Discovery of DNA Aptamers Targeting SARS-CoV-2 Proteins and Protein Binding Epitopes Identification for Label-Free COVID-19 Diagnostics

Suttinee Poolsup

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the Doctor of Philosophy degree

Department of Chemistry and Biomolecular Sciences

Faculty of Science

University of Ottawa

ABSTRACT

The spread of COVID-19 has affected billions of people across the globe with the challenges of less accurate and time-consuming detection approaches. Especially, new variants with various mutations in the spike protein can escape immune responses as well as a low viral load in clinical samples can trigger a false-negative test result. Due to high immunogenicity and abundant expression during viral infection, SARS-CoV-2 nucleocapsid (N) protein could be an alternative diagnostic marker, together with targeting the S1 subunit of SARS-CoV-2 spike protein. This study aimed to develop a label-free optical aptasensor fabricated with novel ssDNA aptamers to detect the N and S1 proteins. The aptamers selected using asymmetric emulsion PCR-SELEX and their binding affinity and cross-reactivity were characterized by using bio-layer interferometry. The tNSP3 aptamer (44 nt) was identified to bind the N protein of wild type and Delta and Omicron variants with high affinity (K_D in the range of 0.6 – 3.5 nM). Furthermore, the aptamers targeting the S1 subunit were also discovered in order to expand an additional tool for aptamer-based biosensors used in COVID-19 detection. The S1-tSP10 aptamer with 40 nt performed better affinity than other S1-targeting aptamers, and this aptamer possessed the K_D about 15 nM binding to the S1 protein of wild type and Omicron variant. Utilizing tNSP3 and S1-tSP10 aptamers to separately detect the N and S1 proteins spiked in human saliva evinced the potential of these aptamers in the function of aptamer-based BLI with LODs of 4.5 nM and 19 nM corresponding to tNSP3 and S1-tSP10 aptamers, respectively. To get insight into where the aptamers bind to their target proteins, mass spectrometry analysis was performed along with the molecular dynamic simulation. The identified epitope peptides are localized within the RNA-binding domain and CTD of the N protein as well as RBD of the S1 protein. Hence, we confirmed the performance of the aptamers as a label-free analytical tool for COVID-19 diagnosis, and also

knowing the binding epitopes can be essential for improving biomedical applications in therapeutics.

Acknowledgments

I would like to express my sincerest appreciation to Dr. Maxim Berezovski for the opportunity to work and study in his laboratory. I am grateful for his support and research direction for almost 5 years. Particularly, I would like to extend my deepest appreciation to my committee members, Dr. Paul Mayer, Dr. Jeffrey W. Keillor and Dr. Maria DeRosa for their time and expertise in guiding my graduate studies. Their engagement and advice greatly contributed to the success that I achieved in my doctoral studies. I would like to acknowledge Dr. Zoran Minic for his assistance in mass spectrometry in our collaboration with John L. Holmes Mass Spectrometry Facility. Also, Dr. Bob Dass and Dylan Tanner for generously providing technical assistance and training for bio-layer interferometry. Moreover, I would like to thank former and current lab members: Dr. Pabir Kulabhusan, Nico Hüttmann, Emil Zaripov, Abdullah Khraibah and other members for kindly assisting with data analysis and experimental discussion. I wish everyone all the best for their future studies and career paths.

This research would not have been possible without financial assistance from the Department of Chemistry and Biomolecular Sciences in the form of doctoral scholarship, bursaries and especially the research and teaching assistance scholarship which have helped me with living expenses throughout my PhD study.

Most importantly, I would like to express my special thanks to my parents, my family and Aldo Jordan for their unconditional love, support as well as positive mindset while I am solely living in Canada. I could not have pursued and accomplished my goals and attained my education without them.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AptaBiD	aptamer-facilitated biomarker discovery
ACE2	angiotensin-converting enzyme 2
BLI	bio-layer interferometry
BSA	bovine serum albumin
COVID-19	coronavirus disease 2019
CTD	C-terminal domain
dNTPs	deoxynucleotide triphosphates
dsDNA	double-stranded deoxyribonucleic acid
DPBS	Dulbecco's phosphate buffer saline
ELISA	enzyme-linked immunosorbent assay
ePCR	emulsion polymerase chain reaction
FW	forward primer
IgG	immunoglobulin G
IgM	immunoglobulin M
K_D	equilibrium dissociation constant
k_{on}	association rate constant
k_{off}	dissociation rate constant
LFA	lateral flow assay
LOD	limit of detection
LSPR	localized surface plasmon resonance
N	nucleocapsid protein

Ni-NTA	nickel-nitrilotriacetic acid
NTD	N-terminal domain
MERS-CoV	middle east respiratory syndrome coronavirus
PCR	polymerase chain reaction
RW	reverse primer
RBD	receptor-binding domain
RNA	ribonucleic acid
SA	streptavidin
SARS-CoV	severe acute respiratory syndrome coronavirus
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
S1	S1 subunit of spike protein
SELEX	systematic evolution of ligands by exponential enrichment
SPR	surface plasmon resonance
ssDNA	single-stranded deoxyribonucleic acid

Chapter 1

Introduction

1.1 COVID-19 Diagnostics

With the rapid spread of a novel coronavirus, identified as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) discovered in late 2019, the number of Coronavirus Disease 2019 (COVID-19) cases has dramatically increased and later become a global pandemic.[1] Thus far, approximately 539 million cases and more than 6.3 million deaths have been confirmed worldwide in 2021.[2] SARS-CoV-2 is an enveloped, single-stranded positive RNA genomic virus belonging to the Betacoronavirus genus in Coronaviridae family. The genomic RNA is surrounded by nucleocapsid-forming proteins and the lipid-like viral membrane.[3,4] The studies have reported that the source of SARS-CoV-2 was zoonotic and most likely from the host animal to human, and its transmission was initiated from the infected person excreting infected micro-aerosol from the respiratory tracts to the other person. The structure of this novel coronavirus is composed of structural proteins including nucleocapsid (N), spike glycoprotein (S), envelope (E) and membrane proteins encoded by the SARS-CoV-2 genome (Figure 1.1).[5,6] The S protein is the largest transmembrane glycoprotein among other structural proteins of SARS-CoV-2 which is protruding from the viral surface. The S protein comprises two functional protein subunits including the S1 and S2 subunits. The S1 subunit is composed of the receptor-binding domain (RBD) that has high affinity to the human angiotensin-converting enzyme 2 (ACE2) on the host surface receptor, and the S2 subunit mediates the cell fusion and the integration of the viral membrane (E and M proteins) into the host cell membrane rendering

the entry of the virus into the cells.[7] The S protein was firstly the main target of COVID-19 diagnostics and therapeutics including neutralizing antibodies and other viral inhibitors. In addition to the S protein, the N protein is the most conserved in the genus Betacoronavirus and it is highly abundant among other four structural proteins. The N protein is an RNA-binding protein that is crucial for packing viral genome into a long helical ribonucleocapsid (RNP) complex and playing an important role in assembling the virion via the interaction of the viral genome and the membrane protein. These predominant characteristics has recently inspired to target the N protein for the SARS-CoV-2 antigen detection. According to various manifestations of COVID-19, including asymptomatic infections which increase the effective transmission from the infected carriers to healthy people, it is necessary to develop robust, fast diagnostic tools that reflect the early stages of viral infection aiming to minimize the needs of hospitalization.[8]

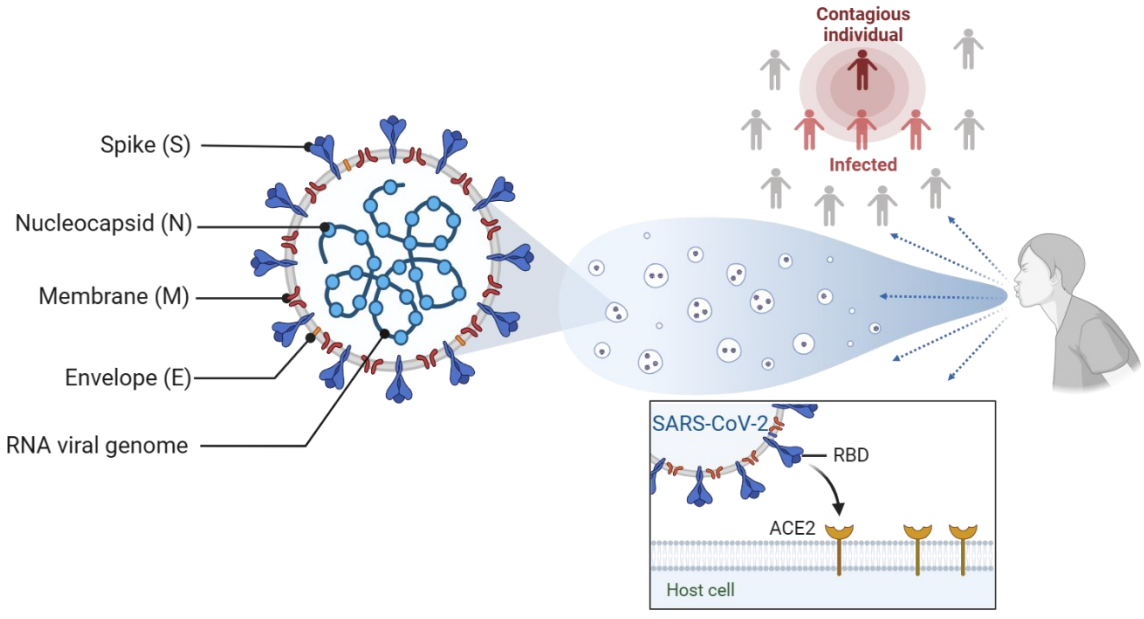


Figure 1.1 Transmission of SARS-CoV-2 from person to person. The scheme shows the spread of novel coronavirus causing COVID-19 through respiratory droplets from an infected person to others when the infected person sneezes, coughs and breathes. The viral infection can occur via the interaction of receptor-binding domain (RBD) of the S1 subunit of spike glycoprotein with the ACE2 cell receptor on host cells. The structure of the SARS-CoV-2 displays the components of the structural proteins of this novel coronavirus (created by BioRender).

In 2021, there were several commercial detections for SARS-CoV-2 granted for Emergency Use Authorization (EUA). Apart from X-ray, a conventional method for detecting lung infection, a nucleic acid amplification technique such as reverse-transcription polymerase chain reaction (RT-PCR) was primarily used for COVID-19 diagnosis,[9-11] and the genes encoding S, E, and N proteins have been targeted for the PCR test.[12-14] Additionally, antibody tests for COVID-19 diagnosis, also known as serological tests, were the alternatives to nucleic acid testing officially approved by the U.S. Food and Drug Administration (FDA) during the outbreak.[15,16] The serological test can rapidly test the antibodies produced by the immune system of infected people in response to the viral infection as well as the serological-based assays, e.g., enzyme-linked immunosorbent assay (ELISA) and lateral flow immunoassay (LFIA), essentially rely on the function of antibodies such as IgG/IgM implemented on the sensing materials for detecting viral antigens(Figure 1.2).[17-19]

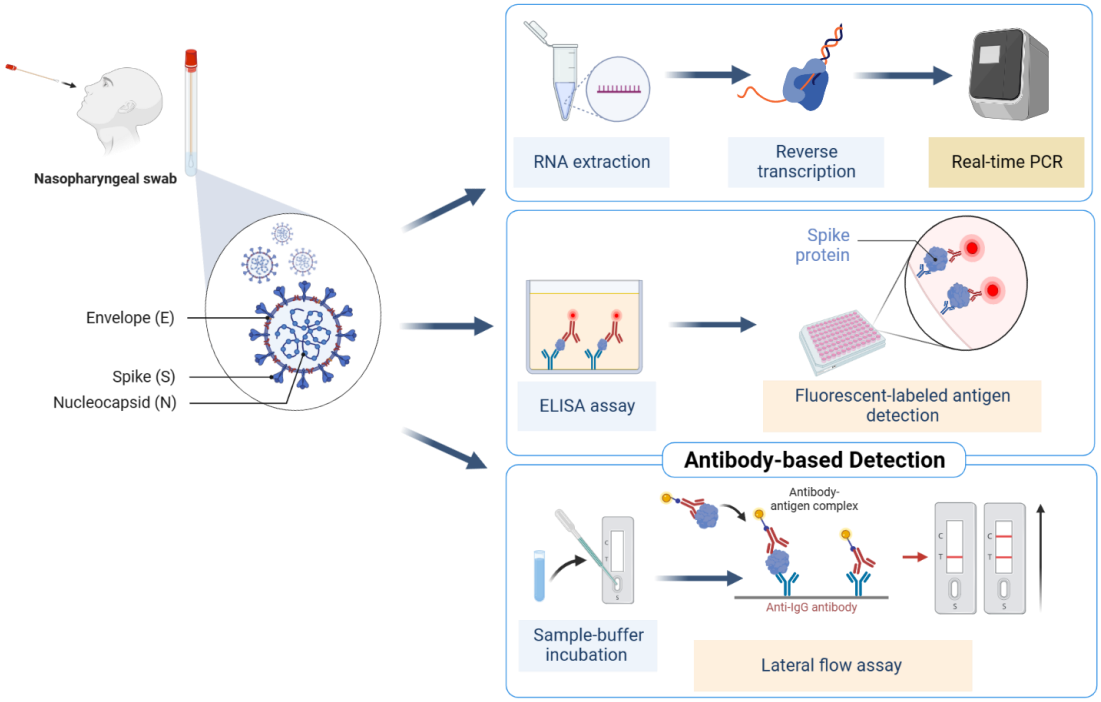


Figure 1.2 Platforms of current COVID-19 diagnosis. The diagnostic tools that have been widely used for detecting viral genes by RT-PCR, human antibody as well as viral antigen detections by ELISA and lateral flow assay (created by BioRender).

1.1.1 RT-PCR for viral gene detection

RT-PCR is a gold standard method used for detecting COVID-19. It stands for reverse transcription polymerase chain reaction. The process involves extracting RNA from a patient sample, such as a swab from the nose.[20] The extracted RNA is then converted into complementary DNA (cDNA) using reverse transcription. The DNA is then amplified by a process called polymerase chain reaction (PCR) using gene-specific primers and fluorescence-labeled probes.[21] The PCR creates millions of copies of the DNA throughout the repeated thermal cycles. The probes produce the fluorescent signals upon each

successful amplification of the target gene regions and the fluorescent signals of the DNA copies are then quantified and analyzed to determine if the gene of SARS-CoV-2 virus is present in the sample.[22] Although the RT-PCR is considered to be the most accurate method for COVID-19 testing, the number of tests per day is limited due to the time-consuming for sample pre-treatment and the complex process as well as the high cost of the test.[23] Also, the sensitivity of RT-PCR significantly relies on the amount of RNA in each sample meaning that an insufficient amount of the virus on the swaps leads to false-negative results, especially in the samples obtained from the very first days of the infection.[24,25] Hence, it is very crucial to develop viral detections with faster outcomes that do not depend on highly trained personnel and complicated reactions. Antibody-based tests targeting viral proteins such as nucleocapsid, spike or envelop protein could be useful approaches to accelerate the detection speed preventing the risk of person-to-person contact which can reduce the spread of the virus among the population.[26,27]

1.1.2 Antibody-based techniques for viral protein detection

The antibody-based detection is an FDA-approved tool in the field of medical research, especially viral antigen detection. Antibodies are proteins produced by the immune system in response to the responsive antigens, and antibodies are also called immunoglobulins (Ig) which can be classified into five classes: IgA, IgD, IgE, IgG and IgM distinguished by their C-terminal regions. During an infection, IgM is first produced while IgG is the most abundant in serum.[28-30] Thus, the presence and concentration of IgG and IgM can be measured in the plasma samples to determine if the antibodies are produced in the body to neutralize pathogens such as contagious viruses and inactivate the viral antigens.[31] To develop an antibody-based test, the recombinant antibody and viral antigen

are produced in the laboratory and used for the construction of lateral-flow assays (LFA) and enzyme-linked immunosorbent assays (ELISA).

In the antibody-based LFA test, the sample which contains viral antigen is deposited onto the sample pad and moved through a polymeric strip by capillary action to the conjugate pad where the antibody-labeled gold nanoparticles mix and form a complex with the viral antigen as a target of interest. Then, the complex migrates toward the first line that has either IgG or IgM antibodies immobilized on the nitrocellulose membrane. These antibodies capture the target complex forming the colour on a test line, and then the excessive antibody-labeled gold nanoparticles are moved further along the strip and captured at the control line as shown in Figure 1.2. As a test result, two lines should be displayed on the strip if the test is positive.[32,33] Even though the sample does not contain the target antigen, the antibody-labeled gold nanoparticles are captured at the control line appearing one line on the strip which proves the validity of the test. For the positive tests, there can be three lines if both IgG and IgM antibodies are immobilized on individual test lines.[34]

For the ELISA test, the recombinant viral antigen is coated onto the surface at the bottom of 96-well plates. Then, the patient sample is added to the well. The binding of the antigen-antibody occurs if the sample contains IgG or IgM antibodies. After washing out the excess sample, the horseradish peroxidase (HRP)-linked secondary anti-human antibody is added, and the binding of the secondary antibody to the primary (IgG/IgM) antibody can be detected by the colour change once the HRP enzyme interacts with its substrate.[35,36] The concentration of the antibody of interest can be quantified using a

spectrometer to record the fluorescence intensity. On the other hand, the colour change cannot be observed if the sample does not contain the antibody of interest after washing out the excessive HRP-linked secondary antibody from the wells. [37]

By implementing the specific antibodies onto LFA and ELISA, these techniques can only be indirect indicators of the viral infection which require authorized laboratories to perform moderate to high complexity tests.[38] Also, these immunoassays are struggling with the lack of specificity for emerging SARS-CoV-2 variants, and they rely on the function of antibodies which can provide a high risk of false-negative results caused by inadequate antibodies in clinical samples, inconsistent immobilization of antigen-coated supporting materials, and the varying range of viral load over the infection period can cause inaccuracy of antigen detection.[40]

Similar to antibodies, aptamers selectively bind and recognize target antigens or specific proteins. However, the production of antibodies requires the use of living animals which has more batch-to-batch variation and a higher chance of bacterial contamination than the production of aptamers.[41] Compared to antibodies, aptamers are smaller, chemically synthesized, and more stable with the ease of chemical modification which can be introduced by attaching spacers or linkers followed by the functional groups and fluorescence at the 5'- or 3'-end of the aptamers in order to expand the applications of the aptamers in diagnosis.[42]

1.2 Aptamers for COVID-19 Diagnosis

The term “aptamer” is derived from the Latin word *aptus* meaning “to fit” that was initiated by Ellington and Szostak in 1990 when they generated synthetic RNA molecules binding specifically to the dye molecules. Later, Tuerk and Gold reported a PCR-based method named Systematic Evolution of Ligands by Exponential Enrichment (SELEX) for aptamer discovery.[43-44] Aptamers are single-stranded DNAs and RNAs or peptides that bind to targets such as biological targets including proteins, small molecules, and whole cells like bacteria, viruses and human cells.[45] The aptamers can form three-dimensional (3D) structures rendering the ability of the aptamers to bind the target analytes with high affinity and specificity.[46] According to the characteristics of the aptamers, they are easy to be chemically synthesized *in vitro* and stably immobilized on the surface of biosensing materials which can be combined with other cutting-edge technologies such as CRISPR/Cas and nanomaterials to enhance the performance of diagnostic applications for infectious diseases.[47] Previously, several aptamers were developed to detect respiratory viral diseases including SARS-CoV in 2003 and MERS-CoV in 2012 which are relatively close to SARS-CoV-2.[48,49] Therefore, the use of aptamers in previous outbreaks has emphasized the potential of aptamer-based technologies in COVID-19 diagnosis.

1.2.1 Systematic Evolution of Ligand Exponential Enrichment (SELEX)

The systematic evolution of ligands by exponential enrichment (SELEX) is a method used to select specific nucleic acids known as aptamers from a highly diverse population called a library[50]. The conventional technique involves iterative rounds of selection, amplification, and isolation to enrich the desired ligands. During each round of

selection, the initial library of synthetic oligonucleotides is exposed to a target molecule, and those DNA oligonucleotides that bind to the target are isolated and amplified by polymerase chain reaction (PCR) for the next round.[51,52] For the RNA library, a reverse transcription step is required to synthesize the DNA for the amplification.[53] Over time, the population becomes enriched with high-affinity ligands that specifically bind to the target molecule, and then the enriched DNA pools are proceed further through next-generation sequencing (NGS) to retrieve the DNA sequences.[54] In addition to the conventional SELEX, the aptamer selection has been modified in order to generate optimal methods that effectively produce aptamers for some particular targets. For instance, capillary electrophoresis-based SELEX is used to select aptamers without immobilizing the target molecules on supporting materials such as magnetic beads, and Cell-SELEX is developed for isolating aptamers that specifically bind to cell surfaces, or exosomes secreted in the cell media.[55-57] Recently, there are several techniques (e.g., surface plasmon resonance (SPR), bio-layer interferometry (BLI), flow cytometry, etc.) used for validating the binding affinity of the aptamers in the term of apparent K_Ds in nanomolar to picomolar ranges to their respective targets or analytes.[58,59] Hence, the SELEX technique has been widely used for many applications in drug discovery, diagnostics, and basic research.[60]

1.3 Development of aptamer-based biosensors for SARS-CoV-2 protein detection

The implementation of aptamers in the optical biosensor is also widely used for COVID-19 diagnostic applications with the advantages of high sensitivity, less complexity of the pre-treatment process and stability[61]. The means of signal production is initiated from the biorecognition interaction of a receptor, or a ligand known as aptamers and analytes (such as viral antigens, cells and enzymes), and the chemical and/or physical

signals are converted into an optical signal on the optical field by signal transducers[62]. Optical aptasensors can be classified into two main categories based on detection modes: fluorescence-based and label-free aptasensors.

1.3.1 Fluorescence-based aptasensors

For the label-based detection, labelling reagents (e.g., fluorescence and radioactive labels) are required for the detection such as colorimetric aptasensor, chemiluminescent (CL) aptasensors, fluorescence aptasensors and surface-enhanced Raman scattering (SERS) aptasensors[63]. Li, J. et al. selected aptamers against the S protein of wild-type and B.1.1.7 SARS-CoV-2. The gamma-[32p]-labeled MSA1 and MSA5 aptamers were applied on dot-blot apparatus to assess the binding affinity to SARS-CoV-2 S1 protein in human saliva, and also colorimetric assay using ELISA was conducted to detect the spike protein presented on the surface of pseudotyped-lentivirus. Interestingly, the binding of MSA1 and MSA5 aptamers to the S protein in the saliva sample and to the pseaudovirus was still observed after the heat treatment. However, the sensitivity of the detection is still required to improve which fluorescent-labeled technique was proposed for further experiments[64].

