity and virulence [45], there is an urgent need to develop smaller and safer molecules to neutralize the virus and complement existing treatment modalities. Furthermore, the previous study [16] has claimed that WT Ta aptamer can efficiently block viral infection at low concentration and provide a promising lead for the detection and treatment of SARS-CoV-2 and emerging variants. Therefore, the designed aptamer $Ta^{G34C}$, as well as other four successful candidates, provides a promising alternative to antibodies for fighting COVID-19.

## 4. Materials and Methods

### 4.1 Generation of aptamer 3D conformation ensemble

To predict the 3D structures of the aptamers Ta (5'-GGCGACAUUU GUAAUUCCUG

GACCGAUACU UCCGUCAGGA CAGAGGUUGC CA-3') and Tc (5'-GGUCCUGGAC CGAUACUUCC GUCAGGACCA-3'), five state-of-the-art RNA 3D structure prediction tools were employed: IsRNA2 [32], FARFAR2 [31], SimRNA [33], iFoldRNA [46], and RNAComposer [47]. The predicted secondary structure from Mfold [30], along with the aptamer sequence, were used as input for 3D structure prediction. IsRNA2 was downloaded and ran locally following the identical procedure as in the previous studies [48]; FARFAR2 was executed in Rosetta 3.12 using the *rna_denovo* application with default parameters; web-based version of SimRNA, iFoldRNA, and RNAComposer were assessed, and their default parameters were used. The top 5 predictions generated by each program (25 structures in total) were collected to construct a 3D conformation ensemble of a given RNA aptamer.

### 4.2 Generation of aptamer-RBD binding complex pool

The binding complex pool for aptamer Ta-RBD was generated using multiple popular docking tools, including HADDOCK [49], HDOCK [50], ZDOCK [51], and RosettaDock [52]. Rosettadock was performed locally using Rosetta software, while the other three docking tasks were completed on their respective webservers. The RBD structure [23] (PDB id: 6LZG) refined by MD simulation, and 25 RNA 3D conformations predicted in the previous step were used as the receptor and ligands for docking, respectively. The top 10 binding poses from each docking were recorded, yielding a total of 1,000 (4×25×10) aptamer Ta-RBD binding complexes in the pool. Of these, 879 conformations shared the same binding interface of ACE2-RBD complex. Then, these 879 conformations were clustered into 6 major groups, each representing a significantly distinct binding conformation. The binding complex pool for Tc-RBD interaction (used as a negative control) was constructed following the same procedure.

### 4.3 Molecular dynamics simulations

All-atom MD simulations were performed using Gromacs [53] (version 2021.5). For each system, a water box with at least 1.5 nm distance from the surface of the complex was used to solvate the systems, and NaCl ions were added to achieve a physiological concentration of 150 mM after neutralizing the system. The Amberff14SB force field was

used for protein and RNA. Water molecules were described by TIP3P model [54] and Li and Merz ion parameters [55] were used. The periodic boundary conditions were applied in all three dimensions. The particle mesh Ewald (PME) method was used to compute the long-range electrostatic interactions while the vdW interactions were truncated at 1.5 nm. The LINCS algorithm was adopted to constrain the H-bonds to allow an integration timestep of 2 fs. Before MD productions, an energy minimization, 100 ps NVT, and 10 ns NPT simulations with a temperature of T = 310.15 K and pressure of 1 atm was executed sequentially to equilibrate the simulation box. To prevent unexpected structural deviations in the beginning, position restraints on backbone atoms of RNA and protein were performed in the NVT and NPT simulations. After that, series of MD simulations were conducted in the NPT ensemble with velocity-rescaled Berendsen thermostat: (1) 500ns MD simulations were performed for the RBD. Snapshots extracted from the last 300 ns MD trajectories (600 snapshots recorded in the duration of 500 ps) were clustered based on linkage method with a 0.2 nm RMSD cutoff to obtain the most probable receptor conformation for docking. (2) 100 ns simulations were performed on the representative Ta/Tc-RBD complex conformations obtained from the docking. 5 mM $MgCl_2$ was added to the simulation system. These trajectories were used to calculate RMSD and MM/GBSA. (3) 100 ns simulations were performed on the ACE2-RBD complex (PDBID: 6LZG) and three antibody-RBD complexes (PDBID: 7CDI, 7DK4 and 7K4N). These also were used to perform MM/GBSA. (4) 500ns simulations were conducted for the $Ta^{G34C}$-RBD and Ta-RBD complexes. 5 mM $MgCl_2$ was added to the simulation system. A summary of MD simulations performed in this study was given in Table S1. The 3D structure models were rendered using UCSF ChimeraX (version 1.6.1) programs.