Importantly, the NTD of the spike protein was reported to play in specific recognition of and immunity to the SARS-CoV-2 protein. Gupta et al. selected the aptamers binding to trimeric S protein and assessed the binding affinity to the patient-derived clinical samples using aptamer-linked immobilized sorbent assay (ALISA) with the LOD of 2 nM as well as ~ 91% sensitivity and ~ 98% specificity[65]. Compared to other two aptamers selected from this group, the G-quadruplex structure of the S14 aptamer has the potential to possess high selectivity and affinity to NTD of trimeric S protein. Notably, this S14 aptamer preferably binds to four amino acids (Asn23, Pro39, Asp53 and Thr376) of NTD

in the S1 region rather than binding to RBD according to the results obtained from the binding prediction in silico modelling. Along with the study of using ALISA, Kacherovsky et al. discovered two aptamers named SNAP1 and SNAP3 binding to NTD of SARS-CoV-2 S1 protein. According to the binding characterization using Cryo-EM, the aptamers bound to the overlapping epitope on NTD which SNAP1 showed higher binding affinity over SNAP3[66]. Thereby, FITC-labeled SNAP1 aptamer was further fabricated on LFA and ELISA to detect SARS-CoV-2 S protein with the LOD of 250 and 10 pM for LFA and ELISA, respectively. Similarly, Aithal et al. constructed ELISA for SARS-CoV-2 S protein detection. They proposed to improve the sensitivity of detection using gold nanoprobe functionalized with aptamers[67]. Once the SARS-CoV-2 virus was detected by the nanoprobe, the agglomeration of the aptamer-functionalized nanoprobe would cause the color change. This was practically assembled as a test kit with a total of 75-min testing time; nevertheless, the spectrophotometer was essentially required to indicate the test result. Owing to the simplicity of an ELISA apparatus, Svobodova et al. developed the ELISA technique for implementing an aptamer-based sandwich assay to detect the S protein of SARS-CoV-2. The pairing optimization of eight aptamers published in the previous study by Song et al. was performed to screen the most potential pair of aptamers that would improve the sensitivity and specificity of the sandwich assay. The pair of Apt1 and Apt5 aptamers showed as the best S1 protein detector with the highest sensitivity compared to other aptamer pairs. In the mode of sensitivity detection, the sandwich-based ELISA assay constructed with the thiol-labeled Apt1 and biotin-labeled Apt5 as a capture aptamer and a reporter aptamer, respectively, was able to detect the spike protein at the LOD of 270 pM in PBS buffer and 190 pM in the viral transportation medium (VTM).

Interestingly, once this aptamer-based sandwich assay was applied to the clinical samples obtained from nasopharyngeal (NP) swabs, there was about 70 – 72% accuracy of COVID-19 positive detection compared to the PCR test results[68].

Nonetheless, the ELISA-based technique that has been widely used for viral protein detection still provided low sensitivity and false-negative results. Deng J. et al. reported the one-step aptamer-based thermophoretic assay as an alternative for rapid and sensitive detection of SARS-CoV-2 spike protein without the requirement of the complex sample preparation. The binding recognition of Cy5-labeled aptamer to the spike on the virus mixed with PEG-thermophoretic enhancement results in the rapid accumulation of viral particles. Then, the PEG-enhanced thermophoretic accumulation of viral particles in PEG solution was processed through a localized infrared laser which then generated a temperature gradient causing the formation of PEG gradient. The intensity of fluorescence obtained from the accumulation of the aptamer and virus complex is linearly correlated to the concentration of viral particles. The LOD of this thermophoretic detection was determined at 26 fM. Besides, protein-induced fluorescence enhancement (PIFE) uses a fluorophore-labeled aptamer binding to a target protein in the sample solution without immobilizing and washing steps as ELISA as well as no need to use antibodies as traditional fluorescence-resonance energy transfer (FRET)[69]. Lee, J.M. et al. labeled an N-binding aptamer with Cy3 at 5'-end and the fluorescence signal was enhanced within 2 min once the N protein was present in the assay solution. It is fascinating that PIFE-based assay requires neither sophisticated sample preparation nor an advanced instrument. Instead, only two simple steps of mixing Cy3-labeled aptamer with a lysate sample in the solution and reading the fluorescence signal are processed. This approach was able to detect N protein

in SARS-CoV-2 lysate as low as 0.05 ng/μL compared to the LFA technique with the LOD of 5 ng/μL which is about a hundred-fold improvement[70].

1.3.2 Label-free aptasensor technique

In contrast to the label-based detection, one-step incubation of aptamers and analytes is potentially sufficient with the minimal requirement of functional groups to immobilize the aptamers on the sensing surface. The measurement of changes in the reflective index or the biolayer thickness on the optical surface is directly detected in real-time upon the analytes specifically interacting with the aptamer. This includes SPR and BLI aptasensors.[71] Stanborough et al. reported the use of SPR and BLI aptasensors to detect recombinant spike protein of SARS-CoV-2 at a physiological condition.[72] First, the two biotinylated aptamers (1C and 4C) specifically binding to the RBD of SARS-CoV-2 spike protein, selected by Song et al[73], were adsorbed on streptavidin fiber optic biosensors which function as a transducer of BLI. With the optimal conditions, the BLI signal response of the 1C aptamer interacting with spike protein was significantly higher than the 4C and scrambled aptamers. Besides, the detection limit of BLI was assessed as low as ~250 nM. To improve the LOD, Cennamo N., et al. used SPR by immobilizing the biotinylated 1C aptamer on a D-shaped plastic optical fiber (POF) coated gold sensor chip. The optical measurement was conducted in PBS buffer and diluted human serum and the limit of detection of spike protein with the varied concentrations was calculated as 36.7 nM.[74] Compared to traditional SPR and SPR-POF, Lewis T. et al. recently investigated the interaction of aptamers and SARS-CoV-2 S1 protein by using LSPR.[75] The gold nanoparticle on the sensor chip of LSPR is highly sensitive to the changes in refractive

index and the binding of biomolecules on the sensor surface.[76] As a result, the LOD of using the LSPR aptasensor to detect S1 protein was reported as 0.25 nM.

1.4 Challenges for developing aptamers-based COVID-19 Diagnosis

Remarkably, the applications of aptamers for viral diagnostics and therapeutics have been drastically increasing in the past decade since the emergence of the SARS coronavirus in 2002.[77,78] While most studies have been paying attention to the RBD and spike protein of SARS-CoV-2[73,79,80], SARS-CoV-2 nucleocapsid, an RNA-binding protein for viral genome encapsulation, is abundantly expressed in infected cells with very low genetic variation over time compared to the spike protein.[81,82] Thereby, the N protein could be used as a potential biomarker for early diagnosis.[83] In the early days of the global pandemic, Zhang and his coworkers[71] reported the DNA aptamers targeting SARS-CoV-2 N protein and implemented the selected aptamers with an anti-N antibody for the N protein detection. After that, several studies used the reported aptamer sequences to implement into various types of biosensors aiming to improve the sensitivity and accuracy in detecting SARS-CoV-2 nucleocapsid.[84-86] These studies are still in progress with disadvantages of requiring multiple steps of sample preparation, and a lack of understanding of molecular interaction between the N protein and aptamers. In addition to the aptamers targeting the N protein, having more varieties of aptamer sequences binding to the S1 protein in the database would be beneficial for the aptamer users who can further develop more advanced approaches for viral protein detections, and apply aptamers for not only COVID-19 diagnosis, but this will be also amendable for therapeutics.

1.5 Thesis aims

In this study, we selected novel ssDNA aptamers using asymmetric emulsion PCR-SELEX to target nucleocapsid and spike proteins of SARS-CoV-2 with different strategies and sequences from the previous studies. The aptamer sequence, possessing the highest affinity to the N and S1 proteins, was attached to the surface of BLI biosensors through the biotin-streptavidin interaction without additional labels. This label-free aptamer-based BLI was successfully able to detect the SARS-CoV-2 N and S1 proteins of wild type as well as Omicron and/or Delta variants. Furthermore, we studied the cross-reactivity of the selected aptamers to other SARS-CoV-2 and MERS-CoV proteins. In addition to the discovery of DNA aptamers targeting the N and S1 proteins, the technology of aptamer-facilitated biomarker discovery (AptaBiD)[87] was amendable to identify the specific binding epitope on the SARS-CoV-2 N and S1 proteins using our selected aptamers. The aptamer-binding digested peptides were processed through a proteomics approach, and the interaction of the binding epitope and the aptamer was additionally assessed by molecular dynamic simulation. These aptamers and corresponding BLI sensors may be used to distinguish between SARS-CoV-2 positive and negative clinical samples. The identified aptamer-protein binding motifs may also be applied to therapeutic purposes in the future. Hence, the use of this aptamer-based BLI has highlighted the potential of our discovered aptamer as a promising candidate for COVID-19 diagnostics.

1.6 Chapter overviews

The specific objectives which the studies aimed to address are described in Chapters 2 – 4 of the thesis. Beginning with Chapter 2, the selection of ssDNA aptamers targeting SARS-CoV-2 nucleocapsid (N) and S1 subunit of the spike protein using asymmetric-

emulsion PCR was explained as well as the aptamer sequences were identified using a software tool named FASTAptamer. There were sixteen and fourteen DNA sequences passed through the criteria used for choosing the sequences targeting the N and S1 proteins, respectively. After that, Chapter 3 describes the evaluation of the binding affinity of the selected DNA aptamers to the target proteins by using BLI, and only the best binding candidates showing the lowest K_Ds which include the N- and S1-binding aptamers named tNSP3 and S1-tSP10 were further used for the development of aptamer-based BLI as a label-free technique for SARS-CoV-2 protein detection in human saliva. Besides, Chapter 4 shows the identification of the binding epitopes on the N and S1 proteins using the aptamer-based pull-down assay, and the digested peptides were identified by using nLC-MS/MS. Lastly, Chapter 5 presents the overall conclusion and future directions of the research are proposed.

Chapter 2

Selection of ssDNA Aptamers via Asymmetric Emulsion PCR-SELEX

A partial version of this chapter has been previously published in

Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics (Mol. Ther. Nucleic. Acids, Vol. 31, P731-743, March14, 2023)

Suttinee Poolsup, Emil Zaripov, Nico Hüttmann, Zoran Minic, Polina V. Artyushenko, Irina A. Shchugoreva, Felix N. Tomilin, Anna S. Kichkailo, and Maxim V. Berezovski

Author contributions for Chapter 2:

Suttinee Poolsup designed and performed the experiments and wrote materials and methods, and also results and discussion.

Suttinee Poolsup contributed Tables 2.1, 2.2, 2.3, 2.4 and Figures 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6

2.1 Introduction

In SELEX method, amplification is one of the vital steps highly affecting the success in aptamer selection as described in Chapter 1. The most general procedure for producing the DNA library is to amplify target-bound ssDNA through conventional PCR, and then convert dsDNA as a PCR product to the ssDNA pool which is used in the next round of selection.[88] Various techniques have been introduced to isolate the ssDNA; for example, the enzyme-based technique uses lambda exonuclease enzyme to eliminate the undesirable complementary strand at the 5'-phosphorylated strand of dsDNA. This technique requires optimization of digestion time and enzyme concentration to produce a high yield of ssDNA.[89] The other technique is the biotin-streptavidin-based technique in which the biotin-labeled reverse primer is used in the PCR and the unwanted strand is removed by streptavidin magnetic beads followed by alkaline denaturation.[90] These two techniques require an additional purification step through DNA precipitation which could lower the yield of ssDNA. To shorten the time consumption of aptamer selection via SELEX, asymmetric PCR is a potential alternative to conventional PCR.[91] Theoretically, the principle of asymmetric PCR is the addition of unequal concentrations of forward and reverse primers in the PCR reaction so that the amplification can continue with the excessive primer to produce the desired ssDNA strands.[92] Still, it is necessary to optimize some parameters including primer ratios, and the number of PCR cycles in order to obtain a sufficient yield of ssDNA which is challenging when it comes to developing the procedure of aptamer selection. The inadequate amount of DNA templates obtained from

the beginning rounds of selection contributes to non-specific amplification and determining the unsuccessful *in vitro* selection process.[93]

To alleviate the over-amplification or primer-dimer hybridization which can generate undesired PCR by-products in the conventional asymmetric PCR (Figure 2.1), emulsion PCR is an alternative method used to solve these problems. The emulsion PCR was initially proposed to overcome the difficulty in amplification of genomic DNA, cDNA library and other complex mixtures of genes which are composed of long DNA fragments and could generate the recombination between homologous regions resulting in the formation of undesired chimeric DNAs.[94] According to the basic principle of emulsion PCR, the DNA template is compartmentalized into numerous droplets of the water-in-oil emulsion. Ideally, each droplet functions as a PCR reactor containing a single molecule of the DNA template, a set of FW/RW primers and other PCR reagents.[95] The formation of the emulsion PCR is based on the mixture components of mineral oil supplemented with non-ionic surfactants including Triton-X100 and Tween-80 and/or other emulsifiers such as Span 80 and ABIL WE 09. Also, BSA is essentially required in the preparation of the PCR mixture to saturate the interface of organic and aqueous phases which can protect the RNA polymerase from being inactivated at the interface.[96] To recover the amplified DNA products from the emulsion PCR, liquid-liquid extraction is performed by adding the volatile organic solvents such as diethyl ether to remove the remaining oil from the emulsion along with vortexing and centrifugation which facilitate to break the emulsion and separate the organic (solvent) phase (upper layer) from the aqueous phase (bottom layer).

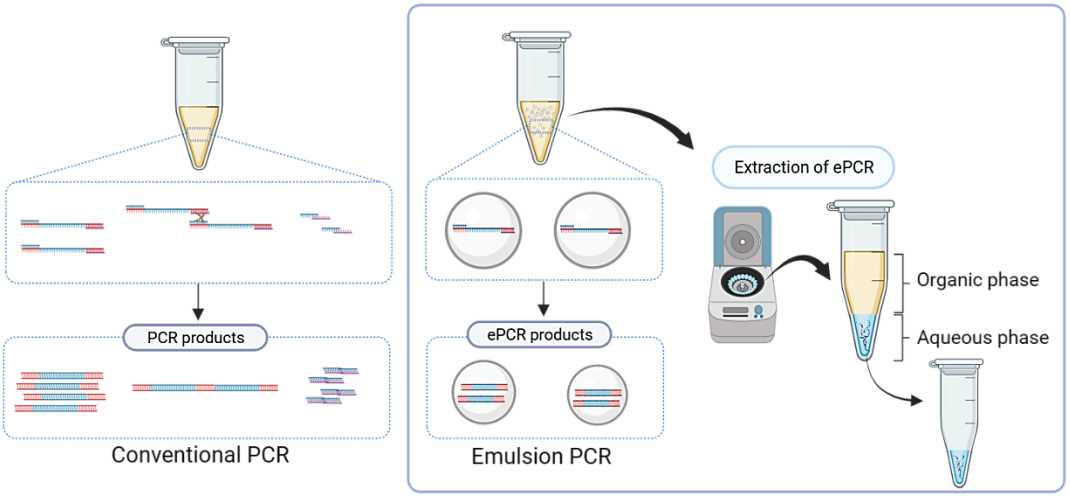


Figure 2.1 Amplification of DNA library by conventional PCR and emulsion PCR. DNA fragments from the previous round of selection are pooled together for amplification in a conventional PCR (left) or emulsion PCR which each droplet contains a single template DNA molecule (right).[94] Then, the amplified DNA from the emulsion PCR was extracted using liquid-liquid extraction by adding organic solvents such as diethyl ether or 2-butanol.

Due to the advantages of emulsion PCR, there are several studies implementing emulsion PCR in the aptamer selection. Yufa et al. (2015) used the emulsion PCR to improve the selection of aptamers through nonequilibrium capillary electrophoresis of equilibrium mixture (NECEEM)-based SELEX. This could fasten the process of aptamer selection targeting the AlkB homologous 2 (ABH2) enzyme by completing within four rounds with the K_D range of 38 nM to 1 μM whereas the binding complex of the aptamer and ABH2 was not detectable when the conventional PCR was used in the NECEEM-based SELEX.[97] Moreover, Liu et al. (2020) applied the emulsion PCR combined with magnetic beads for selecting DNA aptamers targeting a rice viral capsid protein. The optimization

of emulsion PCR in SELEX along with closely monitoring the formation of PCR droplets could successfully improve the aptamer selection for detection of viral protein in plants via aptamer-based dot-blot assay and aptamer-based immunofluorescence.[98]

Apart from the applications of emulsion PCR in the aptamer selection, emulsion PCR has been widely used for solid-phase PCR using microbeads and high-throughput next-generation sequencing including whole genomic sequencing and RNAseq. Despite the advantages of using emulsion PCR in the means by minimizing the chances of self-hybridization of DNA templates as well as primer-dimer formation, the optimization of the emulsion condition (e.g., types of oils and surfactants, concentrations of BSA, RNA polymerase, DNA templates and number of droplets per PCR reaction, etc.) is still crucial in order to increase the PCR efficiency and reduce the PCR artifacts.[99]

In this chapter, asymmetric emulsion PCR was implemented in the SELEX process to select ssDNA aptamers targeting SARS-CoV-2 nucleocapsid and S1 subunit of the spike protein.

2.2 Materials and Methods

The DNA (N40) Library**,** 5'-CTCCTCTGACTGTAACCACG-(N40)-GCATAGGTAGTCCAGAAGCC -3', forward primer (5'-CTCCTCTGACTGTAACCACG-3') and reverse primer (5'-GGCTTCTGGACTACCTATGC-3') were purchased from Integrated DNA Technologies (Newark, NJ, USA). Phosphate buffer saline (PBS) without Ca^{2+}-Mg^{2+} (Cat. No. 10010031) and DPBS with Ca^{2+}-Mg^{2+} (Cat. No. 14040141) were purchased from Gibco. tRNA (Cat. No. 15401-011) was purchased from Thermo Fisher Scientific. BSA (Cat. No. A9418), Tween 20 (Cat. No. 9005-64-5), Span80 (Cat. No. 85548), Tween80 (Cat. No.

P4780), Triton-X100 (Cat. No. T9284) and mineral oil (Cat. No. M8410) were purchased from Sigma-Aldrich. Microcon-10kDA Centrifugal filter (Cat. No. MRCPRT010) was purchased from Millipore. Ni-NTA HisSorb Strips (Cat. No. 1002478) were purchased from Qiagen. All buffers should be filtered through a 0.22 μm filter before use.

2.2.1 Asymmetric-emulsion PCR-SELEX of ssDNA aptamers

2.2.1.1 Immobilization of his-tagged proteins on Ni-NTA strips

The protocol was slightly modified from the manufacturer's protocol. Briefly, His-tagged proteins including nucleocapsid and S1 subunit were separately dissolved in PBS with Ca^{2+}-Mg^{2+} buffer at the concentration of 20 μg/mL, and the protein mixture was added to each well of a Ni-NTA HisSorb strip and incubated at 4°C overnight. After coating the protein on the strip, the wells were rinsed with Ca^{2+}-Mg^{2+} added PBS buffer.

2.2.1.2 Preparation of ssDNA library

2 nmoles of ssDNA (N40) library were mixed in 300 μL of DNA-heating buffer (DPBS with Ca^{2+}-Mg^{2+}) and the DNA library mixture was heated at 95°C for 5 min followed by cooling on ice for 10 minutes. Then, the heat-folded DNA library was prepared in the selection buffer (PBS with Ca^{2+}-Mg^{2+}, 2 μg/μL BSA, 0.2 μg/μL tRNA)

2.2.1.3 Selection of aptamer partitioning target proteins

For the first round of selection, the heat-folded library initially went under the negative selection by adding the library from section 2.2 into each well of the 8-well Ni-NTA strip and incubating at room temperature on an orbital shaker for 30 min. During the total of 6 rounds of aptamer selection, the negative selections were repeated twice including in the first (as described above), and the fourth round of selection. Prior to starting the

fourth round, the negative selection was performed by incubating the enriched DNA pool obtained from the third round with the non-targeted protein-coated 8-well Ni-NTA strips on the orbital shaker at room temperature for 30 minutes. Then, the non-target unbound DNA pool was transferred to the strips which were coated with the target protein and incubated on the orbital shaker at room temperature for 1 hour. After DNA-protein incubation, 200 µL of the washing buffer (DPBS with Ca^{2+}-Mg^{2+} added 0.05% Tween20) was added to each well to wash out the unbound DNAs on the protein-coated strips and the supernatant was discarded. The washing step was repeated twice. The stringency of consecutive rounds could be increased by an additional number of washings as well as the decrease of DNA-protein incubation time as shown in Table S1. Next, 200 µL of the DPBS with Ca^{2+}-Mg^{2+} was added to each well of the strip in order to avoid carrying over Tween20 contained in the washing buffer to the PCR reaction. After that, 200 µL of nuclease-free water was added to each well of the strip and incubated at 95°C for 10 min to elute the targeted protein-bound DNA from the strip followed by transferring the eluted DNA pool to a 1.5 mL centrifuge tube and cooling down at room temperature. To concentrate the DNA eluent, the DNA pool was transferred to a 10 kDa centrifugal filter and centrifuged at 14,000 rpm for 20 min until the final volume of the concentrated DNA remained at 100 µL on the filter.

2.2.1.4 Asymmetric Emulsion PCR (ePCR) amplification

The concentrated DNA pool obtained from the selection steps was subjected to asymmetric emulsion PCR. The PCR master mix was prepared according to the sequences of each reagent (Table 1) in the final volume of 200 µL asymmetric PCR mixture. At the same time, the emulsion oil mixture was prepared by homogenously mixing 4.5% span80,

0.4% Tween80 and 0.05% triton-X100 in mineral oil. Then, the 200 µL of asymmetric PCR mixture was slowly added dropwise to 400 µL of emulsion oil while stirring. After 5-min stirring, 50 µL of emulsion PCR mixture was aliquoted into each PCR tube. After optimization of number of PCR cycles (Figure S1), the PCR reaction was amplified for 25 cycles by using the following PCR program: 98°C for 30 sec; 25 cycles at 98°C for 30 s, 56°C for 15 s, 72°C for 15 s and hold at 4°C.

Table 2.1 Asymmetric emulsion PCR mixture. Amount and concentrations of reagents used to prepare PCR reactions with a total volume of 200 µL.

Reagents	Reaction mixture	Final concentration
5X Phire reaction buffer	40 µL	1X
10 mM dNTPs	4 µL	0.2 mM
10 µM Cy5-forward primer	20 µL	1 µM
10 µM Reverse primer	1 µL	0.05 µM
100 g/L BSA	20 µL	10 g/L
Nuclease-free water	11 µL	-
DNA template	100 µL	~100-150 ng
50X Phire Hot Start II DNA polymerase	4 µL	1X

2.2.1.5 DNA Extraction of Asymmetric Emulsion PCR

The asymmetric ePCR product obtained from the previous steps was combined into a new 1.5 mL microcentrifuge tube. 200 µL of nuclease-free water was added and vortexed to mix followed by adding 500 µL of 2-butanol and vortexed for 30 s. The mixture was centrifuged at 15,000 g for 5 min, and then the top layer of the oil phase suspended in organic solvent was discarded. Then, the 200 µL of nuclease-free water and 500 µL of 2-butanol were repeatedly added and vortexed before centrifuging the mixture at 5,000 g for 5 minutes, and the top layer of the organic solvent was discarded. Again, 500 µL of 2-butanol was added into the mixture, vortexed and shaken horizontally on the orbital shaker

at 750 rpm for 10 min. The mixture was centrifuged at 3,000 g for 2 minutes and the top layer was discarded. At this step, the DNA pellet would be formed and dried on a speed-vac concentrator for 20 minutes at room temperature. After drying the DNA pellet, 100 µL of nuclease-free water was added to thoroughly dissolve the DNA pellet and its concentration was measured by using NanoDrop. During the extraction of DNA from the emulsion PCR mixture, the pellet of the hydrophobic phase, known as the phase of organic solvent, might be formed at the bottom of the tube. If the pellet could not be completely dissolved by pipetting up and down within 15 min at room temperature, the tube of DNA solution could be placed on an incubator shaker at 4°C overnight and the jelly-like pellet could be manually removed from the DNA suspended in water on the following day.