### 4.4 MM/GBSA calculations

In this study, binding energies ($\Delta G$) of ACE2, aptamers (Ta and Tc), and several neutralizing antibodies to RBD were assessed using the end-point Molecular Mechanics Generalized Born Surface Area (MM/GBSA) method, implemented through the gmx_MMPBSA software [56]. Briefly, $\Delta G$ was calculated by summing up the changes in electro-static energies ($\Delta E^{ele}$), the van der Waals energies ($\Delta E^{vdW}$), the electrostatic solvation energy ($\Delta G^{GB}$, polar contribution), the nonpolar contribution ($\Delta G^{SA}$) between the solute and

the continuum solvent, and conformational entropy ($-T\Delta S$) upon ligand binding. The dielectric constants for the solute and solvent were set to 10 and 78.5, respectively. The OBC solvation model (igb = 8) was used. The interaction entropy method was used to calculate the conformational entropy ($-T\Delta S$). Additionally, the lowest energy conformations from the complexes pool clustering were selected based on single-frame conformational MM/GBSA. The static complex structure was employed to calculate the enthalpy using MM/GBSA.

### 4.5 SMD and rupture works

Constant velocity steered molecular dynamics (SMD) was performed to calculate the rupture work required to separate the bound aptamer from the RBD. A steered velocity of 0.1 nm/ns was applied to the center-of-mass (COM) of the aptamer, while keeping the COM of RBD fixed, using the COM distance between aptamer (or antibody) and RBD as the collective variable. Four replicate SMD simulations (about 50 ns simulation time each) were performed for each binding complex. All simulation parameters remain the same as those used in the MD simulations. The force spectra were recorded during the SMD simulations at a 0.1 ps intervals. The rupture works were obtained through integration of the force spectra over the COM distance.

### 4.6 FEP calculations

The binding free energy changes resulting from point mutations of key bases at the interface between the aptamer Ta and RBD were calculated using the free energy perturbation (FEP) method. We estimated the binding free energy changes for single nucleotide mutation in both the bound state (Ta and RBD complex) $\Delta G_{bound}$ and the free state (the aptamer Ta only) $\Delta G_{free}$ using Gromacs 2021.5. Thus, the binding free energy change caused by nucleotide mutation was estimated as $\Delta\Delta G_{calc}=\Delta G_{bound}-\Delta G_{free}$. For each single nucleotide mutation, the dual-topology file was prepared in a pmx-like manner based on the Amberff14SB force field and eighteen λ windows (0.0, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.8, 0.85, 0.9, 0.95, 0.99, and 1.0) with 1 ns/window were used. The vdW and electrostatic interactions were transferred simultaneously during simulations and the soft-core potentials (α=0.3) were used. For each mutation, five independent replicas starting from different

conditions were performed for sufficient sampling and at least 180 ns (1 ns×18 windows×5 runs×2 states) simulation time was generated, which result in reasonable convergence in the free energy calculations. The Gromacs bar analysis tool was used to estimate the binding free energy changes.

### 4.7 EMSA experiments

The sequences of Ta, Tc, $Ta^{G34C}$, $Ta^{G34U}$, $Ta^{G34A}$, $Ta^{C23G}$, $Ta^{C23A}$, and $Ta^{C23U}$ aptamers are shown in Table S2. In these sequences, the uppercase letters A, U, G and C indicated the ribonucleotide. The EMSA was performed according to a previous protocol [48] with a minor modification. Synthesized 5′ end Cy3-labelled RNAs were resuspended with the RNase-free $H_2O$ to a concentration of 100 μM. 5 μl of Cy3-labelled RNAs were annealed with the 5 μl 2×annealing buffer (20 mM Tris-HCl pH 7.5, 200 mM KCl) under a predefined procedure: 68°C for 5min, then annealing at -0.1°C/s to 25°C, and finally at 25°C for 5 min, then diluted to the final concentration of 5 μM. The SARS-CoV-2 Spike RBD protein was purchased from HuaBio (Cat: HA210064). Mouse monoclonal SARS-CoV-2 neutralizing antibody was purchased from Sino Biological (Cat: 40592-R001). 40 μM of RBD protein and 5 μM of annealed Cy3-labelled aptamers were mixed in the EMSA buffer (10 mM Sodium phosphate buffer pH 7.5, 1 U/μl SUPERase-In RNase Inhibitor [Thermo Fisher]). Mixtures were incubated at room temperature for 20 min.

For competitive binding experiments, 40 μM of RBP proteins, 5 μM of annealed Cy3-labelled RNAs and gradient amounts of SARS-CoV-2 neutralizing antibody (40592-R001) were mixed in the EMSA buffer and incubated at room temperature for 20 min. Next, the mixtures were added 6×loading buffer (15% Ficoll 400, 0.25% Bromophenol Blue, 0.25% Xylene cyanol, 1×TBE), then resolved on a native 0.8% agarose gel and imaged with iBright1500 (Thermo Fisher). The images were quantified with Image J software. The dissociation constant $K_d$ for RBD with aptamers were calculated using Prism 8 (GraphPad) software.s

**Acknowledgments**

We thank Zaixing Yang and Teng Xie for their helpful discussions. This work was supported

by funds from the National Key R&D Program of China (2021YFF1200404 to R.Z., 2024YFA1307502 to D.Z.), National Natural Science Foundation of China (Nos. 12474203 to D.Z., U1967217 to R.Z.), National Independent Innovation Demonstration Zone Shanghai Zhangjiang Major Projects (ZJZX2020014 to R.Z.), the Zhejiang Provincial Natural Science Foundation of China (No. LZ25A040001 to D.Z.), Starry Night Science Fund at Shanghai Institute for Advanced Study of Zhejiang University (SN-ZJU-SIAS-003 to R.Z., SN-ZJU-SIAS-006 to L.H.), National Center of Technology Innovation for Biopharmaceuticals (NCTIB2022HS02010 to R.Z.), Shanghai Artificial Intelligence Lab (P22KN00272 to R.Z.), Aoming Biomedical Research (AO-ZJU-SIAS-001 to R.Z.), and Zhejiang University Global Partnership Fund (188170+194452409/004 to R.Z.).

## CRediT authorship contribution statement

R.Z. designed research; Y.Y., and Y.J. performed all molecular simulations; L.Q. performed EMSA experiments under the supervision of Z.W.; Y.Y. contributed new reagents/analytic tools; Y.Y., Z.W., D.Z., and R.Z. analyzed data; Y.Y., L.Q., Z.W., D.Z., D.B., L.H., and R.Z. wrote the paper with contributions from all authors.

## Declaration of competing interest

The authors declare that they have no conflict of interest.

## References

[1] A.D. Ellington, J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands, Nature 346 (1990) 818–822. doi:10.1038/346818a0.

[2] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, Science 249 (1990) 505–510. doi:10.1126/science.2200121.

[3] A.D. Keefe, S. Pai, A. Ellington, Aptamers as therapeutics, Nat. Rev. Drug Discov. 9 (2010) 537–550. doi:10.1038/nrd3141.