2.2.1.6 Emulsion PCR Clean-up (modified from the manufacturer's protocol)

To purify the amplified DNA obtained from the previous steps, 100 µL of DNA solution was inversely mixed with 200 µL of DNA cleanup binding buffer. Then, 300 µL of absolute ethanol (≥ 95%) was added and mixed well by pipetting up and down or flicking the tube followed by loading the DNA-solvent mixture onto the PCR cleanup column. The DNA-containing column was centrifuged at 16,000 rpm for 1 minute, and the flow-through was discarded. Then, 500 µL of DNA wash buffer was added to the column and centrifuged at 16,000 rpm for 1 minute, and the flow-through was discarded before repeating the washing step twice. Lastly, the purified DNA was eluted from the PCR cleanup column into a new 1.5 mL collecting tube by adding 50 µL of nuclease-free water to the center of the filter matrix, incubating at 50°C for 1 minute and centrifuging at 16,000 rpm for 1 minute to collect the flow-through from the column. The purified ssDNA was subjected to

the next round of selection and these aforementioned steps were repeated until the 6th round as the schematic summary of SELEX process shown in Figure 2.2.

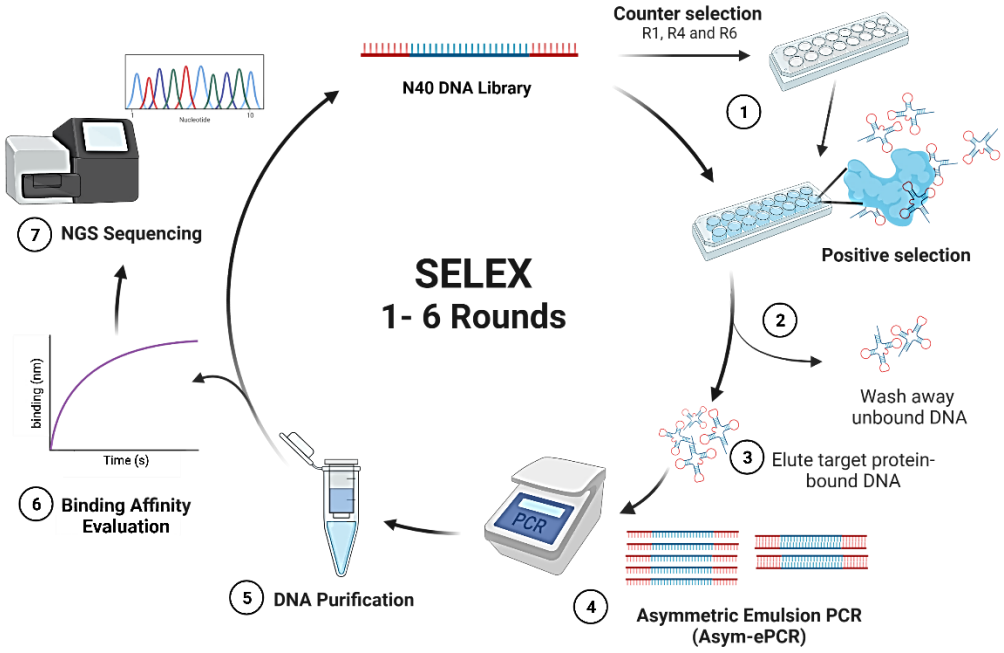


Figure 2.2 Asymmetric-emulsion PCR-SELEX. Schematic representation of asymmetric emulsion PCR (ePCR) SELEX.

2.2.2 Next-Generation Sequencing (NGS)

After six rounds of selection, each ssDNA pool was reamplified by symmetric PCR and purified prior to the next-generation sequencing. The FASTQ raw data obtained from NGS were analyzed by following the guideline from FASTAptamer[100] as shown in Figure 2.3. The data representing each aptamer pool were categorized into its cluster, followed by using the FASTAptamer-enrich tool to retrieve the most enriched sequences. After the aptamer sequences were identified, the multiple sequence alignment tool on Clustal Omega[101] was used to evaluate the similarity and differentiation of individual aptamer sequences which were classified and presented as phylogenetic trees. The secondary

structures of the selected aptamers were then predicted using the RNAstructure web server.[102] The details of retrieving aptamer sequences were clearly described in Appendix.

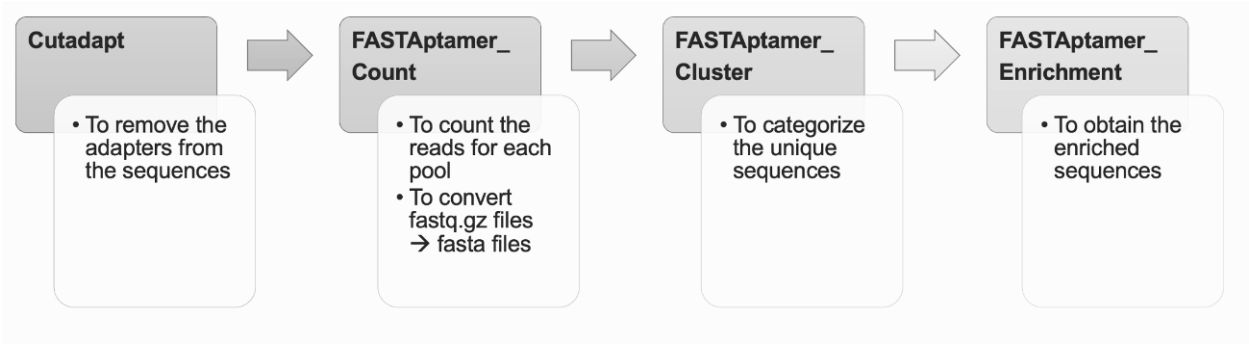


Figure 2.3 Schematic NGS data analysis. The data processing workflow to retrieve the aptamer sequences from NGS raw data.[100]

2.3 Results

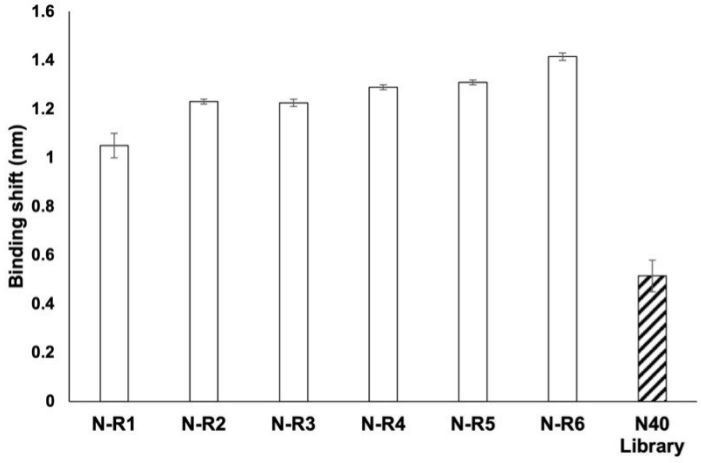
2.3.1 Selection of DNA aptamers to SARS-CoV-2 proteins

The recombinant nucleocapsid protein and S1 subunit of spike protein were separately used as targets for the positive selections of ssDNA aptamers using the asymmetric-emulsion PCR SELEX procedure shown in Figure 2.1. The DNA library, with a 40-nt random region flanked with 20-nt primer binding sequences at 5' and 3' ends, was incubated with the target protein immobilized on Ni-NTA HisSorb strips. As thoroughly described in the Materials and Methods, the unbound DNA sequences were washed out while the target protein-binding DNA pool was further amplified by asymmetric emulsion PCR and purified with a PCR cleanup kit to generate an ssDNA pool subsequently subjected to the next rounds of selection. To eliminate the non-specific sequences, the counter-selection took place in the first, fourth, and sixth rounds of selection by incubating the enriched DNA pool with a non-target protein before starting the positive selection. After each round of selection, one additional wash was added to remove the weak-binding

ssDNA sequences and increase the stringency of the selection as well as the number of PCR cycles was optimized to achieve maximum amplification efficiency.

To assess the presence of the desired ssDNA sequences with 80-nt length in each enriched pool from every round of selection, gel electrophoresis was carried out, and the results showed two bands of ssDNA as a major product and dsDNA as a minor product located at 40 bp and 80 bp, respectively, as shown in Figure S1. After six selection rounds, the enriched pools were subjected to preliminary binding affinity tests using BLI, whose principle was clearly described in Chapter 3. All six enriched pools showed relatively higher binding shifts (Figure 2.4) with a 2-fold increase of the association curves compared to the initial N40 DNA library after 300-s partitioning with N protein on the Ni-NTA biosensor tips (Figure S2A). Similarly, the association curves of the enriched DNA pools also showed higher binding onto the S1 protein compared to the initial DNA library (Figure S2B). Therefore. all enriched pools were sequenced and analyzed using next-generation sequencing (NGS).

A



B

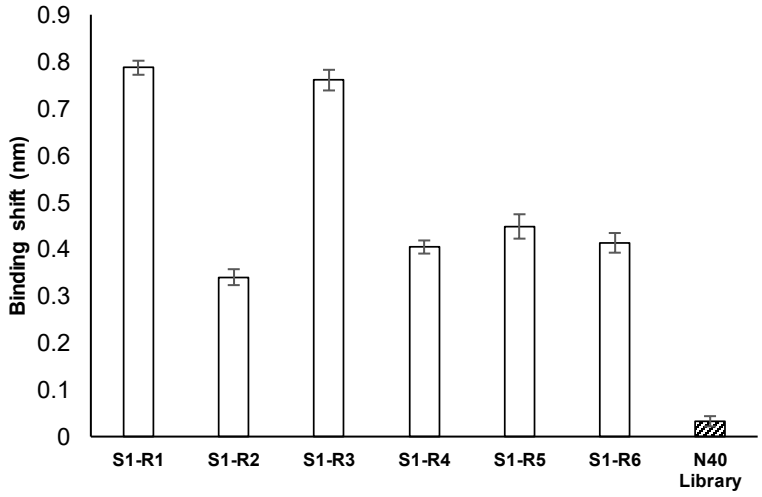


Figure 2.4 Affinity screening test of enriched DNA pools after SELEX. Binding evaluation of six enriched DNA pools obtained from asymmetric-ePCR SELEX targeting (A) SARS-CoV-2 nucleocapsid and (B) SARS-CoV-2 S1 subunit proteins compared to an initial DNA N40 library using BLI. Error bars represent means ± standard deviations (SDs). Also, see Figure S2A and S2B.

2.3.2 NGS data analysis of DNA aptamer sequences targeting SARS-CoV-2 nucleocapsid

The NGS results for all six pools were processed using the FASTAptamer toolkit (Figure 2.3).[100] Sequences were counted and clustered into families based on the neighboring distance and were further subjected to the enrichment tool to calculate the fold-enrichment of individual sequences across six pools. The retrieved sequences were chosen based on three criteria, including (i) the total length of 80-82 nucleotides containing 40-42 nt random region flanked with forward and reverse primers, (ii) the identical

sequences enriched in more than one pool, and (iii) the sequence composition without primer-dimer formation in the random region. While three enriched sequences named NSP1, NSP2, and NSP3 were obtained from the cluster-enrichment tool, the count-enrichment tool resulted in sixteen unique sequences (Table 2.2) based on the criteria above, including the NSP1, NSP2, and NSP3 aptamer sequences. These sixteen unique sequences were then aligned to analyze the differentiation and similarity using Clustal Omega.[101] Relative to the construction of the N40 DNA library with a 40-nt randomized region, the insertion of two nucleotides was observed in two sequences named NSP6 and NSP10. Phylogenetic tree analysis based on primary sequence homology categorized the sixteen aptamer sequences into three groups (Figure 2.5).

Table 2.2 Full-length sequences (80 nt) of the N-binding selected aptamer. The enriched sequences were obtained from NGS data enrichment. The primer-binding regions are written in bold.

Aptamer	Aptamer sequence
NSP1	**CTCCTCTGACTGTAACCACG**GCGCAAGCCGGGGTGTACGTGTTA TACGTGCGTGTATCGA**GCATAGGTAGTCCAGAAGCC**
NSP2	**CTCCTCTGACTGTAACCACG**TATTGCGTTCCAGTCCCTATGACC AACGTCACAATAAGTC**GCATAGGTAGTCCAGAAGCC**
NSP3	**CTCCTCTGACTGTAACCACG**CAGCGTCACGTGTTGTTCCCCATT GTACTGATTCGTCGTG**GCATAGGTAGTCCAGAAGCC**
NSP4	**CTCCTCTGACTGTAACCACG**CGTTGAGCGTGTTGTCCCTACATG CAATAGACCCTCCTTC**GCATAGGTAGTCCAGAAGCC**
NSP5	**CTCCTCTGACTGTAACCACG**ATTCATGCGCCAATAGTGGTTTGG AAATGTCTCCCCATAC**GCATAGGTAGTCCAGAAGCC**

NSP6	**CTCCTCTGACTGTAACCACG**GCATAGGTAGTCAAGAAGCCATCT CCTATGACTGTAACCACG**GCATAGGTAGTCCAGAAGCC**
NSP7	**CTCCTCTGACTGTAACCACG**TGCTAGATTACGTAACTGGTTGCT ACGGGTTTCATCCTCG**GCATAGGTAGTCCAGAAGCC**
NSP8	**CTCCTCTGACTGTAACCACG**AGGATTGCGCACATACGGTTGCTC ACCATTCTACTGTCCT**GCATAGGTAGTCCAGAAGCC**
NSP9	**CTCCTCTGACTGTAACCACG**CTTATGCACCAGGGGCCCGGTTCG GGCAGGCGCAAACACG**GCATAGGTAGTCCAGAAGCC**
NSP10	**CTCCTCTGACTGTAACCACG**GCATAGGTAGTCAAGAAGCCATCT CATCTGACTGTAACCACG**GCATAGGTAGTCCAGAAGCC**
NSP11	**CTCCTCTGACTGTAACCACG**GCGCACCCTGCGACGTCCTCATGC CAGGCTTCAGCCTCAA**GCATAGGTAGTCCAGAAGCC**
NSP12	**CTCCTCTGACTGTAACCACG**TGGTTCGCCCAGTTCCGTTCAATA CACTCAACCCTGCACG**GCATAGGTAGTCCAGAAGCC**
NSP13	**CTCCTCTGACTGTAACCACG**TATTGCGTTCCATTCCCTATGACCA ACGTCACAATAAGTC**GCATAGGTAGTCCAGAAGCC**
NSP14	**CTCCTCTGACTGTAACCACG**GGCTTGGTTTCGGGATCTGCACCC CTGAAACATTTCTCAG**GCATAGGTAGTCCAGAAGCC**
NSP15	**CTCCTCTGACTGTAACCACG**CTCGTCCAGCGGCTGCGTGTTTCA CCGTGCATTGTTGATA**GCATAGGTAGTCCAGAAGCC**
NSP16	**CTCCTCTGACTGTAACCACG**TTCAAGAGCGGTTCGTAATGTTAA TTGACTGATCCCTACC**GCATAGGTAGTCCAGAAGCC**

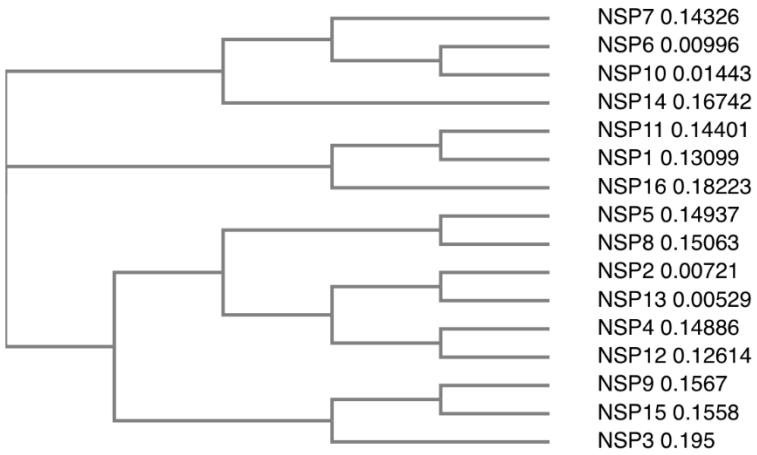


Figure 2.5 Phylogenetic tree of N-binding DNA aptamers. The diagram obtained from Clustal Omega represents the three main groups of sixteen enriched aptamer sequences targeting SARS-CoV-2 nucleocapsid and phylogenetic distances are displayed to represent the number of pairwise differences in nucleotides between the two close sequences. The sequences were retrieved from count-enrich FASTAptamer.

Among those sixteen unique sequences, the seven aptamers representing each subgroup were chosen for further validation of their affinity binding to the N protein, and their structures were truncated and named tNSP1, tNSP2, tNSP3, tNSP5, tNSP9, tNSP10, and tNSP12 (Table 2.3). Based on the predicted secondary structures by RNAstructure, this results in different nucleotide lengths of the selected aptamers in the range of 47 – 60 nt.

Table 2.3 Truncated sequences of the N-binding aptamers. The sequences of selected aptamers binding to the N protein were shortened based on the prediction of their secondary structures obtained from RNAstructure web server[102].

Aptamer	Aptamer sequence
tNSP1 (49nt)	TAACCACGGCGCAAGCCGGGGTGTACGTGTTATACGTGCGTGTACGAG
tNSP2 (64nt)	CTGACTGTAACCACGTATTGCGTTCCAGTCCCTATGACCAACGTCCAATAAGTCGCATAGGTA
tNSP3 (44nt)	CAGCGTCACGTGTTGTTCCCCATTGTACTGATTCGTCGTGGCAT
tNSP5 (46nt)	ATTCATGCGCCAATAGTGGTTTGGAAATGTCTCCCCATACGCATAG
tNSP9 (60nt)	TGTAACCACGTGCTAGATTACGTAACTGGTTGCTACGGGTTTCATCCTCGGCATAGGTA
tNSP10 (47nt)	CTTATGCACCAGGGGCCCGGTTCGGGCAGGCGCAAACACGGCATAGG
tNSP12 (60nt)	TGTAACCACGTGGTTCGCCCAGTTCCGTTCAATACACTCAACCCTGCACGGCATAGGTA

2.3.3 NGS data analysis of DNA aptamer sequences targeting SARS-CoV-2 S1 subunit

To retrieve the DNA sequences binding to the S1 subunit of the spike protein, the count-enrichment was used, and fourteen sequences passed through the criteria as mentioned earlier (Table 2.4). Similar to the N-binding sequences, the fourteen unique sequences obtained from the FASTAptamer enrichment tool were then aligned to analyze the differentiation and similarity using Clustal Omega. Phylogenetic tree analysis based on primary sequence homology categorized the fourteen aptamer sequences into three groups (Figure 2.6).

Table 2.4 Full-length sequences (80 nt) of the S1-binding selected aptamer. The enriched sequences were obtained from NGS data enrichment. The primer-binding regions are written in bold.

Aptamer	Aptamer sequence
S1-SP1	**CTCCTCTGACTGTAACCACG**AACCACGTATTGTGCACTGTCTCAT TAGTAATCGCTCCCA**GCATAGGTAGTCCAGAAGCC**
S1-SP2	**CTCCTCTGACTGTAACCACG**GCGCAAGCCGGGGTGTATGTGTTAT ACGTGCGTGTATCGA**GCATAGGTAGTCCAGAAGCC**
S1-SP3	**CTCCTCTGACTGTAACCACG**GCGCAAGCCGGGGTGTACGCGTTA TACGTGCGTGTATCGA**GCATAGGTAGTCCAGAAGCC**
S1-SP4	**CTCCTCTGACTGTAACCACG**TAATGCCTAACTCTTTTTGTGTTTGC GATCTTTGCACATA**GCATAGGTAGTCCAGAAGCC**
S1-SP5	**CTCCTCTGACTGTAACCACG**TATCTCCGCACCGGCATGGAACGA CAGATTCCCAACTTCC**GCATAGGTAGTCCAGAAGCC**
S1-SP6	**CTCCTCTGACTGTAACCACG**CGGTTACGATCGGGTTACAAACAC GGCTGACACGAATTCA**GCATAGGTAGTCCAGAAGCC**
S1-SP7	**CTCCTCTGACTGTAACCACG**TCCTCATCCCGCAAACCGCCTTGTC AGAAGAACTTTGCAT**GCATAGGTAGTCCAGAAGCC**
S1-SP8	**CTCCTCTGACTGTAACCACG**TTCCATTTTACGGATCCCCTGGTGT TATAGAATGGTGTAT**GCATAGGTAGTCCAGAAGCC**
S1-SP9	**CTCCTCTGACTGTAACCACG**TTTCTTTCGCGTGTGGTCTCAAGAA GAGTACCGTGACTGC**GCATAGGTAGTCCAGAAGCC**
S1-SP10	**CTCCTCTGACTGTAACCACG**CACACTTTCTGCCCGCCTTCTCCCT CCGTTCCCCTCCCCG**GCATAGGTAGTCCAGAAGCC**
S1-SP11	**CTCCTCTGACTGTAACCACG**ATGTCCTCGCACACCCAAACGCACT CATCTCCCCACCCAT**GCATAGGTAGTCCAGAAGCC**
S1-SP12	**CTCCTCTGACTGTAACCACG**ATTCATGCGCCAATAGTGGTTTGGA AATGTCTCCCCATAC**GCATAGGTAGTCCAGAAGCC**
S1-SP13	**CTCCTCTGACTGTAACCACG**TATCGCGTCATTCGATCCATTTGTA CATCATTGTGCATAG**GCATAGGTAGTCCAGAAGCC**
S1-SP14	**CTCCTCTGACTGTAACCACG**TCTTTACAAGTTCACACCCTTGGTA CATGACTACATTCAC**GCATAGGTAGTCCAGAAGCC**

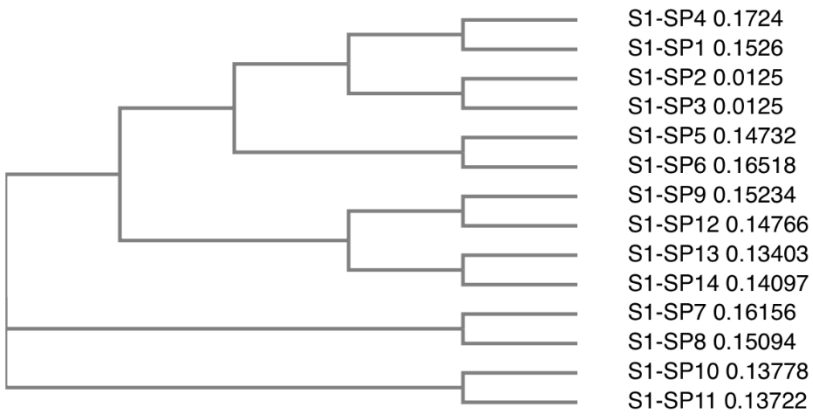


Figure 2.6 Phylogenetic tree of S1-binding DNA aptamers. The diagram obtained from Clustal Omega represents the three main groups of fourteen enriched aptamer sequences targeting the SARS-CoV-2 S1 subunit of the spike protein and phylogenetic distances are displayed to represent the number of pairwise differences in nucleotides between the two close sequences. The sequences were retrieved from count-enrich FASTAptamer.

Moreover, the secondary structures of these aptamer sequences were predicted using an RNAstructure and MFold web servers for designing the truncated motifs of aptamers.[102,133] Based on the observation of the predicted structures, some additional nucleotides at the 5'- and 3'- primer binding regions remained to conserve the stem-loop structures. The secondary and tertiary structures of the three most enriched sequences with truncated structures were predicted using MFold and MD simulations, respectively (Figure S5 and S6).