[4] G. Zhou, G. Wilson, L. Hebbard, W. Duan, C. Liddle, J. George, et al., Aptamers: A promising chemical antibody for cancer therapy, Oncotarget 7 (2016) 13446–13463.

doi:10.18632/oncotarget.7178.

[5] H. Sun, Y. Zu, A highlight of recent advances in aptamer technology and its application, Molecules 20 (2015) 11959–11980. doi:10.3390/molecules200711959.

[6] S.-K. Guo, C.-X. Liu, Y.-F. Xu, X. Wang, F. Nan, Y. Huang, et al., Therapeutic application of circular RNA aptamers in a mouse model of psoriasis, Nat. Biotechnol. (2024) 1–11. doi:10.1038/s41587-024-02204-4.

[7] J.P. Dassie, X. Liu, G.S. Thomas, R.M. Whitaker, K.W. Thiel, K.R. Stockdale, et al., Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors, Nat. Biotechnol. 27 (2009) 839–846. doi:10.1038/nbt.1560.

[8] J. Zhou, J. Rossi, Aptamers as targeted therapeutics: current potential and challenges, Nat. Rev. Drug Discov. 16 (2017) 181–202. doi:10.1038/nrd.2016.199.

[9] J. Abatemarco, M.F. Sarhan, J.M. Wagner, J.-L. Lin, L. Liu, W. Hassouneh, et al., RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes, Nat. Commun. 8 (2017) 332–340. doi:10.1038/s41467-017-00425-7.

[10] Y. Ning, J. Hu, F. Lu, Aptamers used for biosensors and targeted therapy, Biomed. Pharmacother. 132 (2020) 110902–110922. doi:10.1016/j.biopha.2020.110902.

[11] The Eyetech Study Group, Anti-vascular endothelial growth factor therapy for subfoveal choroidal neovascularization secondary to age-related macular degeneration: Phase II study results, Ophthalmology 110 (2003) 979–986. doi:10.1016/S0161-6420(03)00085-X.

[12] G. Zhu, H. Zhang, O. Jacobson, Z. Wang, H. Chen, X. Yang, et al., Combinatorial screening of DNA aptamers for molecular imaging of HER2 in cancer, Bioconjug. Chem. 28 (2017) 1068–1075. doi:10.1021/acs.bioconjchem.6b00746.

[13] D.-L. Wang, Y.-L. Song, Z. Zhu, X.-L. Li, Y. Zou, H.-T. Yang, et al., Selection of DNA aptamers against epidermal growth factor receptor with high affinity and specificity, Biochem. Biophys. Res. Commun. 453 (2014) 681–685. doi:10.1016/j.bbrc.2014.09.023.

[14] S.E. Lupold, B.J. Hicke, Y. Lin, D.S. Coffey, Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen1, Cancer Res. 62 (2002) 4029–4033.

[15] E. Fang, X. Liu, M. Li, Z. Zhang, L. Song, B. Zhu, et al., Advances in COVID-19 mRNA

vaccine development, Signal Transduct. Target. Ther. 7 (2022) 1–31. doi:10.1038/s41392-022-00950-y.

[16] J. Valero, L. Civit, D.M. Dupont, D. Selnihhin, L.S. Reinert, M. Idorn, et al., A serum-stable RNA aptamer specific for SARS-CoV-2 neutralizes viral entry, Proc. Natl. Acad. Sci. 118 (2021) e2112942118. doi:10.1073/pnas.2112942118.

[17] S.H. El-Nabi, M. Elhiti, M. El-Sheekh, A new approach for COVID-19 treatment by micro-RNA, Med. Hypotheses 143 (2020) 110203–110205. doi:10.1016/j.mehy.2020.110203.

[18] T. Yang, Y.-C. Wang, C.-F. Shen, C.-M. Cheng, Point-of-care RNA-based diagnostic device for COVID-19, Diagnostics 10 (2020) 165–167. doi:10.3390/diagnostics10030165.

[19] C. Zhu, Z. Feng, H. Qin, L. Chen, M. Yan, L. Li, et al., Recent progress of SELEX methods for screening nucleic acid aptamers, Talanta 266 (2024) 124998–125018. doi:10.1016/j.talanta.2023.124998.

[20] A. Brown, J. Brill, R. Amini, C. Nurmi, Y. Li, Development of better aptamers: structured library approaches, selection methods, and chemical modifications, Angew. Chem. Int. Ed. 63 (2024) e202318665. doi:10.1002/anie.202318665.

[21] D.R. Bell, J.K. Weber, W. Yin, T. Huynh, W. Duan, R. Zhou, In silico design and validation of high-affinity RNA aptamers targeting epithelial cellular adhesion molecule dimers, Proc. Natl. Acad. Sci. 117 (2020) 8486–8493. doi:10.1073/pnas.1913242117.

[22] A.C. Stelzer, A.T. Frank, J.D. Kratz, M.D. Swanson, M.J. Gonzalez-Hernandez, J. Lee, et al., Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble, Nat. Chem. Biol. 7 (2011) 553–559. doi:10.1038/nchembio.596.

[23] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2, Cell 181 (2020) 894–904. doi:10.1016/j.cell.2020.03.045.

[24] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (2020) 1444–1448. doi:10.1126/science.abb2762.

[25] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, et al.,

SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2020) 271–280. doi:10.1016/j.cell.2020.02.052.

[26] B. Ju, Q. Zhang, J. Ge, R. Wang, J. Sun, X. Ge, et al., Human neutralizing antibodies elicited by SARS-CoV-2 infection, Nature 584 (2020) 115–119. doi:10.1038/s41586-020-2380-z.

[27] Y. Cao, B. Su, X. Guo, W. Sun, Y. Deng, L. Bao, et al., Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells, Cell 182 (2020) 73–84. doi:10.1016/j.cell.2020.05.025.

[28] D.D. Boehr, R. Nussinov, P.E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, Nat. Chem. Biol. 5 (2009) 789–796. doi:10.1038/nchembio.232.

[29] P. Qu, K. Xu, J.N. Faraone, N. Goodarzi, Y.-M. Zheng, C. Carlin, et al., Immune evasion, infectivity, and fusogenicity of SARS-CoV-2 BA.2.86 and FLip variants, Cell 187 (2024) 585–595. doi:10.1016/j.cell.2023.12.026.

[30] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415. doi:10.1093/nar/gkg595.

[31] A.M. Watkins, R. Rangan, R. Das, FARFAR2: improved *de novo* rosetta prediction of complex global RNA folds, Structure 28 (2020) 963–976. doi:10.1016/j.str.2020.05.011.

[32] D. Zhang, S.-J. Chen, R. Zhou, Modeling Noncanonical RNA Base Pairs by a Coarse-Grained IsRNA2 Model, J. Phys. Chem. B 125 (2021) 11907–11915. doi:10.1021/acs.jpcb.1c07288.

[33] M.J. Boniecki, G. Lach, W.K. Dawson, K. Tomala, P. Lukasz, T. Soltysinski, et al., SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction, Nucleic Acids Res. 44 (2016) e63. doi:10.1093/nar/gkv1479.

[34] M.A. Tortorici, M. Beltramello, F.A. Lempp, D. Pinto, H.V. Dang, L.E. Rosen, et al., Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms, Science 370 (2020) 950–957. doi:10.1126/science.abe3354.

[35] J. Ge, R. Wang, B. Ju, Q. Zhang, J. Sun, P. Chen, et al., Antibody neutralization of SARS-CoV-2 through ACE2 receptor mimicry, Nat. Commun. 12 (2021) 250–258. doi:10.1038/s41467-020-20501-9.

[36] C. Zhang, Y. Wang, Y. Zhu, C. Liu, C. Gu, S. Xu, et al., Development and structural basis of a two-MAb cocktail for treating SARS-CoV-2 infections, Nat. Commun. 12 (2021) 264–279. doi:10.1038/s41467-020-20465-w.