2.4 Discussion

The novel sequences of DNA aptamers were discovered from a synthetic DNA library which the two selections were separately targeted the nucleocapsid and S1 subunit of spike protein of wild-type SARS-CoV-2. The methods of selection were designed and

developed with the main idea of reducing the processing time of each selection round while producing effective yields in the term of high affinity aptamers binding to the targets. Using asymmetric PCR could shorten the process of producing single-stranded DNA pool used for the next round of selection from a few days to one day of SELEX completion.[103-104]

Owing to the early by-product generation obtained from using conventional asymmetric PCR, the emulsion PCR was used to amplify the randomized DNA library under the optimal asymmetric PCR conditions including optimized PCR cycles and FW/RW primer ratio. Even though the emulsion PCR was expected to reduce the non-specific amplification, the aptamer selection was stopped at the sixth round because of emerging PCR by-products caused by PCR artifacts (such as a primer-dimer and primer partially hybridized DNA template). Also, the carryover of dsDNA as a minor product of asymmetric PCR and the high stringency used in the aptamer selection could create the competitive environment in the DNA-protein partitioning step contributing to insufficient amounts of targeted DNA functioning as a PCR template in the PCR reaction. To address this issue, the DNAs eluted from the protein-coated strips were concentrated by using a 10 kDa filter in order to remove the excessive amount of elution buffer before performing the PCR. Theoretically, the amplification of the homogenous DNA can yield an exponential increase of quantity of the PCR products until it reaches the maximum level and stays plateau. On the contrary, the PCR amplification of the enriched DNA pools containing the random DNA templates is more problematic which the quantity of the PCR products can drop from the maximum level down to zero.[105] This has pointed out how crucial to design a better structure of initial DNA library, especially the sequence of nucleobases in the primer-binding region, and to optimize individual PCR parameters (e.g., template

concentrations, primer concentrations, annealing temperatures and the effect of the polymerase) which significantly impact the PCR efficiency.[106] These can clearly explain why the PCR by-products could still occur while the emulsion PCR was used in this asymmetric PCR-SELEX.

For the extraction of PCR products obtained from the emulsion PCR, the PCR cleanup kit was applied to extract and purify the amplified DNAs which could yield better purity of DNA, faster process and less hazardous than using organic solvents such as diethyl ether.[107] In addition to using the DNA-binding buffer provided in the PCR cleanup kit, 2-butanol and nuclease-free water were firstly added prior to the centrifugation in order to facilitate the better separation of the oil phase suspended in 2-butanol from the aqueous phase containing the target DNA in water.

Notably, adding more selection rounds did not strongly guarantee that the aptamer with the best binding could be obtained since the affinity screening showed the increased binding of the enriched pool at the beginning of round 1 to round 3. Afterward, the binding to the N protein was still maintained whereas binding to the S1 protein was decreased after round 3. As mentioned above, the presence of the PCR by-product in the asymmetric emulsion PCR and the addition of extremely high stringency could possibly interfere with the binding efficiency between the targeted DNA aptamers and the proteins.[108-109]

Ultimately, it is essential to set up the criteria that can be used to choose the DNA sequences from the NGS raw data which could facilitate by means of eliminating the impact of PCR artifacts. As aforementioned criteria, the sequences with the high abundance of the last round of selection were diminished due to the incorrect constructions of the DNA sequences which contained repetitions of dimeric or trimeric sequences of FW and

RW primer binding regions located before or after the randomized region when compared to the correct construction of the initial N40 DNA library. Instead, the enriched sequences of all six rounds with the correct composition of the sequence structure were targeted and considered to be retrieved as aptamer candidates for binding affinity tests reported in Chapter 3.

Chapter 3

Characterization of Binding Affinity and Label-free Detection of SARS-CoV-2 Proteins by using Aptamer-based Bio-layer Interferometry (BLI)

A partial version of this chapter has been previously published in

Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics (Mol. Ther. Nucleic. Acids, Vol. 31, P731-743, March14, 2023)

Suttinee Poolsup, Emil Zaripov, Nico Hüttmann, Zoran Minic, Polina V. Artyushenko, Irina A. Shchugoreva, Felix N. Tomilin, Anna S. Kichkailo, and Maxim V. Berezovski

Author contributions for Chapter 3:

Suttinee Poolsup designed and performed the experiments and wrote materials and methods as well as results and discussion. Emil Zaripov produced the in-house expressed SARS-CoV-2 nucleocapsid protein in *E. coli* used for section 3.3.5.

Suttinee Poolsup contributed the results shown in Tables 3.1, 3.2, 3.3, 3.4 and Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11 and 3.12.

3.1 Introduction

Affinity and kinetics of DNA and protein interaction can be characterized via several techniques. Some standard methods can only determine the equilibrium dissociation constant (K_D) or IC_{50} values under the optimal conditions without providing the kinetic parameters such as on-rate (k_{on}) and off-rate (k_{off}).[110] Prior to characterizing the binding affinity of the selected aptamers and their target proteins, a method that is suitable for providing the expected information needs to be taken into consideration because the similar affinity constants or K_Ds can be yielded from the greatly different binding profiles of k_{on} and k_{off}. Recently, there are many techniques that have been developed in order to not only determine the thermodynamic metrics such as affinity, entropy and enthalpy of binding, but also kinetic parameters.

For instance, isothermal titration calorimetry (ITC) is one of the earliest techniques used to study the aptamer-protein binding affinity via a series of micro-injection by loading an analyte into the syringe while being titrated to the aptamer loaded in the isothermal cell. Then, the calorimeter is used to measure the absorbed or released heat of the binding event during each injection until the saturation of the titrated aptamer is reached at a certain point where the minimal heat changes are observed upon the injection of the titrant. The heat values are integrated with respect to time which the resulting data is plotted against the molar ratio of analyte-to-aptamer.[111] However, high concentrations of the aptamers (10 – 100 μM) and the analytes (at least 100 μM) are required to obtain high accuracy from ITC results which contributes to more cost and may affect the solubility of some hydrophobic analytes. In addition to the requirement of large quantity of reagents, the cost of the ITC is generally expensive compared to other low-throughput methods.

The other standard method that is the most widely used to determine the affinity and kinetic parameters of aptamers binding to the target proteins is surface plasmon resonance (SPR). This optical technique can yield the kinetic interaction of the aptamer and the target analyte through the change in resonance angle caused by the alteration of reflective index on the gold chip surface. This phenomenon occurs when the target molecule is interacting with the aptamer immobilized on the thin gold film resulting in the changes in the intensity of light reflected on the gold film surface at various angles.[112] Once the equilibrium stage of non-bound and target-bound aptamer is reached, the free analyte is washed away by the assay buffer circulating throughout the microfluidic system, then the refractive index returns to its initial values. The resonance angle change can be plotted in the pattern of resonance signal (RU) over time from which k_{on} and k_{off} can be determined and used to calculate the apparent K_D (Figure 3.1). Nevertheless, there are some challenging of using a microfluidic flow-based technique like SPR including a risk of cross-contamination of the components while running different samples and clogging if the unpurified or crude samples are loaded in the sample compartment and the tubing system.

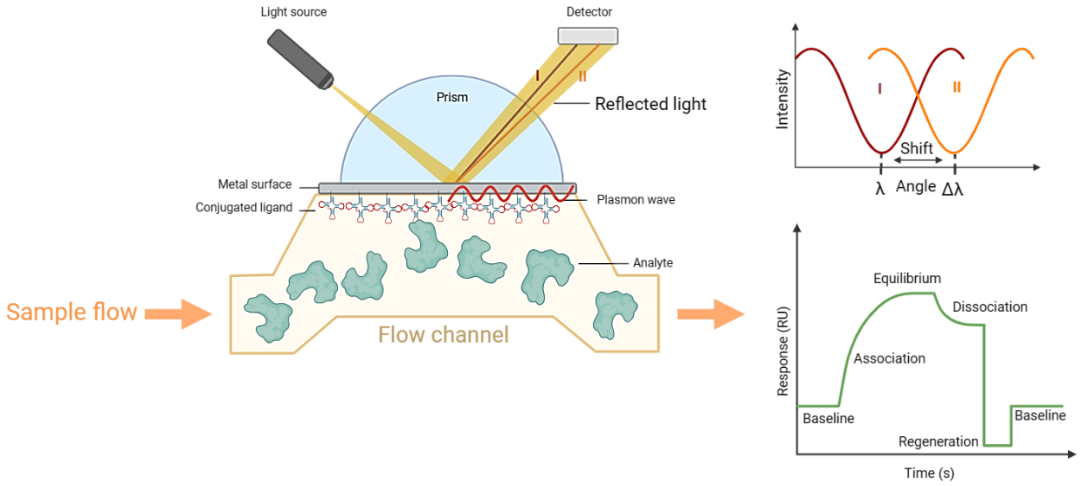


Figure 3.1 Schematic experimental setup of SPR. A gold sensor chip is placed on a prism. The light source shines polarized light on the sensor chip and the light is reflected which the intensity of the reflected light is measured in the detector. At a certain angle of incidence, excitation of surface plasmons occurs resulting in a dip in the intensity of the reflected light (I). A change of refractive index at the surface of the gold film coated on the sensor chip will cause an angle shift from I to II.

To overcome the challenges of using ITC and SPR, bio-layer interferometry (BLI) enables to characterize the binding affinity and calculate the K_Ds of the aptamers without consuming the samples (aptamers and its targets) which can be recovered after the assay is completed. Owing to the dip and read-based system using biosensors, BLI does not require the micro-injection, degassing the buffers as well as unclogging and flushing the tubing. This can relatively reduce the analysis time as well as the costs of reagents and consumables.[113] The BLI is an optical technique that analyzes the changes in interference patterns of the light waves when the binding of ligand and analyte occurs. To perform the analysis of the interaction between target biomolecules, a fiber-optic biosensor is

essentially required to attach the aptamer as a ligand prior to detecting the target analyte.[114] In principle, the biosensor is constructed with two interfaces which are an internal reference layer and a biocompatible layer on the biosensor surface. These interfaces are separated by a thin layer at the tip of the biosensor. When light reflects from each of the two layers, it creates constructive and deconstructive interference patterns that can be detected at the CCD array detector. To process this technique, the tip of the biosensor is dipped into a sample in either a tube or a 96-well microplate depending on the version of the instrument. After loading the ligand, the analyte then binds to the ligand-coated surface forming a molecular layer that increases in thickness as more molecules bind causing a shift in the interference pattern of the reflected light. This shift is monitored at the detector and reported on a sensorgram as a change in wavelength (nm) over time (second). By monitoring the interference pattern in real time, it is possible to obtain the kinetic data on the molecular interactions.[115-116]

Kinetic binding analysis is used to measure association (k_{on}) and dissociation (k_{off}) rate for non-covalent interaction which can be the combinations of electrostatic interactions, hydrogen bonds, van der Waals forces and hydrophobic effects. In the term of kinetics, the association rate of the interaction determines how fast one molecule of the analyte binds to a ligand, and the dissociation rate measures how fast a ligand-analyte complex falls apart from one another.[117] On the Octet N1 instrument used in this study, the interpretation of kinetic data is based on the 1:1 binding model of the biomolecular interaction where one aptamer or ligand molecule interacts with one analyte molecule, and the formation of the aptamer-analyte complex follows the pseudo-first order kinetics which association rate of the analyte binds equally and independently to all binding sites of the

ligand.[118] Once the binding of the analyte and the ligand reaches equilibrium, the association rate of the analyte-ligand complex formation is the same as the rate of the analyte dissociating from the ligand. This 1:1 binding model can be represented by Equation 1.

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AB \qquad \text{Eq.1}$$

(A: analyte, B: ligand, k_a: association rate and k_d: dissociation rate)

When the biosensor is dipped into the analyte solution, the rate of complex formation (AB) is measured in real-time which can be described in the terms of an association rate and the concentrations of unbound ligand and analyte. The rate of association is a function of the decreasing concentration of unbound ligand while the binding of the analyte occurs (Equation 2).

$$\frac{d[AB]}{dt} = k_a[A]_t[B]_t - k_d[AB]_t \qquad \text{Eq.2}$$

Equation 3 shows the integration of a differential equation used to fit the association phase of the sensorgram (Figure 3.3). In the association phase, the formation of the complex AB happens at the same time while it decays back to A and B so that k_{obs} reflects the overall rate of the combined association and dissociation of the analyte and the ligand (aptamer).

$$Y = Y_o + A(1 - e^{-k_{obs}*t}) \qquad \text{Eq. 3}$$

$$\left(A = R_{max}\frac{1}{1+\frac{k_d}{k_a*[Analyte]}}\right)$$

(Y=level of binding, Y_0 = binding at the starting time of association, A = an asymptote, t = time (second) and k_{obs} = the observed rate constant)

In the dissociation phase, the phenomenon that only occurs is the analyte decays from the AB complex since the biosensor is dipped into the assay buffer without having

the analyte. This phenomenon is reflected by the decline in the dissociation rate over time as the concentration of bound complexes on the surface decreases represented by Equation 4. The term k_d measured the fraction of the complexes decaying per second which is expressed in the unit of s^{-1}.

$$Y = Y_o + A(e^{-k_d * t}) \qquad \text{Eq. 4}$$

$$(\text{A} = R_{max} \frac{1}{1+\frac{k_d}{k_a * [Analyte]}})$$

Y_0 = the binding at the start of dissociation, k_d = the dissociation rate.

To obtain the values of k_{obs} and k_d, the concentration of the analyte is known which is required to added to the Octet software prior to performing the kinetic assay on the BLI instrument, and the software continuously determines the Y, Y_0 and the asymptote (A) at a certain time point and it chooses the best data that fits to the 1:1 binding model (R^2 is close 1) once the assay is completed per run. Thus, the Equations 3 and 4 presenting in Figure 3.3 can be solved to yield the values of k_{obs} and k_d, and then the association rate, k_a, can be calculated by using Equation 5.

$$k_a = \frac{k_{obs} - k_d}{[Analyte]} \qquad \text{Eq. 5}$$

The association rate represents the number of AB complexes formed per second in a 1 molar solution of A and B and is expressed in $M^{-1} \cdot s^{-1}$. With the calculated k_a and k_d obtained from the equations above, the equilibrium constant, K_D, can be calculated from the ratio of the off-rate to the on-rate by using k_a and k_d (Equation 6) which can be expressed in the unit of molar (M).

$$K_D = \frac{[A][B]}{[AB]} = \frac{k_d}{k_a} \qquad \text{Eq. 6}$$

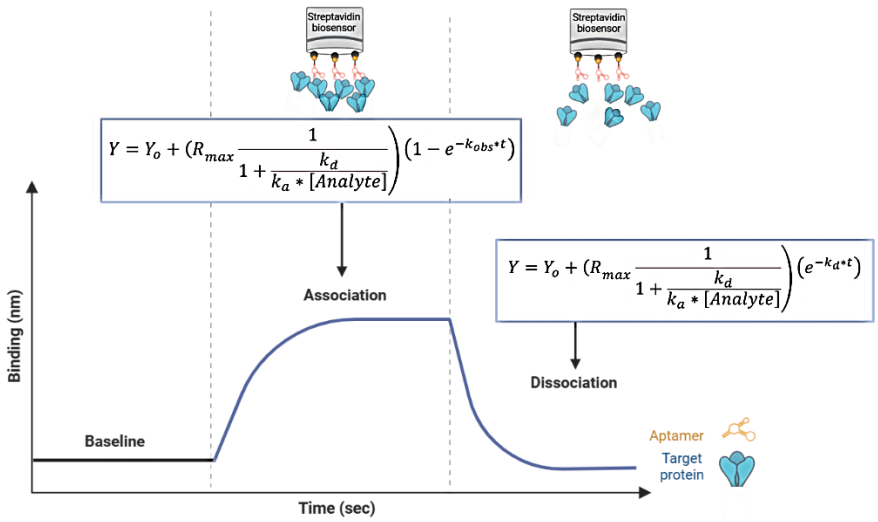


Figure 3.2 The summary of equations fitting the association and dissociation phases. The BLI scheme shows the relevant equations that the BLI software uses to calculate the k_a and k_d by choosing the data that fit 1:1 binding model in order to characterize the K_D of the kinetic binding between aptamers and the target proteins.

In this chapter, BLI is the main analytical technique for screening the affinity of the selected aptamers binding to the target proteins as well as the development of aptamer implemented on the BLI biosensor for assessment of cross-selectivity of the aptamers toward their targets over the non-target proteins, and mainly for the detection of SARS-CoV-2 proteins in human saliva.

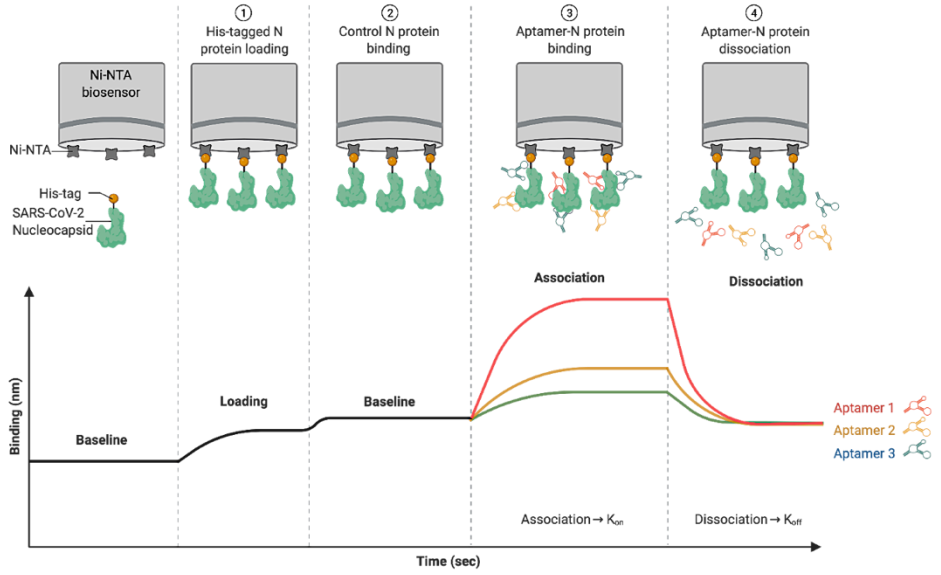
3.2 Materials and Methods

3.2.1 Affinity screening of DNA aptamers binding to SARS-Cov-2 nucleocapsid and spike S1 subunit proteins by BLI

The binding affinity of aptamer pools and SARS-CoV-2 nucleocapsid protein (ACROBiosystems, Newark, DE) was performed using Ni-NTA biosensors (Figure 3.4A) and a BLI instrument named Octet N1 (Sartorius, Bohemia, NY). Firstly, the DNA-

enriched pools were prepared in the assay buffer (DPBS with 0.5mM $MgCl_2$ and 1mM $CaCl_2$, pH 7.4, Gibco) at a concentration of 500 nM and heated at 95°C for 5 min, then ice-cooled. The binding assay was performed in 0.5 mL light-blocking tubes at 25°C with a total volume of 250 µL. In the next steps, 20 µg/mL of his-tagged N protein was loaded on Ni-NTA biosensor tips. During the binding assay, a baseline was established in the binding buffer followed by monitoring the association of the DNA-enriched pools with the N protein on the biosensors. After retrieving the sequences of the most enriched aptamers named NSP1, NSP2 and NSP3, 200 nM of these three aptamers with 80nt-length was prepared as mentioned above and the association step was performed to compare the binding shift of these enriched aptamers to the binding of the initial N40 DNA library and the 80nt-scrambled DNA (Sc80) was used as a negative control. The assay was carried on Octet N1 in the assay buffer at 25°C with a 30-s baseline, 100 s for the association, 150 s for dissociation, and glycine-HCl, pH 1.7 was used for biosensor regeneration. The published aptamers A48 and A58 were included in this experiment as positive controls.

A



B

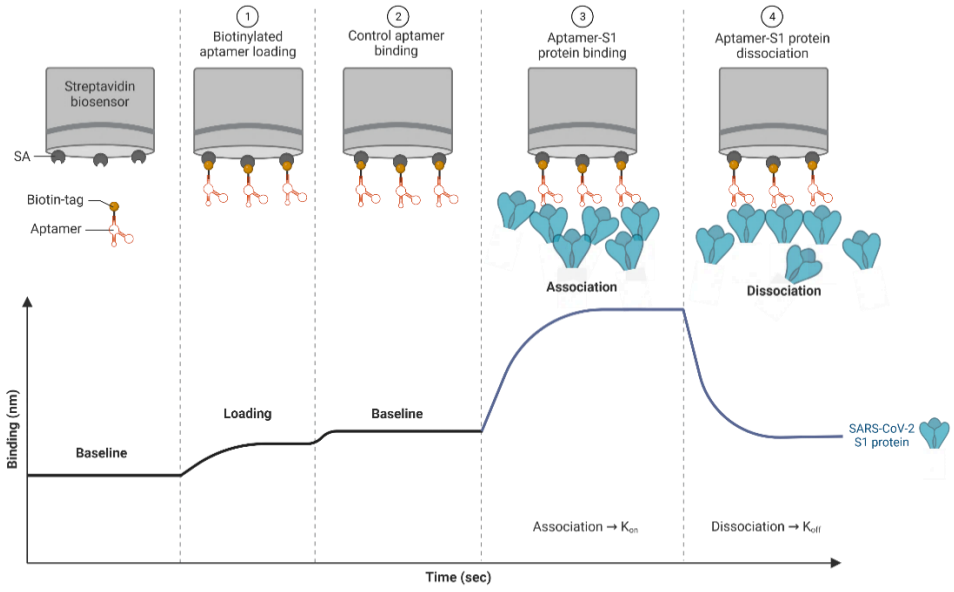


Figure 3.3 Schematic BLI workflow of aptamer-protein affinity tests. (A) Screening affinity of the enriched aptamers binding to the N protein on Ni-NTA biosensor which the his-tagged N protein was immobilized on the biosensor and the unlabeled aptamer

candidates were loaded onto the N-coated biosensor in the association step. (B) Affinity screening of the enriched aptamers targeting the S1 protein which the aptamers were labeled with each biotinylated forward or reverse primer at different times and immobilized on the streptavidin biosensor prior to loading the S1 protein in the association step on the aptamer-coated streptavidin biosensor (created by BioRender).

For the binding test of the selected aptamers against the SARS-CoV-2 spike S1 subunit, streptavidin biosensors were used on the Octet N1 (Figure 3.4B). To prepare biotinylated aptamers, the 80nt-length DNA sequences were hybridized with biotin-labeled FW primer and RW primer, separately, in DPBS at 95°C for 5 min and the temperature was gradually lowered to 25°C for 30 minutes, hold at 4°C for 30 minutes on the thermocycler. The ratio of biotin-labeled primers and the unlabeled 80nt-length DNA sequences was 1:1 (400 nM:400 nM) in the total volume of 100 µL. Then, the biotin-FW-hybridized DNA sequences and the biotin-RW-hybridized DNA sequences were prepared in the assay buffer as the same preparation as the preparation of the N protein-binding aptamers as mentioned above. Briefly, 400 nM of the biotinylated aptamer was heat-folded at 95°C for 5 min, then ice-cooled before loading as a ligand on a streptavidin biosensor. Then, the aptamer-loaded streptavidin biosensor was dipped into 20 µg/mL of the S1 protein prepared in the assay buffer with 300 s for association and 300 s for dissociation to evaluate the affinity of the selected aptamers to the S1 protein. The initial N40 DNA library and the scrambled DNA were also tested as negative controls.

3.2.2 Determination of K_Ds using BLI

The BLI experiments for determining the K_Ds were initially performed in two setups. One setup was to immobilize the his-tagged N protein (20 µg/mL) onto the Ni-NTA biosensors, and the non-labeled aptamers were loaded in the association step. The association step was performed with six different concentrations (0, 12.5, 25, 50,100, and 200 nM) of the 80nt-length aptamers (NSP1, NSP2, and NSP3) which were prepared in the assay buffer and heated at 95°C for 5 min, then ice-cooled prior to transferring 250 µL of the heat-folded aptamer solution to 0.5 mL light-blocking tubes. The binding assay was performed on Octet N1 at 25°C with a 30-s baseline, 100 s for the association, 150 s for dissociation, and Glycine-HCl, pH 1.7 was used for biosensor regeneration. Moreover, the truncated aptamers including tNSP1, tNSP2 and tNSP3 representing the most enriched sequences as well as the other four aptamers (named tNSP5, tNSP9, tNSP10 and tNSP12) representing each subgroup among 16 unique sequences were also evaluated their binding affinity to the N protein by performing under the same experimental setup as the 80nt-length aptamers.

Another BLI setup was to imitate the COVID-19 diagnosis; therefore, the binding assay was performed by immobilizing the aptamers on the biosensor tips to detect the presence of the N protein in the assay buffer. The serial dilutions of the N protein (7.69, 15.4, 30.8, 61.5 and 123 nM) were spiked in the optimized assay buffer (PBST, DPBS with 0.5mM $MgCl_2$ and 1mM $CaCl_2$, pH 7.4, added 0.2% BSA and 0.05% Tween 20). The detection was performed by dipping a streptavidin biosensor (Sartorius, Bohemia, NY), which was immobilized with the biotinylated aptamer, into the N protein spiked in the PBST buffer. The assay was carried out at 25°C with 180 s for association and 120 s for dissociation.

The assay buffer was used to wash out the unbound biomolecules from the biosensor tip during the baseline step for 60 s. The apparent dissociation constant (K_D) of each aptamer was calculated by Octet N1 software (version 1.3.0.5) using a 1:1 global binding model. Additionally, the binding affinity of tNSP3 aptamer to our in-house expressed SARS-CoV-2 nucleocapsid protein and the N proteins of Delta (7.69, 15.4, 30.8, 61.5 and 123 nM) as well as Omicron (7.73, 15.5, 30.9, 61.8 and 123 nM) variants was also performed under the same condition described above.

3.2.3 Cross-reactivity assessment of aptamers to different proteins

The selectivity of three aptamers (tNSP1, tNSP2, and tNSP3) to the N protein was examined by investigating the binding affinity of the aptamers to other viral proteins, including SARS-CoV-2 S1 protein and the nucleocapsid protein of MERS-CoV (ACROBiosystems, Newark, DE). Following the optimized condition of the binding assay as previously mentioned, the cross-reactivity was performed in the assay buffer (PBST, DPBS with 0.5mM $MgCl_2$ and 1mM $CaCl_2$, pH 7.4, added 0.2% BSA and 0.05% Tween 20) on Octet N1 at 25°C with a 60-s baseline, 180 s for the association and 120 s for dissociation. 400 nM of each biotinylated aptamer was immobilized on the streptavidin biosensor tips. The binding affinity of the aptamers to the non-targeted proteins (SARS-CoV-2 S1 and MERS N proteins) was compared to the SARS-CoV-2 N protein at the maximum concentration of 123.04 nM.

The cross-reactivity of S1-tSP10 was also investigated by testing the selectivity of the aptamer to the S1 protein of SARS-CoV-2 compared to MERS-CoV which is non-target protein at the concentration of 167 nM. The S1-tSP10 aptamer was prepared in the same preparation as the N-binding aptamer in the assay buffer as described above. The binding

assay was performed at 25°C with a 60-s baseline, 300 s for the association step and 300 s for dissociation step. The apparent K_D values were calculated using a 1:1 global binding model as mentioned above. To verify non-specific binding to the protein tag, a hexa-histidine peptide was also included as a control in the binding assay.

3.2.4 Heat inactivation effects on binding affinity of the aptamer and the target proteins

To study the effect of heat-inactivation on the aptamer-based detection of N proteins, 400 nM of the biotinylated tNSP3 aptamer was heated at 95°C for 5 min, then ice-cooled, and immobilized on the streptavidin biosensor tips. The binding assay of the aptamer to the N proteins of Delta (123 nM) and Omicron (123 nM) variants that were heated in the assay buffer at two different temperatures (65°C for 15 min and 80°C for 5 min) was performed on Octet N1 with a 60-s baseline, 180 s for the association and 120 s for dissociation.

3.2.5 Applications of aptamers-based BLI for SARS-CoV-2 protein detection in human saliva

The serial dilutions of N protein (7.69, 15.4, 30.8, 61.5, and 123 nM) were spiked in 10% saliva assay buffer, which was prepared by diluting the concentrated human saliva (Bioivt, Westbury, NY) in PBST buffer (DPBS with 0.5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.2% BSA and 0.05% Tween 20, pH 7.4). The detection was performed by immobilizing the biotinylated tNSP3 aptamer on a streptavidin biosensor tip. As mentioned in the previous assay, the N protein-spiked in 10% saliva was detected on Octet N1, and the 10% saliva assay buffer without adding the N protein was used as the background signal. The apparent K_D of the aptamer was calculated with Octet N1 software (version 1.3.0.5) using a 1:1

global binding model, and the limit of detection (LOD) was also determined based on the standard deviation (SD) of the binding curves and the slope of binding (nm) vs concentration of the N protein (nM), LOD = 3.3SD/ slope

3.3 Results

3.3.1 Affinity screening of the enriched aptamers binding the N protein

According to the highly enriched characteristic of the NSP1, NSP2, and NSP3 aptamers, the K_D values of the full-length structures of these three aptamers were determined using BLI. The initial N40 DNA library and the scrambled DNA with 80 nt were tested as controls. The DNA library showed low binding affinity and a low BLI shift (about 0.5 nm) to the N protein and the insignificant binding behavior of the scrambled DNA (Figure 3.4).

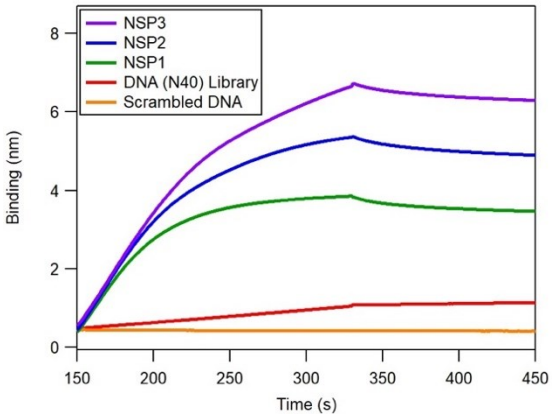


Figure 3.4 Binding evaluation on BLI. Binding affinity tests of NSP1, NSP2, and NSP3 aptamers with the N protein**.** The binding signals were detected in real-time as the association (180 s) and dissociation (120 s) of aptamers and the N protein. The initial DNA N40 library and scrambled DNA were used as controls.

3.3.2 K_D determination of the selected aptamers

To study the real-time binding assay on the BLI instrument (named Octet N1), serial dilutions of NSP1, NSP2, NSP3, and their truncated structures were prepared in the assay buffer with five different concentrations of aptamer from 200 nM to 12.5 nM. The reference assay was prepared in the absence of an aptamer. The his-tagged N protein was loaded on a Ni-NTA biosensor; then, the sensor tip was dipped into the aptamer solution as an analyte for association and dissociation steps. The results demonstrated the binding of full-length NSP1, NSP2, and NSP3 aptamers to the N protein with the K_D values of 7.86 ± 0.39, 4.38 ± 0.40 and 8.60 ±0.35 nM, respectively. The K_D values of truncated structures of these three aptamers named tNSP1, tNSP2, and tNSP3 were also determined to study the impact of primer-binding regions on the target recognition giving K_Ds of 4.53 ± 0.32, 3.04 ± 0.51, 5.77 ± 0.39 nM, respectively (Table 3.1). These results indicate that the primer-binding regions did not significantly affect the binding affinity. Also, the other four aptamers (tNSP5, tNSP9, tNSP10, and tNSP12) were truncated, and their K_Ds ranged from 6 to 22 nM (Table S2). In addition to the binding assay of the selected aptamers, the binding affinity of two positive controls as previously published (A48 and A58 aptamers with the reported K_Ds of 0.49 ± 0.95 and 0.70 ± 0.06 nM, respectively)[71] were measured on Octet N1, and their K_Ds were 6.46 ± 0.35 and 3.02 ± 0.32 nM, respectively. According to our study, Mg^{2+} and Ca^{2+} ions played a crucial role in maintaining the binding affinity between DNA aptamers and the protein compared to the Mg/Ca-omitted buffer condition so that these divalent ions were always added into the binding and assay buffers. Due to the high affinity of the three truncated aptamers (tNSP1, tNSP2, and tNSP3), they were further used for subsequent studies.

Table 3.1 K_Ds of truncated N-binding aptamers (tNSP1, tNSP2, and tNSP3) in the assay buffer. The N protein was immobilized on the Ni-NTA biosensor, and the aptamers were partitioned in the assay buffer while loading onto the N-coated biosensor in the association step. The K_D determination was performed in triplicates. The kinetic binding parameters of the aptamers to SARS-CoV-2 N protein were calculated by using a 1:1 binding model and reported in a format of mean ± standard error (SE).

Aptamer	Aptamer partitioning in the assay buffer		
	K_D (nM)	k_a (1/M·s)	k_d (1/s)
tNSP1	4.53 ± 0.32	(4.53 ± 0.07) x 10^5	(2.1 ± 0.19) x 10^{-3}
tNSP2	3.04 ± 0.51	(4.26 ± 0.10) x 10^5	(1.3 ± 0.38) x 10^{-3}
tNSP3	5.77 ± 0.39	(7.20 ± 0.15) x 10^5	(4.2 ± 0.41) x 10^{-3}

3.3.3 Binding evaluation of aptamers to SARS-CoV-2 N protein

To establish the detection of the N protein using aptamers, the affinity studies were also performed by individually immobilizing the biotinylated tNSP1, tNSP2, and tNSP3 aptamers on a streptavidin biosensor, SARS-CoV-2 N protein was spiked in the PBST buffer with the varied protein concentrations of 7.69, 15.4, 30.8, 61.5 and 123 nM, and the control assay was prepared in the absence of the protein. There is not a significant difference in K_D values obtained from either attaching the truncated aptamers or the N protein on the biosensors. The analysis setup with the aptamer-loaded sensor (Figure 3.5A) yielded a better curve fit ($R^2 = 0.999$) with the use of the default binding model (1:1) of the BLI software, which refers to the higher accuracy of the calculated K_D. Kinetic binding of these three aptamers (Figure 3.5B-D) showed a potential capability of N protein detection

in PBST buffer with the K_D of 4.83 ± 0.63, 4.51 ± 0.50, and 2.91 ± 0.34 nM for tNSP1, tNSP2, and tNSP3, respectively (Table 3.2).

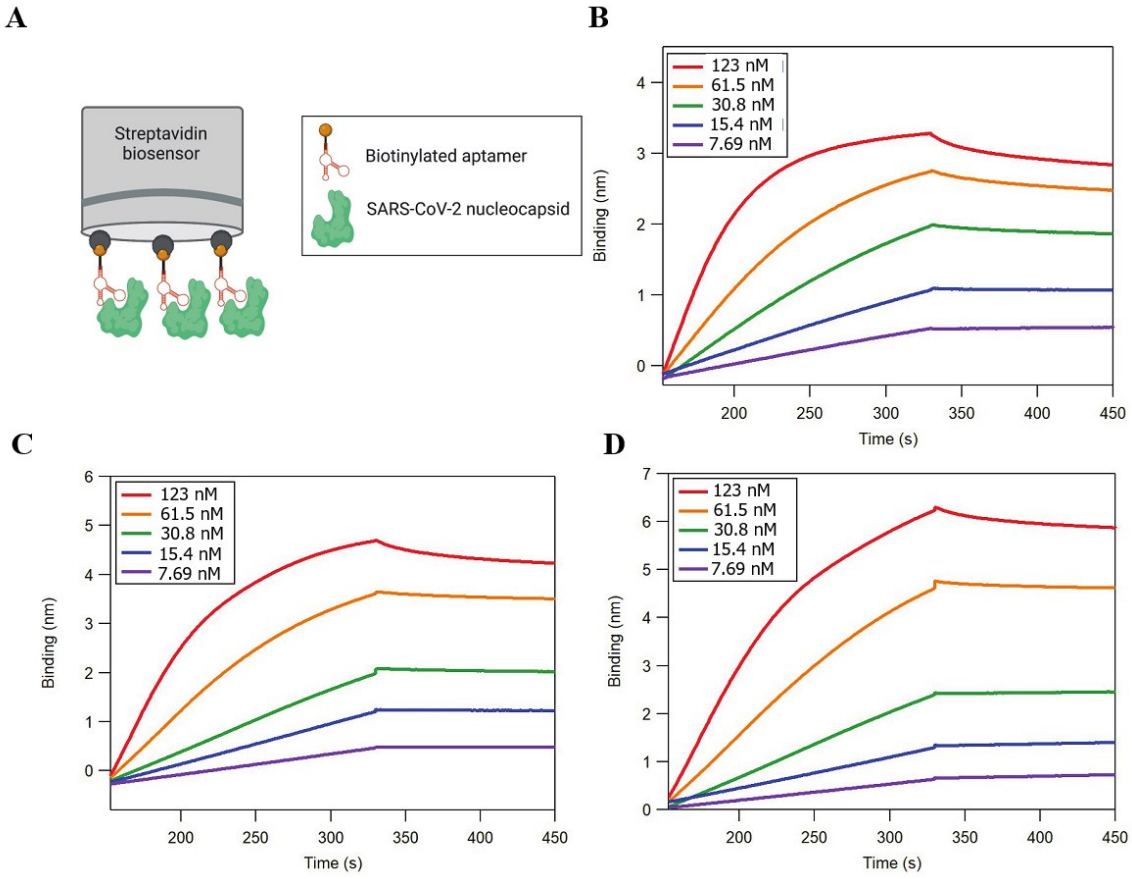


Figure 3.5 Real-time detection of aptamers binding to SARS-CoV-2 N protein. (A) Scheme of an aptamer-based BLI detection setup. The biotinylated aptamer was immobilized on a streptavidin biosensor and incubated with SARS-CoV-2 N protein. The binding signals were detected in real-time as the association (180 s) and dissociation (120 s) of aptamers and the N protein. BLI sensorgrams showed the increase of concentration-dependent target binding aptamers tNSP1 (B), tNSP2 (C), and tNSP3 (D) after being subtracted from the assay buffer with 0 nM of the N protein as a reference.

Table 3.2 K_Ds of truncated N-binding aptamers (tNSP1, tNSP2, and tNSP3) on SA biosensors. The biotinylated aptamer was immobilized on the streptavidin biosensor, and the N protein was partitioned in the assay buffer while loading onto the aptamer-coated biosensor in the association step. The K_D determination was performed in triplicates. The kinetic binding parameters of the aptamers to SARS-CoV-2 N protein were calculated by using a 1:1 binding model and reported in a format of mean ± standard error (SE).

Aptamer	Aptamer immobilized on the streptavidin biosensors		
	K_D (nM)	k_a (1/M·s)	k_d (1/s)
tNSP1	4.83 ± 0.63	(1.67 ± 0.04) x 10^5	(8.05 ± 0.15) x 10^{-4}
tNSP2	4.51 ± 0.50	(1.25 ± 0.10) x 10^6	(5.63 ± 0.30) x 10^{-4}
tNSP3	2.91 ± 0.34	(1.05 ± 0.01) x 10^5	(3.05 ± 0.04) x 10^{-4}

3.3.4 Assessment of aptamers selectivity for SARS-CoV-2 nucleocapsid

To verify the selectivity of the tNSP1, tNSP2, and tNSP3 aptamers to the SARS-CoV-2 N protein, their binding affinity was tested to other viral proteins, including SARS-CoV-2 S1 protein and the MERS-CoV N. The purity of each protein was visualized on SDS-PAGE (Figure S9), and each protein displayed a single band representing high homogeneity. All three aptamers exhibited strong binding to SARS-CoV-2 N protein (Figure 3.6A-C) with no significant cross-reactive binding to the SARS-CoV-2 S1. Some binding (about 50%) was observed to the N protein of MERS coronavirus, highlighting the higher affinity of these aptamers to SARS-CoV-2. Notably, these three truncated aptamers did not exhibit the binding to hexa-histidine peptide used as a tag on the recombinant proteins in this study. Considering the lowest K_D and highest selectivity to the N protein,

the tNSP3 aptamer was chosen for detecting the N protein in human saliva and identifying aptamer-protein binding epitopes.

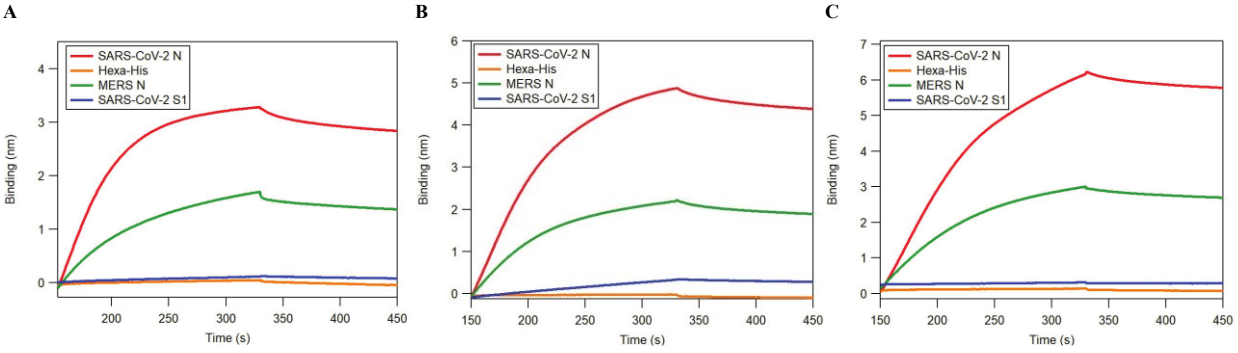


Figure 3.6 Selectivity of aptamers binding to SARS-CoV-2 N protein. BLI sensorgrams of association and dissociation of the aptamers (A) tNSP1, (B) tNSP2, and (C) tNSP3 with 20 µg/mL SARS-CoV-2 N, 20 µg/mL MERS-CoV N and 20 µg/mL SARS-CoV-S1 proteins in PBST buffer as well as 20 µg/mL hexa-histidine peptide as non-specific binding controls.

3.3.5 Assessment of tNSP3 aptamer binding to SARS-CoV-2 nucleocapsid in varied conditions

With different sources of the wild-type N protein production, the binding affinity of the tNSP3 aptamer to our in-house expressed N protein using *E. coli* was tested under the same condition and showed a K_D of 7.52 ± 0.17 nM (Figure S10) compared to the commercial recombinant protein modified with glycosylation and expressed in HEK293 cells. In addition to detecting the N protein obtained from wild-type SARS-CoV-2, the affinity study of our tNSP3 aptamer to the Delta and Omicron variants was also performed. The calculated K_Ds of the aptamer to the Delta and Omicron variants were 0.63 ± 0.10 and 3.55 ± 0.06 nM, respectively (Figure 3.7A and 3.7B). Taking the safety of the clinical laboratory into consideration, viral inactivation is currently required prior to processing the

samples reported by other studies.[37,38] Heat inactivation of the SARS-CoV-2 virus has been commonly used to treat the viral-containing serum.[39] Thus, the binding affinity of the tNSP3 to the N proteins in the BLI assay buffer (PBST) was investigated after heating the N protein of Delta and Omicron variants at 65°C for 15 min and 80°C for 5 min. The results showed that the incubation of the variants N proteins at 65°C had decreased about 20% and 40% binding signal of tNSP3 aptamer to the Delta and Omicron variants, respectively. Incubation at 80°C showed a similar effect on the binding affinity to the N protein of the Omicron variant. On the other hand, this resulted in a 60% loss of binding signal to detect the N protein of the Delta variant compared to the binding signals of nonheated N proteins (Figure 3.7C and 3.7D).

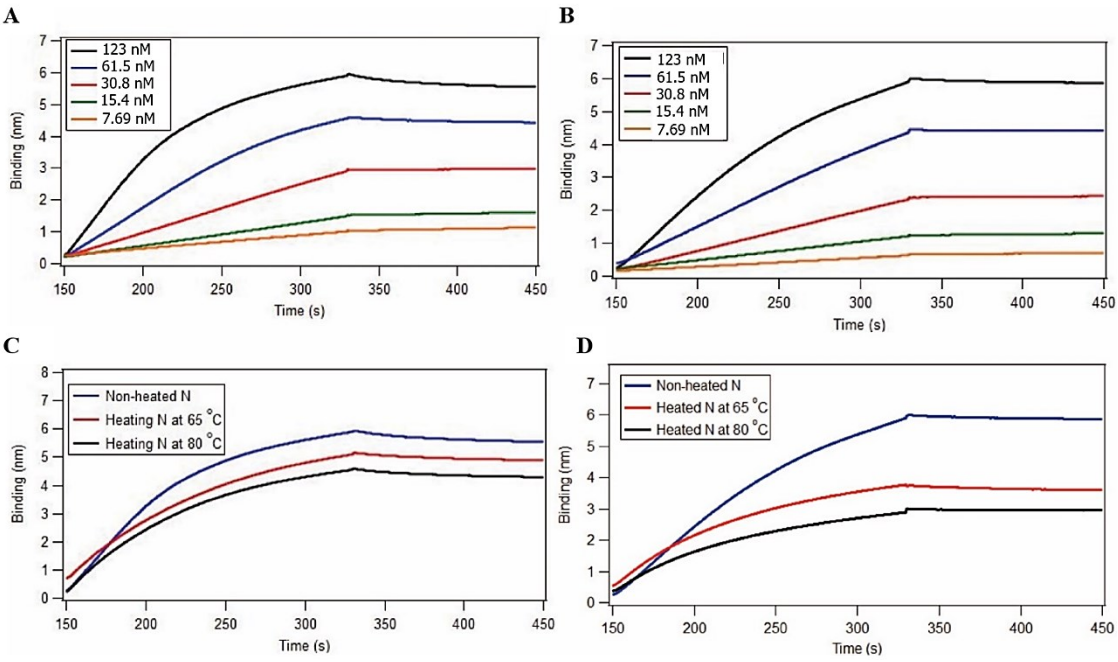


Figure 3.7 Detection of SARS-CoV-2 N protein of Delta and Omicron variants by tNSP3-based BLI. BLI sensorgrams show the increased signals of concentration-dependent target N proteins of (A) Omicron and (B) Delta variants binding to the tNSP3 aptamer after being subtracted from the assay buffer with 0 nM of the N protein as a

reference. Under the heat treatment, the binding affinity of tNSP3 aptamer to the N proteins of (C) Omicron and (D) Delta variants incubated at 65°C for 15 min and 80°C for 5 min.