[37] R. Ahmed, Z. Omidian, A. Giwa, B. Cornwell, N. Majety, D.R. Bell, et al., A public BCR present in a unique dual-receptor-expressing lymphocyte from type 1 diabetes patients encodes a potent T cell autoantigen, Cell 177 (2019) 1583–1599. doi:10.1016/j.cell.2019.05.007.

[38] L.C. Nguyen, D. Yang, V. Nicolaescu, T.J. Best, H. Gula, D. Saxena, et al., Cannabidiol inhibits SARS-CoV-2 replication through induction of the host ER stress and innate immune responses, Sci. Adv. 8 (2022) eabi6110. doi:10.1126/sciadv.abi6110.

[39] M.H. Ali, M.E. Elsherbiny, M. Emara, Updates on aptamer research, Int. J. Mol. Sci. 20 (2019) 2511–2533. doi:10.3390/ijms20102511.

[40] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589. doi:10.1038/s41586-021-03819-2.

[41] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature 630 (2024) 493–500. doi:10.1038/s41586-024-07487-w.

[42] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713. doi:10.1021/acs.jctc.5b00255.

[43] M. Zgarbová, J. Šponer, M. Otyepka, T.E.I. Cheatham, R. Galindo-Murillo, P. Jurečka, Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA, J. Chem. Theory Comput. 11 (2015) 5723–5736. doi:10.1021/acs.jctc.5b00716.

[44] S. Gao, X. Zheng, B. Jiao, L. Wang, Post-SELEX optimization of aptamers, Anal. Bioanal. Chem. 408 (2016) 4567–4573. doi:10.1007/s00216-016-9556-2.

[45] W.S. Lee, A.K. Wheatley, S.J. Kent, B.J. DeKosky, Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies, Nat. Microbiol. 5 (2020) 1185–1191. doi:10.1038/s41564-020-00789-5.

[46] A. Krokhotin, K. Houlihan, N.V. Dokholyan, iFoldRNA v2: folding RNA with constraints, Bioinformatics 31 (2015) 2891–2893. doi:10.1093/bioinformatics/btv221.

[47] J. Sarzynska, M. Popenda, M. Antczak, M. Szachniuk, RNA tertiary structure prediction using RNAComposer in CASP15, Proteins Struct. Funct. Bioinforma. 91 (2023) 1790–1799. doi:10.1002/prot.26578.

[48] D. Zhang, L. Qiao, X. Lei, X. Dong, Y. Tong, J. Wang, et al., Mutagenesis and structural studies reveal the basis for the specific binding of SARS-CoV-2 SL3 RNA element with human TIA1 protein, Nat. Commun. 14 (2023) 3715–3726. doi:10.1038/s41467-023-39410-8.

[49] G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, et al., The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (2016) 720–725. doi:10.1016/j.jmb.2015.09.014.

[50] Y. Yan, D. Zhang, P. Zhou, B. Li, S.-Y. Huang, HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy, Nucleic Acids Res. 45 (2017) W365–W373. doi:10.1093/nar/gkx407.

[51] B.G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, Z. Weng, ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers, Bioinformatics 30 (2014) 1771–1773. doi:10.1093/bioinformatics/btu097.

[52] S. Lyskov, F.-C. Chou, S.Ó. Conchúir, B.S. Der, K. Drew, D. Kuroda, et al., Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE), PLOS ONE 8 (2013) e63906. doi:10.1371/journal.pone.0063906.

[53] S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, et al., Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, J. Chem. Phys. 153 (2020) 134110–134124. doi:10.1063/5.0018516.

[54] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935. doi:10.1063/1.445869.

[55] P. Li, L.F. Song, K.M.Jr. Merz, Systematic parameterization of monovalent ions

employing the nonbonded model, J. Chem. Theory Comput. 11 (2015) 1645–1657. doi:10.1021/ct500918t.

[56] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, J. Chem. Theory Comput. 17 (2021) 6281–6291. doi:10.1021/acs.jctc.1c00645.