3.3.6 Detection of SARS-CoV-2 nucleocapsid protein in human saliva using tNSP3 aptamer

To investigate the aptamer-based detection of COVID-19, the tNSP3 aptamer was used to detect the presence of SARS-CoV-2 N protein in human saliva. Concentrated saliva obtained from unspecified-gender pools was diluted in PBST buffer to 10%. The biotinylated aptamer was loaded on a streptavidin biosensor, and the testing saliva sample containing different concentrations of the N protein was completed with the aptamer. The binding measurement was achieved within 8 min per sample, and the control assay was performed with 10% saliva without N protein. Notably, the tNSP3 aptamer could detect the N protein in the human-pooled saliva with the K_D of 50.9 ± 0.43 nM (Figure 3.8A). The binding of the scrambled DNA to the N protein was not observed. In the controlled experiment, the streptavidin biosensor without tNSP3 aptamer was tested for binding with the highest concentration of N protein (123 nM) in the 10% saliva. Accordingly, the binding of the N protein to the biosensor was not observed, so the non-specific binding was negligible for this aptamer-based detection. Figure 3.8B shows the increased binding shifts with the concentrations of N protein in the range of 7.79 - 123 nM and LOD of 4.5 nM or about 6.77 x 10^{11} molecules in 1.13 pmol of the N protein.

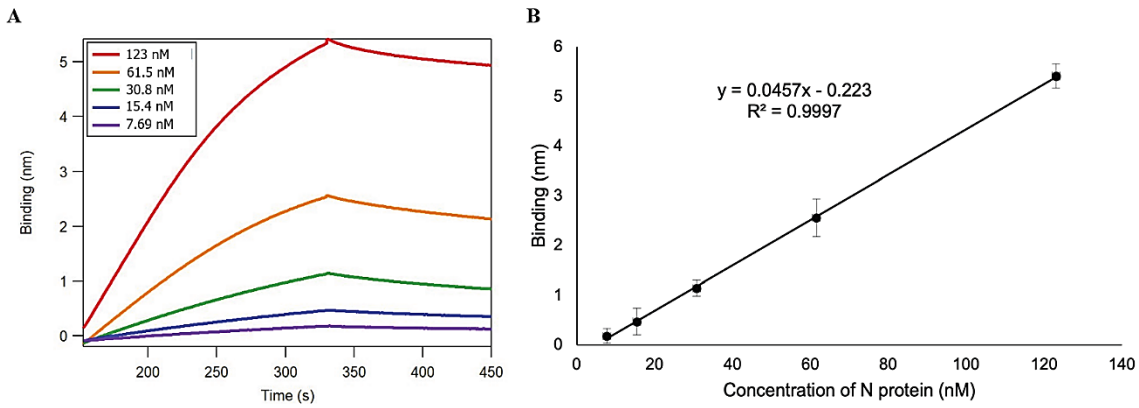


Figure 3.8 Aptamer-based BLI detection of SARS-CoV-2 N protein in human saliva. (A) BLI sensorgram displaying the binding assay of tNSP3 with the N protein spiked in human saliva with 2-fold dilution to determine the K_D. (B) Linear regression curve of the binding shifts corresponding to the N protein in the concentration range of 7.69 - 123 nM.

3.3.7 Binding evaluation of aptamers to SARS-CoV-2 S1 protein

Unlike the BLI experimental setup of the aptamers binding to the N protein using Ni-NTA biosensors, the SARS-CoV-2 spike S1 subunit has three times larger molecular weight than the size of the aptamers selected against the S1 protein. This might reduce the ability of evaluating the affinity of the selected aptamers binding to the S1 protein on BLI if the aptamers were loaded as ligands on Ni-NTA. Hence, the streptavidin biosensor was used to immobilize the biotinylated DNA sequences to measure the binding shift to the S1 protein. As a result, aptamer sequences named S1-SP4, S1-SP10 and S1-SP11 showed higher binding shifts than the other eleven aptamer clones in both primer-hybridization conditions (Figure S3 and S4). According to the hybridization of each primer onto the 5'-FW primer- and 3'RW primer binding regions of the 80nt-length aptamers, this could confirm that the primer binding regions do not play an important role in binding of the aptamer to the target protein as previously discussed in the study of the aptamers targeting

the N protein. Consequently, the truncated sequences of these three aptamers (Table 3.3) were used for the K_Ds determination by individually immobilizing the biotinylated aptamers named S1-tSP4, S1-tSP10, and S1-tSP11 on a streptavidin biosensor (Figure 3.9A). Then, the SARS-CoV-2 S1 protein was spiked in the PBST buffer with the varied protein concentrations of 10.4, 20.8, 41.7, 83.3 and 167 nM, and the control assay was prepared in the absence of the protein. The kinetic binding assay (Figure 3.9B-D) with the use of the default binding model (1:1) of the BLI software yielded the K_D values of 58.9 ± 0.90, 14.2 ±0.70 and 27.7 ± 0.43 nM for S1-tSP4, S1-tSP10, and S1-tSP11, respectively (Table 3.4). According to the high affinity of the S1-tSP10 aptamer to the S1 protein showing the lowest K_D among other selected sequences, the S1-tSP10 aptamer was further used in the following studies.

Table 3.3 Truncated sequences of the S1-binding aptamers. The sequences of selected aptamers binding to the S1 protein were shortened based on the prediction of their secondary structures obtained from RNAstructure web server.[98]

Aptamer	Aptamer sequence
S1-tSP4 (48nt)	CACGTAATGCCTAACTCTTTTTGTGTTTGCGATCTTTGCACATAGCAT
S1-tSP10 (40nt)	CACACTTTCTGCCCGCCTTCTCCCTCCGTTCCCCTCCCCG
S1-tSP11 (45nt)	ATGTCCTCGCACACCCAAACGCACTCATCTCCCCACCCATGCATA

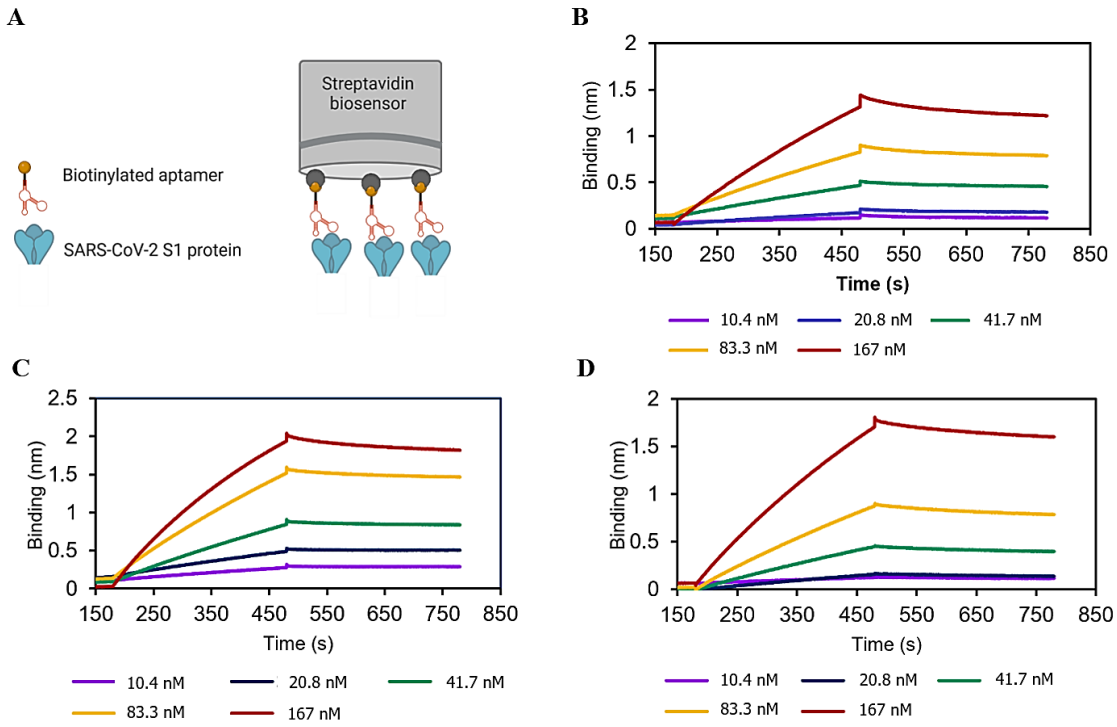


Figure 3.9 Real-time detection of aptamers binding to SARS-CoV-2 S1 protein. (A) Scheme of an aptamer-based BLI detection setup. The biotinylated aptamer was immobilized on a streptavidin biosensor and incubated with SARS-CoV-2 S1 protein. The binding signals were detected in real-time as the association (300 s) and dissociation (300 s) of aptamers and the N protein. BLI sensorgrams showed the increase of concentration-dependent target binding aptamers S1-tSP4 (B), S1-tSP10 (C), and S1-tSP11 (D) after being subtracted from the assay buffer with 0 nM of the S1 protein as a reference.

Table 3.4 KDs of truncated aptamers (S1-tSP4, S1-tSP10, and S1-tSP11) binding to SARS-CoV-2 S1 subunit protein. The S1 protein was partitioned in the assay buffer while loading onto the aptamer-coated streptavidin biosensor in the association step. The K_D determination was performed in triplicates. The kinetic binding parameters of the aptamers

to SARS-CoV-2 S1 protein were calculated by using a 1:1 binding model and reported in a format of mean ± standard error (SE).

Aptamer	K_D (nM)	k_a (1/M·s)	k_d (1/s)
S1-tSP4	58.9 ± 0.90	(1.16 ± 0.50) x 10^4	(6.78 ± 0.20) x 10^{-4}
S1-tSP10	14.2 ± 0.70	(2.79 ± 0.31) x 10^4	(3.99 ± 0.51) x 10^{-4}
S1-tSP11	27.7 ± 0.43	(1.08 ± 0.43) x 10^5	(5.09 ± 0.30) x 10^{-4}

3.3.8 Assessment of aptamers selectivity for SARS-CoV-2 S1 subunit

Similar to the N protein-binding aptamers, the selectivity of S1-tSP10 aptamer binding to the SARS-CoV-2 S1 protein was also assessed by comparing the binding shift (nm) of the aptamer to the S1 subunit protein of MERS-CoV and hexa-histidine peptide tagged on the C-terminal of the S1 protein. The binding of S1-tSP10 aptamer to hexa-histidine peptide was not observed. In particular, this aptamer showed stronger binding to the S1 protein of SARS-CoV-2 S1 over MERS-CoV which possessed about 30% binding onto the S1-tSP10 aptamer (Figure 3.10). This might be caused by the similarity of the amino acid sequences of the S1 subunit of SARS-CoV-2 and MERS-CoV which was further discussed in Chapter 4 (section 4.3.3).

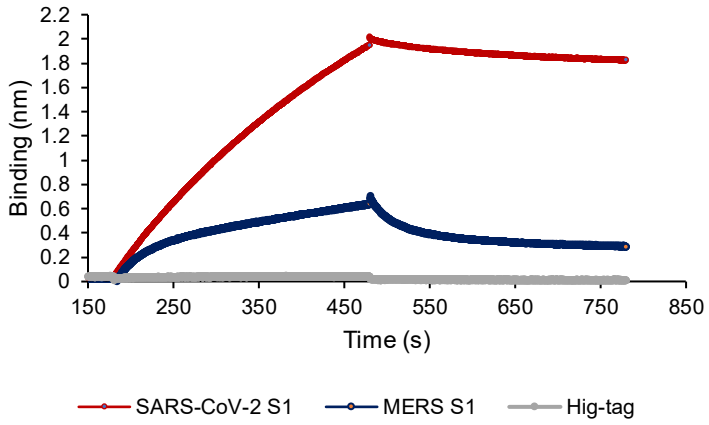


Figure 3.10 Selectivity of aptamers binding to SARS-CoV-2 S1 protein. BLI sensorgrams of association and dissociation of the S1-tSP10 aptamer with 20 µg/mL wildtype SARS-CoV-2 S1 and 20 µg/mL MERS-CoV S1 in PBST buffer as well as 20 µg/mL hexa-histidine peptide as a control of the non-specific binding.

3.3.9 Assessment of S1-tSP10 aptamer binding to SARS-CoV-2 S1 of Omicron variant in varied conditions

Additionally, the affinity study of the S1-tSP10 aptamer to the Omicron variant was also observed on Octet N1. The calculated K_Ds of the aptamer to the Omicron variant was 15.9 ± 0.25 nM, (Figure 3.11A). Besides, the effect of heat treatment on the binding affinity of the S1-tSP10 to the S1 protein of the Omicron variant was also investigated after heating the S1 protein at 65°C for 15 min and 80°C for 5 min. As the BLI sensorgrams shown in Figure 3.11B, the incubation of the S1 protein at 65°C and 80°C did not decrease the binding signal of the S1-tSP10 aptamer to the heat-inactivated S1 protein of the Omicron variant. This could be evidence of the high stability of S1 protein of Omicron variant which was detected by using this S1-tSP10 aptamer implemented BLI approach.

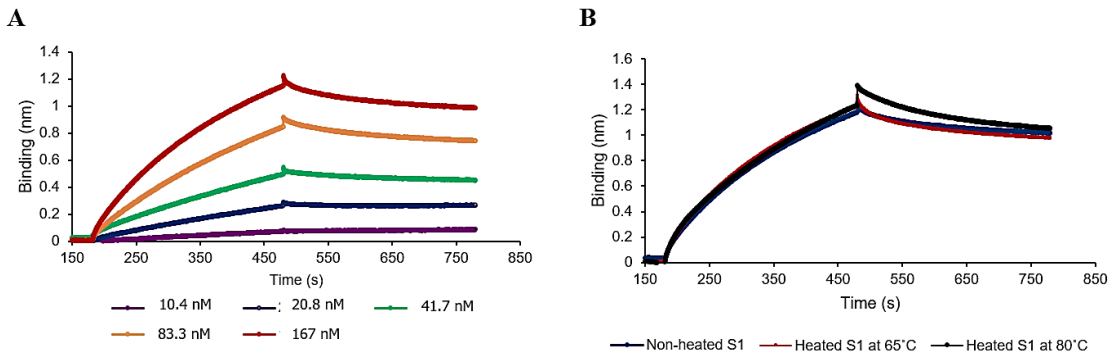


Figure 3.11 Detection of SARS-CoV-2 S1 protein of Omicron variant by S1-tSP10 based BLI. (A) BLI sensorgrams show the increased signals of concentration-dependent target S1 proteins of the Omicron variant binding to the S1-tSP10 aptamer after being subtracted from the assay buffer with 0 nM of the S1 protein as a reference. (B) the binding affinity of S1-tSP10 aptamer to the S1 proteins of the Omicron variant was observed under the heat treatment at 65°C for 15 min and 80°C for 5 min.

3.3.10 Detection of SARS-CoV-2 S1 protein of Omicron variant in human saliva using S1-tSP10 aptamer

The S1-tSP10 aptamer was similarly used to detect the SRS-CoV-2 S1 protein of Omicron in human saliva. The biotinylated aptamer was loaded on a streptavidin biosensor, and the different concentrations of the S1 protein prepared in 10% saliva diluted in the PBST assay buffer was tested. The binding measurement was completed within 14 min per sample, and the control assay was performed with 10% saliva without the S1 protein. As a result, the S1-tSP10 aptamer could detect the S1 protein in the concentration range of 10.4 - 167 nM spiked in the human-pooled saliva and the K_D was calculated at 127 ± 0.84 nM (Figure 3.12A). Additionally, the non-specific binding was also tested as a control experiment. The highest concentration of the S1 protein in the saliva assay buffer did not

show the increase of the association curve binding to the bare streptavidin biosensor without S1-tSP10 aptamer, so the non-specific binding was negligible for this aptamer-based detection of the S1 protein. The limit of detection (LOD) was calculated as 19.2 nM (Figure 3.12B).

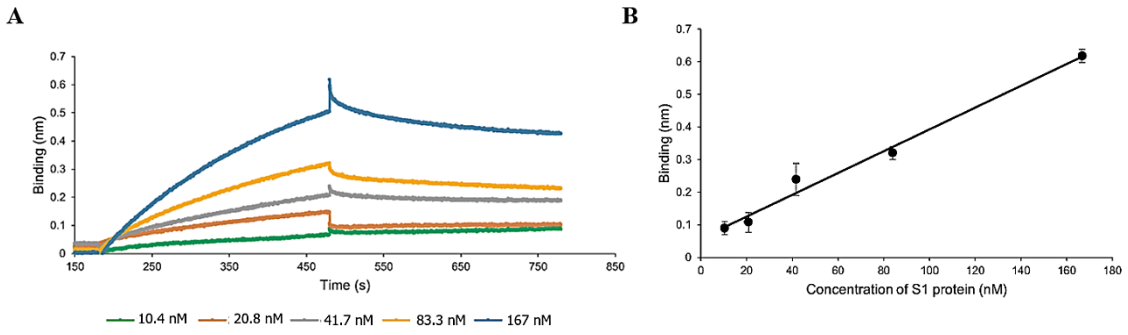


Figure 3.12 Aptamer-based BLI detection of SARS-CoV-2 S1 protein in human saliva. (A) BLI sensorgram displaying the binding assay of S1-tSP10 aptamer with the S1 protein spiked in human saliva with 2-fold dilution to determine the K_D. (B) Linear regression curve of the binding shifts corresponding to the S1 protein in the concentration range of 10.4 – 167 nM.

3.4 Discussion

During the early pandemics, antibody-based tests were widely used to detect the SARS-CoV-2 antigens, especially the spike protein which could be able to detect the viral antigens within the K_D range of 1 – 10 nM. However, the less sensitivity and low reproducibility as well as high cost in production of using the monoclonal antibody as a capture and/or detection molecule in the antigen tests have inspired to develop other antigen-recognition molecules such as aptamers. Within the past few years, the aptamers targeting N and S1 proteins have been rapidly generated from many laboratories around

the world with the similar research perspectives of developing the COVID-19 diagnostics, and these aptamers can alternatively be utilized for blocking viral entry in the terms of therapeutic applications.

In this study, BLI is the main technique used for affinity screening, kinetic assay for determination of K_Ds and the development of aptamer-based label-free biosensor for the detection of SARS-CoV-2 N and S1 proteins. The effect of nucleotide length of aptamer sequences on the binding affinity was first compared between full-length sequences with 80 nucleotides versus the shortened lengths of the aptamers. The truncated structures of these selected aptamers evinced a high binding affinity to the target proteins with a similar range of K_D values. The binding affinity of truncated structures was not significantly affected by trimming out the primer-binding region. While aptamers with their full length may form various unfavorable secondary structures, the smaller size of aptamers renders the higher binding affinity due to less steric hindrance of the aptamer structures.[119,120] These findings were coherent with previously reported studies.[121]

The three truncated aptamers targeting the N protein did not exhibit cross-reactivity to the SARS-CoV-2 S1 subunit and hexa-histidine peptide. The N-binding aptamers (tNSP1, tNSP2 and tNSP3) revealed a two-fold decreased binding to the nucleocapsid protein of MERS-CoV while the S1-binding aptamer (S1-tSP10) displayed a three-fold decline in binding to the S1 subunit of MERS-CoV. This can be explained by the similarity of peptide sequences between SARS-CoV-2 and MERS-CoV, as previously reported.[122,123] To prevent false-positive detection of our aptamer-based BLI detection to proteins from MERS-CoV, a cut-off diagnostic threshold may be set up by using the binding signal of the aptamer to the MERS-CoV N protein as a background signal.

In the comparison of the calculated K_Ds published in the previous studies (Table 3.5), the K_D of the tNSP3 aptamer was higher than the K_Ds of the A48 and A58 aptamers reported by Zhang et al.[71] which performed the K_D determination via SPR. Using the different techniques to determine the K_D values could significantly provide the different K_Ds. To evaluate this hypothesis, the kinetic assays with the use of A48 and A58 aptamers were performed on BLI technique and their K_Ds were calculated as 6.5 nM and 3.1 nM, respectively. This has proved that our tNSP3 aptamer showed relatively similar K_D to the A48 and A58 aptamers binding to the N protein (Table S2). Additionally, the dissociation rate of tNSP3 displayed 13 times slower than A48 aptamer and about 3 times slower than A58 aptamer which evidences the affinity of the tNSP3 aptamer binding to the N protein.

Likewise, the K_Ds of the aptamers targeting the S1 subunit of spike protein were determined using the different methods. While comparing the K_Ds obtained from the BLI technique, our S1-binding aptamers, especially S1-tSP10, showed lower K_Ds than SNAP1.50 and SNAP1.66 aptamers.[66] Hence, our selected aptamers targeting the N and S1 proteins can be comparable to the previously published aptamers. In accordance with the K_D values of our aptamers, the LODs of the aptamer-based label-free BLI technique using tNSP3 and S1-tSP10 were determined to be 4.50 nM and 19.2 nM for detecting the N and S1 protein, respectively, in human saliva. Even though it is quite challenging to precisely compare the limits of detection between different procedures and the LOD of the aptamer-based label-free BLI technique did not exhibit as the best method with the highest sensitivity, the aptamer-based BLI technique can still be amendable to detect the spike protein which has the molar concentration of 5.68 ± 0.22 μM in the SARS-CoV-2 spike protein reference material reported by the National Research Council Canada.[124]

Moreover, our aptamers can be further applied to other analytical techniques in order to improve the sensitivity of the detections.

Table 3.5 The summary of representative DNA aptamers targeting SARS-CoV-2 N and S1 subunit proteins for aptamer-based biosensors. The KDs of previously published aptamers are reported and the aptamers with the lowest KD were implemented onto the biosensors for the detection of SARS-CoV-2 proteins.

Target Protein	DNA Aptamer*	K_D (nM)/ Method used for K_D determination	Biosensor type (LOD)	Ref.
SARS-CoV-2 Nucleocapsid	A15	4.38 ± 0.06 (SPR assay)	ELISA and LFA with anti-Np antibody (LOD: 20 pM detected in human urine)	[71]
	A48	0.49 ± 0.05 (SPR assay)		
	A58	0.70 ± 0.06 (SPR assay)		
	A61	2.74 ± 0.08 (SPR assay)		
	tNSP1	4.83 ± 1.09 (BLI assay)	Label-free BLI (LOD: 4.5 nM detected in human saliva)	This study [148]
	tNSP2	4.51 ± 1.50 (BLI assay)		
	tNSP3	2.91 ± 1.02 (BLI assay)		
SARS-CoV-2 S1 subunit	SNAP1.50	35.8 ± 0.16 (BLI assay)	LFA (LOD: 250 pM detected in assay buffer)	[66]
	SNAP1.66	31.7 ± 0.16 (BLI assay)		
	XN-268s	4.26** (SPR assay)	Nanochannels (LOD: 1 fM detected in assay solution)	[126]

	CoV-S1-Apt1	49.71** (SPR assay)	AuNPs colorimetric assay (LOD: 3.1 nM detected in human serum)	[125]
	MSA1T	2.8 ± 0.5 (dot-blot assay)	Colorimetric sandwich assay (LOD: 400 fM*** in human saliva)	[64]
	MSA5T	10.1 ± 1.4 (dot-blot assay)		
	S1-tSP4	58.9 ± 2.70 (BLI assay)	Label-free BLI (LOD: 19.2 nM in human saliva)	This study
	S1-tSP10	14.2 ± 2.10 (BLI assay)		
	S1-tSP11	27.7 ± 1.35 (BLI assay)		

* The aptamer sequences and their secondary structures are presented in Appendix (Table S3, Figure S11 and S12).
** The standard error of the calculated K_D was not reported in the literature.
*** The LOD of this aptamer-based colorimetric assay was determined from the detection of the pseudoviral particles unlike others that detected the purified recombinant S1 protein.

Surprisingly, the tNSP3 aptamer could bind to the wild type similarly to the Omicron variant with the K_D below 4 nM, and the K_D of 0.6 nM to Delta variant. Also, the binding of the S1-tSP10 to the S1 subunit of wild type showed a similar K_D to its binding to the Omicron variant. This means that the N and S1 subunit proteins are thermally stable, and this emphasizes the compatibility of the tNSP3 and S1-tSP10 to the safety protocol involving the viral deactivation at 65℃ and 80℃ for SARS-CoV-2 detection in the clinical samples.

In addition to the aptamers selected through the asymmetric-ePCR SELEX, the apparent K_Ds of the N-binding aptamers selected by using symmetric or conventional PCR were also determined using BLI. The symmetric-ePCR SELEX targeting the N protein was performed with ten rounds of selection, and this method yielded the aptamer

with the lowest apparent K_D of 6 nM (Table S6). This has supported the statement discussed in Chapter 2 that increasing the number of selection rounds did not improve the better binding affinity of the selected aptamers. Also, using asymmetric PCR could effectively reduce the time consumption of aptamer selection over the symmetric PCR with a similar outcome in the term of binding affinity to the target protein.

Apart from the development in the terms of methodology of asymmetric-emulsion PCR, virus-based SELEX or viro-SELEX is one of the promising methods which has been currently used to select the aptamers targeting the spike protein on the surface of the viral membrane. With this viro-SELEX, the initial DNA library is partitioned directly onto the native conformation of the spike protein. This is beneficial for improving the better affinity and specificity of the selected aptamers to the structural proteins instead of using the purified proteins which might be partially modified during the protein expression. Rahman et al.[127] demonstrated the aptamer selected from the modified nucleotides, 5-[N-(1-naphthylmethyl) carboxamide]-20 -deoxy uridine (Nap-dU), by using the combination of the virus-based SELEX, which the wild-type spike protein-modified baculovirus (surrogate virus) was used as a target, and the conventional SELEX targeting the purified trimeric spike protein of wildtype. The K_Ds was determined by using the filter-based binding assay and the best two aptamers, named AM0320 (80 nt) and AM0470 (52nt), binding to the trimeric spike protein of wild-type yielded the K_D values of 9.50 ± 1.40 nM and 0.11 ± 0.02 nM, respectively.[128]

Chapter 4

Identification of binding epitopes using nLC-MS/MS

A partial version of this chapter has been previously published in

Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics (Mol. Ther. Nucleic. Acids, Vol. 31, P731-743, March 14, 2023)

Suttinee Poolsup, Emil Zaripov, Nico Hüttmann, Zoran Minic, Polina V. Artyushenko, Irina A. Shchugoreva, Felix N. Tomilin, Anna S. Kichkailo, and Maxim V. Berezovski

Author contributions for Chapter 4:

Suttinee Poolsup designed the experiment and prepared samples for mass spectrometry. Nico Hüttmann designed and analyzed the data, and Polina V. Artyushenko performed MD simulations. Suttinee Poolsup, Nico Hüttmann and Polina V. Artyushenko wrote materials and methods as well as results.

Suttinee Poolsup contributed Figure 4.1

Nico Hüttmann and Abdullah Khraibah contributed Figure 4.2 and 4.4

Polina V. Artyushenko contributed Figure 4.3

4.1 Introduction

With the complex structure of the target proteins, the binding epitopes of the proteins where the aptamer binds remain in doubt. Despite the applications of the aptamer in diagnosis as previously reported, aptamers have been used as a promising tool to discover biomarkers. The technology of aptamer-facilitated biomarker discovery (AptaBiD)[87] was initiated by Berezovski and his co-workers to utilize the enriched aptamer pools in identifying the biomarkers on the cell surface. In recent years, the combination of affinity-mass spectrometry (affinity-MS) has been fascinating as a hybrid tool for the characterization of molecular interaction.[129,130] In this study, the development of the affinity-MS approach is based on the hypothesis that the aptamer will bind an epitope sequence that is then extracted from the peptide pool after being cleaved by proteolytic digestion. The epitope excision consists of binding the biotin-labeled aptamer to the target protein followed by adding a specific protease. The complexes of aptamer-peptide fragments are formed and captured on streptavidin beads. The aptamer-unbound fragments are washed away, and the complex of the peptide-bound aptamer is denatured to dissociate the peptide epitope by urea denaturation. This elution fraction containing the peptide is purified and analyzed through mass spectrometry to determine the peptide sequence of the epitope. Herein, the affinity-mass spectrometry approach can be a key element to identify the epitope structure of the SARS-CoV-2 target proteins.

4.2 Materials and Methods

4.2.1 Pull-down assay

First, the biotinylated aptamers were heat-folded at 95°C for 5 min in the incubation buffer (DPBS with 0.5mM $MgCl_2$ and 1 mM $CaCl_2$, 0.2 μg/μL tRNA) prior to use. 20 pmol

of SARS-CoV-2 N and S1 proteins in 25 mM HEPES, pH 7.5, were incubated with 100 pmol of biotinylated tNSP3 aptamer and S1-aptamers (S1-tSP4, S1-tSP10 and S1-tSP11), respectively, for 30 min at 25°C. Then, the DNA-protein mixture was transferred to a Microcon-30kDa filter (Millipore, Burlington, MA) and centrifuged at 14,000 g for 10 min to clean up the unbound DNAs. The solution of aptamer-protein complexes retained on the filter was transferred to a 1.5 mL microcentrifuge tube. The aptamer-bound protein complexes were digested by adding 1.5 µL of 0.2 ug/µL trypsin/Lys-C, followed by incubation at 37°C for 30 min. The digested aptamer-protein complexes were added to the pre-washed Dynabeads MyOne streptavidin magnetic beads (Invitrogen, Waltham, MA) according to the manufacturer's protocol. Briefly, the biotinylated aptamer-digested peptide complexes were incubated with the streptavidin magnetic beads on a rotator at room temperature for 15 min, and the unbound biotinylated aptamer was washed out twice with the washing buffer (DPBS with 0.5mM $MgCl_2$ and 1mM $CaCl_2$). Then, 100 µL of 8 M urea was added and incubated for 10 min in order to elute the aptamer-bound peptides at room temperature. The eluted peptides were collected, and 2 µL of formic acid was added to the sample, then stored at -20°C until further processing (Figure 4.1). Another control sample was initially performed by incubating the heat-folded scrambled DNA with the N protein for 30 min at 25°C before transferring the aptamer-protein complexes to the Microcon-30kDa filter. The digestion of aptamer-protein complexes was performed using trypsin as described previously. Each sample was prepared in triplicates for statistical analysis.

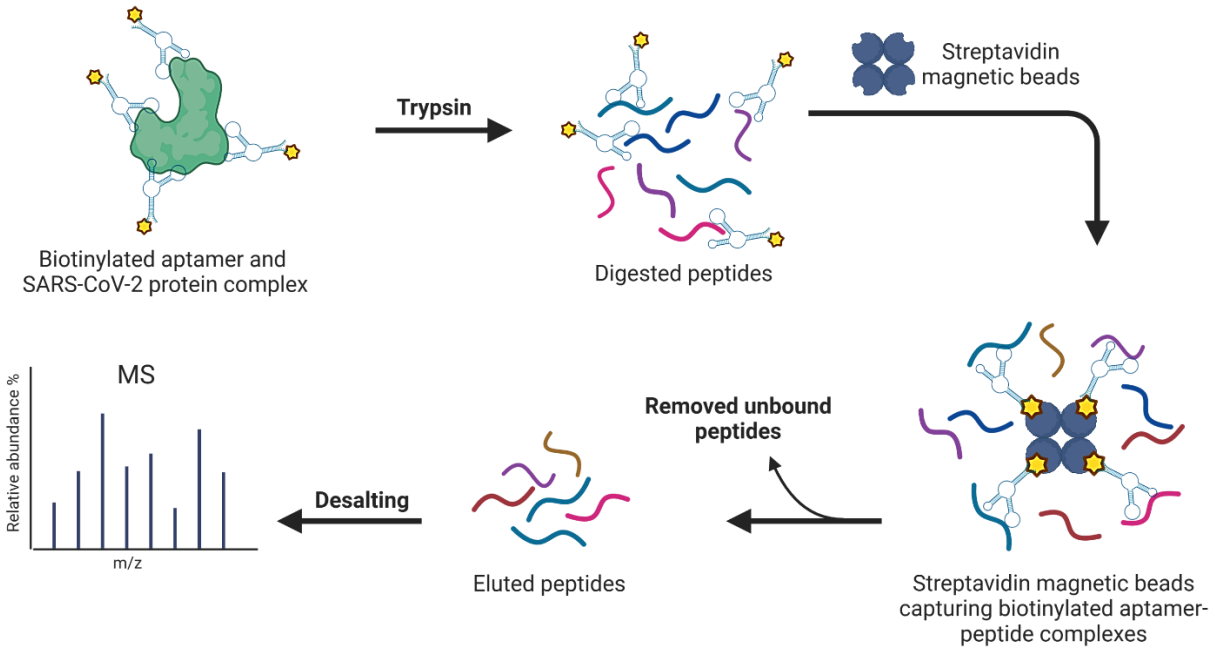


Figure 4.1 Scheme of pull-down assay for binding epitope identification. The experimental workflow used for isolating binding epitopes of the target proteins using the selected aptamer and the digested peptide fragments were analyzed by using nLC-MS/MS.

4.2.2 Sample preparation for nLC-MS/MS

The digested peptides were cleaned up by desalting on C18 stage tips (TopTip) before injection into the nLC-MS/MS following the manufacturer's protocol. Briefly, C18 TopTips were pre-washed by adding 50 µL of 70%ACN/0.1%FA and using a syringe to push through the washing reagent. The pre-washing step was repeated three times followed by washing with 50 µL of 0.1% FA three times. Then, the digested peptide solution was slowly transferred to the pre-washed C18 TopTips at a maximum volume of 70 µL. This step was repeated again until all the sample solution was loaded to the C18 TopTips. After that, the unbound peptides were washed out from the tips three times with 50 µL of 0.1% FA each time. To elute the bound peptides from the C18 TopTips, 50 µL of

70%ACN/0.1%FA was added to the tips and the flow-through was collected in a 1.5 mL tube, repeated three times prior to drying the eluted samples on a speed-vac concentrator for 3 hours at room temperature. The evaporated peptide samples could be stored at -20°C until further application onto nLC-MS/MS proceeded by John L. Holmes Mass Spectrometry Facility (JHMSF).

4.2.3 Epitope identification by proteomics data analysis

MS raw files were analyzed with MaxQuant (version 2.0.3.0)64 and the Andromeda search engine.[131] Peptides were searched against a human UniProt FASTA file containing 20,408 entries (21.04.2021) and a default contaminants database. Default parameters were used if not mentioned otherwise. N-terminal acetylation and methionine oxidation were set as variable modifications, and cysteine carbamidomethylation was set as a fixed modification. A minimum peptide length of 6 amino acids was required. The false discovery rate (FDR) was set to 0.01 for the protein and peptide levels, determined by searching against a reverse sequence database. Enzyme specificity was set as C-terminal to arginine and lysine with a maximum of two missed cleavages. Peptides were identified with an initial precursor mass deviation of up to 10 ppm and a fragment mass deviation of 0.5 Da. The "Match between runs" algorithm in MaxQuant[132] was performed between all samples to increase the peptide identification rate. Proteins and peptides matching the reverse database were discarded. For label-free protein quantitation (LFQ), a minimum ratio count of 2 was required.[133] To identify the aptamer epitope on the N protein, the log2 fold-change and a p-value for each peptide were computed from the label-free quantification intensity between eluted fractions obtained from the tNSP3 aptamer and the scrambled DNA. Peptides significantly more abundant ($p < 0.05$) in the fraction eluted

from the aptamer were considered potential epitope peptides. The sequence position of individual peptides was visualized with R and the pOmics and ggplot2 packages.

4.2.4 Molecular dynamic simulation of peptide-aptamer interaction

The secondary structures of the aptamers were predicted based on their sequences using the MFold web server.[134] Protein folding was performed using I-TASSER server.[135] A model with the highest C-score was chosen for further refinement by MD simulations to improve structure quality. Aptamer-protein complexes were obtained by HDOCK molecular docking web server.[136] Molecular dynamic simulations of the aptamers, N protein, and aptamer-protein complexes were conducted the same way as previously described[119] by using GROMACS 2019.8 software[137] with the Amber14sb force field[138] and the TIP3P model for water.[139] The negative charge of the complexes was neutralized with Na^+ ions. Then, Na^+ and Cl^- ions were added to the system at the concentration of 0.15 M. After MD simulations, clusters of structures and their centers were computed using the quality threshold algorithm[140] with the VMD program.[141]

4.3 Results

4.3.1 Identification of nucleocapsid binding epitopes-targeting aptamer by nLC-MS/MS

To identify the binding sites of tNSP3 aptamer on the SARS-CoV-2 nucleoprotein, restricted proteolytic digestion of the aptamer-protein complex was performed, and the streptavidin magnetic beads were used to pull down peptides bound to the biotinylated aptamer, and non-binding peptides were washed out (Figure 4.1). The resulting binding peptides were eluted and subjected to nLC-MS/MS analysis. The intensities of the eluted

peptides from the tNSP3 aptamer were compared to a scrambled-DNA control experiment. Abundance differences of individual peptides were compared (Figure 4.2A) and visualized along the protein sequence (Figure 4.2B). The eluted peptide at a location of 376 – 385 (denoted position in SARS-CoV-2 nucleoprotein, UniProt: P0DTC9) revealed the highest abundance, and this peptide with the amino acid sequence of ADETQALPQR (named AA10) was identified as a possible binding epitope. Apart from the AA10 peptide, another peptide appeared at the location of 69 – 89 with the amino acid sequence of GQGVPINTNSSPDDQIGYYRR (named AA21); however, its abundance difference showed a 2-fold lower abundance compared to the AA10 peptide.

A B

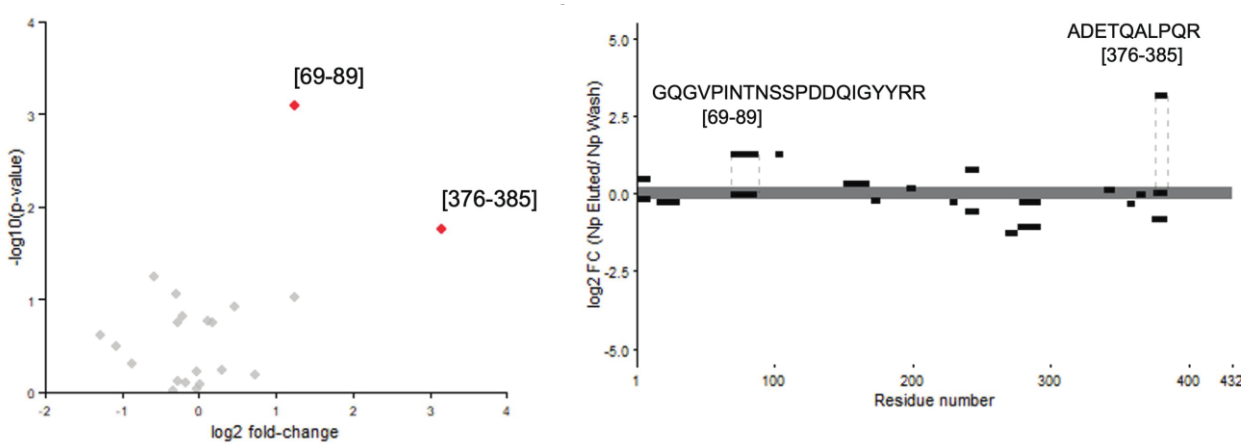


Figure 4.2 Identification of epitope peptides binding to tNSP3 aptamer by nLC-MS/MS. (A) Volcano plot showing fold-change (x-axis) and statistical significance (y-axis) of individual peptides based on label-free quantification (LFQ) intensity. Red points indicate peptides-of-interest appearing in the eluted peptides from tNSP3 aptamer in insignificantly higher abundance. (B) Peptide plots of log2 fold-change (FC) differences

of peptides visualized along the SARS-CoV-2 N sequence. The peptides that are significantly more abundant in the fractions eluted from tNSP3 aptamer than eluted from Sc64 were identified by a paired t-test and annotated by dotted lines.

4.3.2 Molecular modeling of the aptamer-protein binding interaction

The primary sequence of tNSP3 aptamer was used to predict its secondary structure by the mFold web server (Figure 4.3A).[142] The corresponding tertiary structure of the tNSP3 was modeled using SimRNA[143] and VMD[144] programs. After that, 200-ns molecular dynamics (MD) simulations were performed using GROMACS package[137] to obtain the spatial structure of the tNSP3 aptamer in the solution (Figure 4.3B). Owing to the high binding affinity of the tNSP3 aptamer, molecular docking for this aptamer was performed to predict the spatial structure of the protein-ligand complex and to reveal the critical nucleotides providing the binding. There is no known experimental structure of the whole N protein to date. In this study, the N protein tertiary structure was predicted by modeling using Iterative Threading ASSEmbly Refinement (I-TASSER) service.[135] The model with the best score referred to a reported modeling template[144] was selected and subjected to 200-ns MD simulations followed by cluster analysis of MD trajectories. The obtained equilibrated structure of the N protein (Figure 4.3C) was used for molecular docking. Considering AA10 and AA21 peptides as the binding sites on the N protein, molecular docking was performed using HDOCK web server.[136] As a result, the HDOCK program yielded ten different aptamer-protein complexes for both AA10 and AA21 binding sites. The top-scoring binding poses were taken to study aptamer-protein interaction. The models of these binding sites were further refined using MD simulations to obtain the best structures in the solution condition (Figures 4.3D and 4.3E). Both considered complexes

were stable during 200-ns long MD simulations and showed no dissociation. Two clusters with a large population in MD simulation were chosen for a more detailed evaluation of residue-nucleotide interaction for each complex. Detailed information on the hydrogen bonds is given in Table S5, and both complexes did not reveal a significant difference in the total number of hydrogen bonds. Different nucleotides are involved in the binding with AA10 and AA21 epitopes. For the AA10 binding epitope, the binding was primarily driven by G15, T17, C18, and C19 nucleotides, whereas the binding of the AA21 epitope was mainly due to T32, A22, and T23 nucleotides. All these nucleotides are located in the loop of the tNSP3 aptamer.

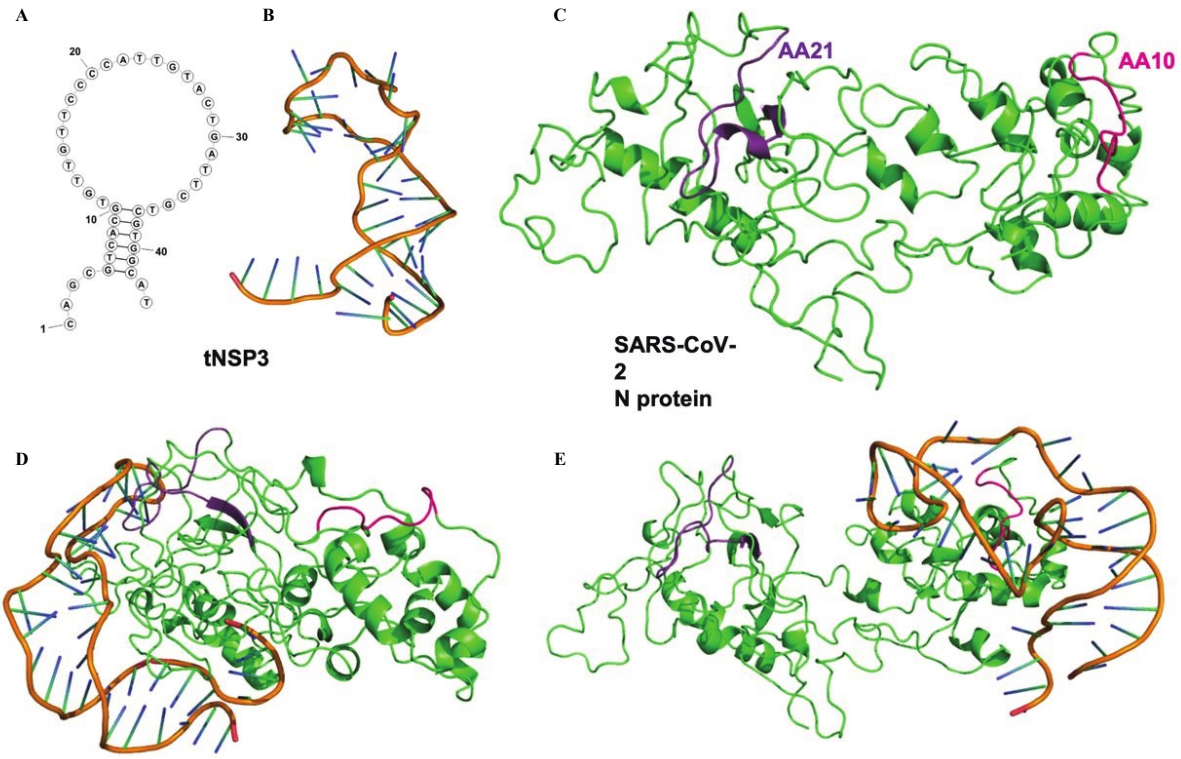


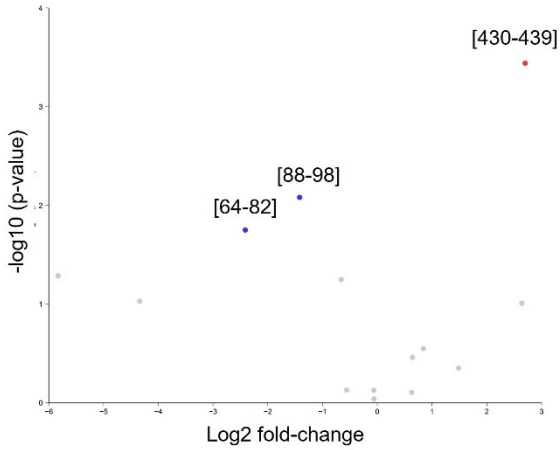
Figure 4.3 Structure of tNSP3-N protein complexes. (A) Secondary and (B) tertiary structures of tNSP3 aptamer. (C) The tertiary structure of SARS-CoV-2 nucleocapsid

protein. (D-E) Complexes of tNSP3 binding to AA21 and AA10 motifs on the N protein. The AA21 and AA10 peptides motifs are colored in purple and red, respectively.

4.3.3 Identification of S1 subunit binding epitopes-targeting aptamer by nLC-MS/MS

To determine the peptide sequences of binding epitopes on S1 protein, the experiment was performed as same as the protocol used for identifying the binding epitopes of the N protein. The intensities of eluted peptides obtained from the S1-tSP10 aptamer were statistically compared to the negative control (Sc64) which were the peptides eluted from the unknown DNA with the p-value below 0.05. The differences of individual peptide abundance were displayed as volcano plots and visualized along the protein sequences. Figure 4.4 reveals the peptides (VGGNYNYLYR) named SA10 with the highest abundance at locations of 430 – 439 as a binding epitope of S1-tSP10 aptamer over the negative control. Hence, this peptide position could be a potential binding epitope of the S1-tSP10 aptamer on the S1 subunit of SARS-CoV-2 spike protein.

A



B

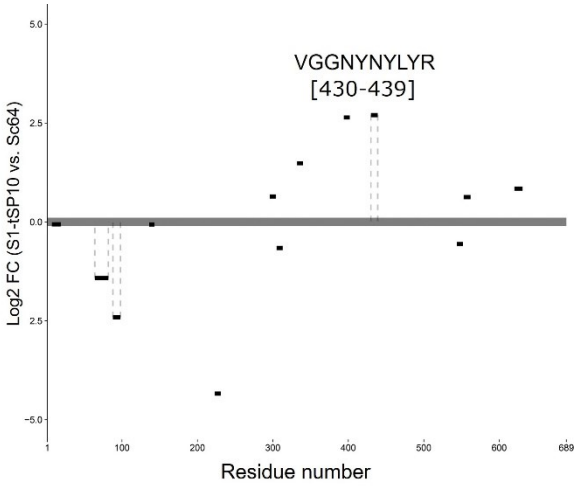


Figure 4.4 Identification of epitope peptides binding to tPS10 aptamer by nLC-MS/MS. (A) Volcano plot showing fold-change (x-axis) and statistical significance (y-axis) of individual peptides based on label-free quantification (LFQ) intensity. Red points indicate peptides-of-interest appearing in the eluted peptides from S1-tSP10 aptamer in insignificantly higher abundance than eluted from the unknown DNA (negative control). (B) Peptide plot of log2 fold-change differences of peptides visualized along the SARS-CoV-2 S1 sequence (denoted position in SARS-CoV-2 Spike protein, UniProt: P0DTC2). The peptide that is significantly more abundant in the fractions eluted from the aptamers than eluted from the unknown DNA was identified by a paired t-test and annotated by dotted lines.

4.4 Discussion

The binding epitopes on the N and S1 proteins of the tNSP3 and S1-tSP10 aptamers were identified as consisting of 10 amino acids. This binding epitope named AA10 was localized in the C-terminal domain of the nucleocapsid of SARS-CoV-2. The interaction of the AA10 binding site with the tNSP3 was modeled by molecular simulation. Interestingly, there are no identical nucleotide-amino acid pairs between tNSP3 to AA10 and to AA21 in the aptamer-protein complexes assuming that an individual nucleotide may preferably bind to the specific amino acids on the binding sites. Also, the tertiary structures of the N protein (Figure 4.3C) shows the interaction of the tNSP3 aptamer to the AA10 epitope forming the random coil configuration (magenta color) which is resistant to the heat denaturation.[145] This supports the results obtained from the study of heat inactivation effect on the binding affinity between the tNSP3 aptamer and the heated N protein. The study exhibited the heat resistance of the random coil conformation after

being heated at 65℃ and 80℃ which could still maintain the 80% binding of the protein to the aptamer on the BLI biosensor. Even though the tNSP3 aptamer showed partial binding to the AA21 peptide localized in the RNA-binding domain, this might be the non-specific interaction mainly through electrostatic interactions between positive residues on the N protein and a phosphate backbone of the DNA aptamer.

Owing to the future propose of incorporating the S1-binding aptamer for downstream applications, the binding epitope (SA10) on the S1 protein targeted by the S1-tSP10 aptamer was also identified as the amino acid position at 430 – 439 which was published as a part of RBD located in the S1 subunit which binds to the cell receptor ACE2.[146,147] This could emphasize the potential for applying this aptamer onto not only the biosensing devices, but also for therapeutic purposes.

Speaking of the mutation in amino acid sequences occurred among SARS-CoV-2 wild type and its variants, the study of MD simulation has also claimed that the mutation of the amino acid sequences in the variants does not significantly affect the binding affinity of the tNSP3 aptamer to the N protein of the variants of concerns such as Delta and Omicron. This statement can be supported by the molecular docking of the tNSP3 aptamer interacting with the AA10 peptide sequences on the N protein through H-bonds between a glutamine (Q385) and G15 and C18 as well as an arginine (R385) and T17 and C18 on the loop of the tNSP3 aptamer. Along with the summary of amino acid changes in the N and S1 proteins reported by Quaglia et al. [148], the amino acid region at the position of 376-385 on the N protein of SARS-CoV-2 wild type has only one amino acid at the position of 377 changed from aspartic acid (D) to tyrosine (Y) in the Delta variant, but none of amino acid at this position is changed in the Omicron variant. This means that the two amino acids of

interest (Gln384 and Arg385) on the identified epitope (AA10) on the N protein are not mutated among wild-type and its variants. Similarly, the amino acid sequence of the identified epitope (AA11) on the S1 protein at aa430 – aa439 was also fully conserved among the wild-type and its variants as shown in the table below. These results have contributed to the reason why the K_Ds of the tNSP3 and S1-tSP10 aptamers respectively binding to the N and S1 proteins of the Omicron variant were determined within the similar range as the K_D of the aptamers binding to the proteins of wild-type. Therefore, the mutations occurred in SARS-CoV-2 proteins, especially the N and S1 protein, do not significantly affect the binding affinity and the specificity of tNSP3 and S1-tSP10 aptamers to their individual target proteins of both wild type and its variants.

The peptide sequences of the binding epitopes (AA10 and SA10) on the N and S1 proteins of SARS-CoV-2 have only one and two amino acids, respectively, that are the same as the amino acid of MERS-CoV (Figure S7 and Figure S8). This can explain the specific interaction of the tNSP3 and S1-tSP10 aptamers with SARS-CoV-2 over MERS-CoV due to the dissimilarity of the amino acid sequences of the N and S1 proteins.

In this study of aptamer-binding epitope identification, there are some limitations in terms of different binding of the aptamers to their target proteins and to the digested peptides. After trypsin digestion, the conformational structures of specific peptide motifs in the protein could change. In addition to the main interaction of the aptamer with the specific binding epitopes, the aptamer can also interact with other parts of the protein. Moreover, the loss of potential binding epitopes that contain lysine and/or arginine can occur during proteolytic digestion.[149]

Chapter 5

Thesis Conclusions and Future Directions of Research

5.1 Thesis conclusions

This thesis mainly focuses on generating ssDNA aptamers to capture SARS-CoV-2 proteins especially nucleocapsid and S1 subunit of spike protein that could be utilized to develop a label-free biosensor for COVID-19 diagnosis. The rationale of this project was proposed that the enriched DNA aptamer with high affinity in the nanomolar range binding to the targets can be a detecting agent with the ease of functional modification in order to be implemented onto several types of the biosensing surface platform as well as can be alternatives to antibodies for therapeutic applications.

In Chapter 2, the novel DNA aptamers binding to the SARS-CoV-2 nucleocapsid and S1 subunit of spike protein were successfully selected by using asymmetric-ePCR SELEX (Figure 2.2) and applying FASTAptamer to extract the DNA sequences with the most enrichment over the abundance (Figure 2.3). The developed method proved that using asymmetric PCR with the optimized conditions including primer concentrations, FW/RW primer ratio and PCR cycles would be amendable to produce the aptamers with high affinity while less processing time compared to using conventionally symmetric PCR (Figure 2.4). With the NGS data analysis, there were sixteen (Figure 2.5) and fourteen (Figure 2.6) aptamer candidates targeting the N and S1 proteins, respectively, that could be carried over through the criteria for choosing aptamer sequences.

In Chapter 3, The target-binding affinity of the enriched aptamers was screened by using BLI technique (Figure 3.3). Among those sixteen sequences, the most enriched

sequences (NSP1, NSP2 and NSP3) showed higher binding to the N protein over the controls (Figure 3.4). The K_Ds of these top three aptamers with the varied concentrations of the N protein were below 6 nM (Figure 3.5, Table 3.1 and Table 3.2) along with the assessment of their selectivity toward the N protein of SARS-CoV-2 (Figure 3.6). With the lowest K_D (< 4nM) and high specificity of the tNSP3 aptamer to the N protein, the cross-reactivity to SARS-CoV-2 S1 was negligible; however, the binding of this aptamer to the N protein of MERS was observed with a 50% reduction. Owing to the rapid spread of variants of SARS-CoV-2 during 2021 – 2022, the ability of tNSP3 aptamer was evaluated to detect the N protein of Delta and Omicron variants (Figure 3.7A-B). The effect of heat-inactivation of the protein on the binding affinity of the aptamer was additionally studied and the results turned out that the tNSP3 aptamer could maintain its ability to detect the heat-inactivated protein of variants on BLI (Figure 3.7C-D) as well as the detection of the target N protein in human saliva with the low LOD (~4.5 nM) has highlighted the potential of the aptamer-based BLI detection (Figure 3.8).

During the early pandemic, spike protein was a main antigen that was fascinating in the aptamer study. Particularly, the S1 subunit consisting of the RBD domain plays an important role in the part of viral infection into the host cells. Thus, the production of the DNA aptamers targeting the S1 protein was encouraged to complete the development of viral protein detection. After affinity screening of the enriched aptamers targeting the S1 protein on BLI, the target binding of S1-tSP4, S1-tSp10 and S1-tSP11 could be distinguishable from other eleven aptamer sequences and their K_Ds were below 60 nM (Figure 3.9 and Table 3.4). Unexpectedly, the S1-tSP10 aptamer provided the lowest K_D among others, so its selectivity was also investigated. As a result, the S1-tSP10 exhibited

the specificity toward the S1 protein of SARS-CoV-2 with 70% reduction binding to the S1 subunit of MERS coronavirus (Figure 3.10). Moreover, this aptamer was further implemented onto the BLI biosensor to detect the S1 protein of the Omicron variant undergoing heat-inactivation; consequently, heating the protein did not fall off the association of the S1-tSP10 to its target (Figure 3.11). Last but not least, the performance of S1-tSP10 was further evinced by its ability to detect the S1 protein of the Omicron variant in human saliva with a LOD below 20 nM (Figure 3.12).

Apart from the development of aptamer-based BLI for COVID-19 detection, the use of mass spectrometry could elucidate the inside details of molecular interaction between DNA and protein along with the confirmation obtained from the MD simulations described in Chapter 4. The key advantage of mass spectrometry in this study is to reveal the amino acid sequences of the peptide-binding epitopes targeted by the aptamers (Figure 4.1). The tNSP3 aptamer showed the specific binding to the peptide sequence located in CTD of the nucleocapsid protein (Figure 4.2), and this phenomenon was simulated under the molecular dynamics showing the evidence of highly specific interaction of tNSP3 with the CTD of the N protein (Figure 4.3). Interestingly, the S1-tSP10 aptamer targeting the S1 protein bound the receptor-binding domain on the S1 subunit of spike protein (Figure 4.4). These could be pieces of fundamental knowledge for biomedical applications in therapeutics and drug development in the future.

5.2 Future directions of the current research

Despite the aptamer-based labeled free detection using BLI, the selected aptamers including tNSp3 and S1-tSP10 have been further applied to other types of biosensors and still undergoing the preliminary studies described below (Figure 5.1).

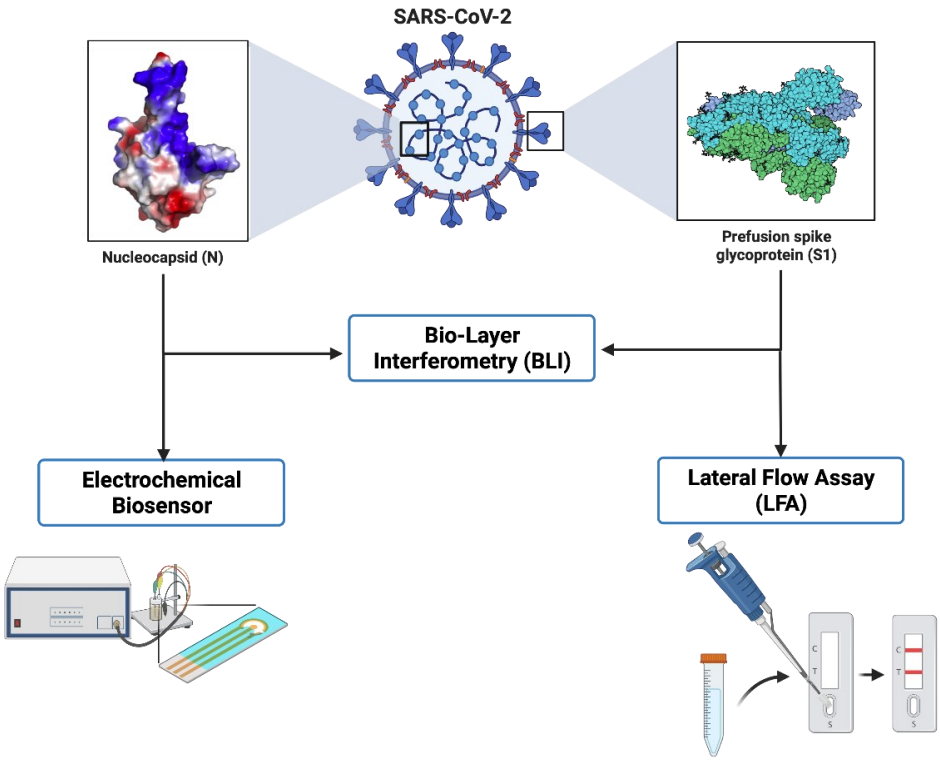


Figure 5.1 Applications of aptamer-based biosensors for SARS-CoV-2 protein detection. The scheme shows three types of biosensors using an aptamer as a capturing agent on the different sensing materials. The aptamer-based BLI technique can be utilized for detecting both the N and S1 proteins. The aptamer targeting the N protein will be implemented onto the electrochemical biosensors as well as the S1-binding aptamer will be applied for the LFA-based assay.

5.2.1 Development of electrode microarray platforms using aptamer-based electrochemical biosensor to detect SARS-CoV-2 proteins

With the ease of chemical modifications of the aptamers, aptamer-based biosensors can be used in a wide range of applications such as drug development, environmental monitoring and in vivo monitoring. By incorporating aptamer-based sensors into an electrode microarray platform, it becomes possible to simultaneously detect several

analytes at the same time and in a localized manner.[149] The previous study has demonstrated the use of three aptamers for monitoring biological neurotransmitters including adenosine triphosphate, dopamine and serotonin in blood samples.[150] The aptamer was modified with thiol linked (-SH) at the 5' end to adsorb via self-assembly monolayer (SAM) onto a gold electrode surface, and methylene blue, a redox-active reagent was linked at the 3' end of the aptamer to generate the redox current while detecting the target analytes. This has encouraged the idea of integrating the tNSP3 and S1-tSP10 conjugated with thiol linker and methylene blue onto the gold electrode to monitor the contamination of SARS-CoV-2 N and S1 proteins in clinical samples and/or wastewater. Currently, the preliminary study of this project has been started with the use of thiolated tNSP3 aptamer linked with methylene blue to optimize the SAM condition and concentration of the aptamer and the proteins as well as the incubation times (Figure 5.2).

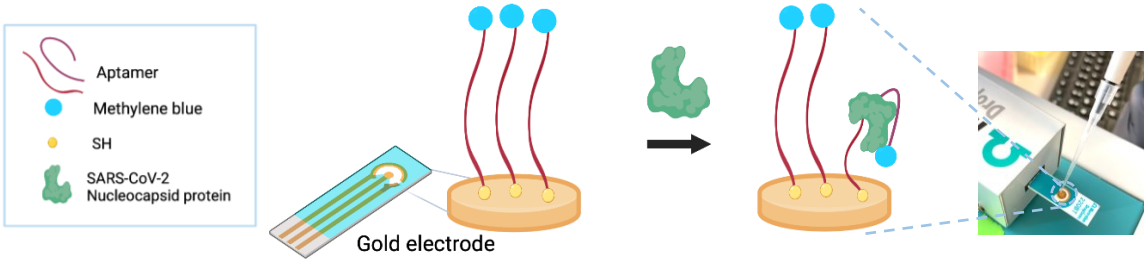


Figure 5.2 Aptamer-modified electrode microarray for sensing the presence of the N protein of SARS-CoV-2. The 5'ends of thiolated aptamers are attached to the gold electrode and their 3'ends are linked with methylene blue for generating the redox current when the conformation of the aptamer is changed while detecting the N protein.

5.2.2 Aptamer-based Lateral Flow Assay for COVID-19 detection targeting spike protein

Among various platforms for SARS-CoV-2 detection, lateral flow assays gain more interest with the advantage of providing point-of-care results rapidly from the patient samples without the requirement of highly trained personnel. Pun and his group developed an aptamer-based lateral flow assay (ALFA) incorporated with the use of an anti-RBD antibody[151], and later another LFA assay was generated without using an antibody.[152] In the aspect of enhancing the sensitivity and accuracy of ALFA, our selected aptamers targeting the S1 protein of SARS-CoV-2 can be implemented to the LFA by linking the S1-tSP4 and S1-tSP11 aptamers that bind to the CTD and RBD of the S1 subunit together, then the aptamers-S1 protein complex will be sandwiched with the S1-tSP10 aptamer targeting the NTD of the S1 subunit as shown the scheme in Figure 5.3.

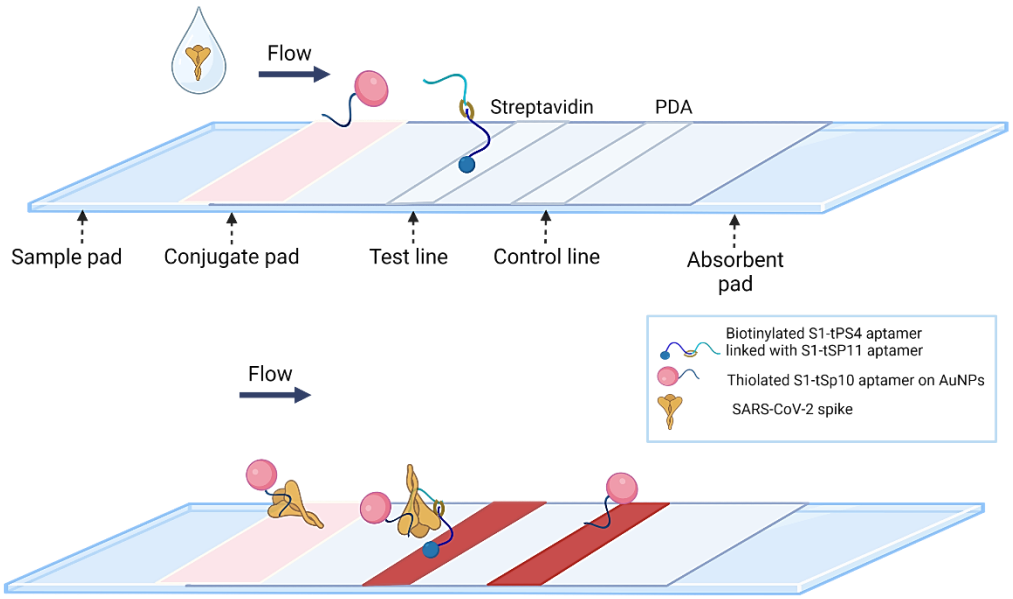


Figure 5.3 Scheme of aptamer-based LFA on SA-strips for SARS-CoV-2 spike protein detection. Once the sample solution is dropped on the sample pad, the solution flows toward the absorbent pad and the control line appears. The test line gradually reveals when the complex of dimeric aptamers binding to the S1 subunit of spike protein and the aptamer-linked AuNPs is formed on the SA strip.

5.2.3 Therapeutic and diagnostic applications of the S1- aptamers for detecting viral particles and blocking viral entry.

According to the different epitopes of the S1-binding aptamers on the S1 subunit protein, this anticipated result would suggest the development of a detecting agent by connecting three aptamers to form a trimeric structure with the integration of these three aptamers (S1-tSP4, S1-tSP10 and S1-tSP11) with DNA spacers which can be expected to potentially enhance the binding signal and sensitivity of aptamer-based biosensors to detect viral particles. To support this future study, the use of a homotrimeric aptamer composed of a symmetrical structure of three monomers of an aptamer was reported to target the spike protein.[153] However, the structure of the S protein of mature SARS-CoV-2 consists of S1 and S2 subunit which the receptor-binding motif located in the S1 subunit interacts with the host cell receptor, ACE2.[154] This has also brought to the attention of utilizing the heterotrimeric aptamer as a blocking agent to inhibit the interaction of the RBD of spike protein with ACE2, which can prevent viral entry into the host cells. The preliminary study will be tested on the pseudoviral particles transfected and replicated in HEK293 cells, and the assessment of aptamer-blocking viral entry will be examined on the ACE2-expressing Huh7 cells.

Appendix

Table S1. Conditions of asymmetric ePCR-SELEX. The stringency of aptamer selection by increasing the incubation time and the number of washes during the 1st to 6th rounds of selection.

Selection round	Incubation time (min)	No. of wash
1st	60	3
2nd	45	3
3rd	45	4
4th	30	5
5th	30	6
6th	25	7

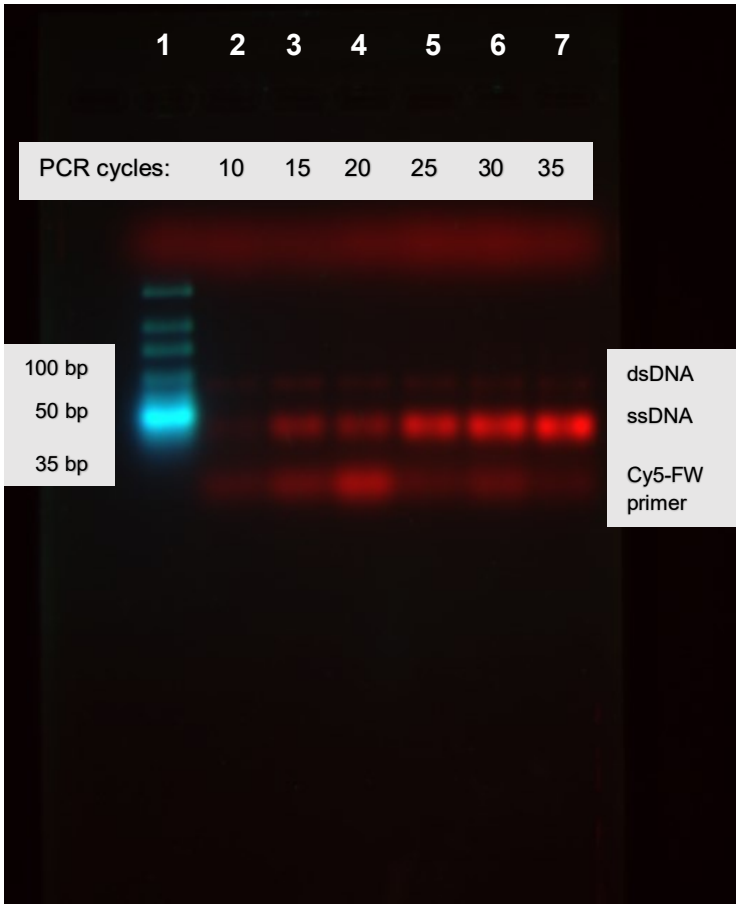


Figure S1. Gel image representing PCR products obtained from asymmetric-ePCR. Wells No. 1 - 7 (left to right) were loaded with an UltraLow range DNA ladder and the ePCR samples amplified at 10, 15, 20, 25, 30 and 35 cycles, respectively. The titration of PCR cycles yielded an optical thermal cycle for amplifying the entire eluted DNA. The PCR samples were run on 2% agarose gel at 110 V for 60 min. Two bands of each PCR product were observed on the gel for dsDNA and ssDNA with lengths of 80 bp and 80 nt, respectively, as labeled.

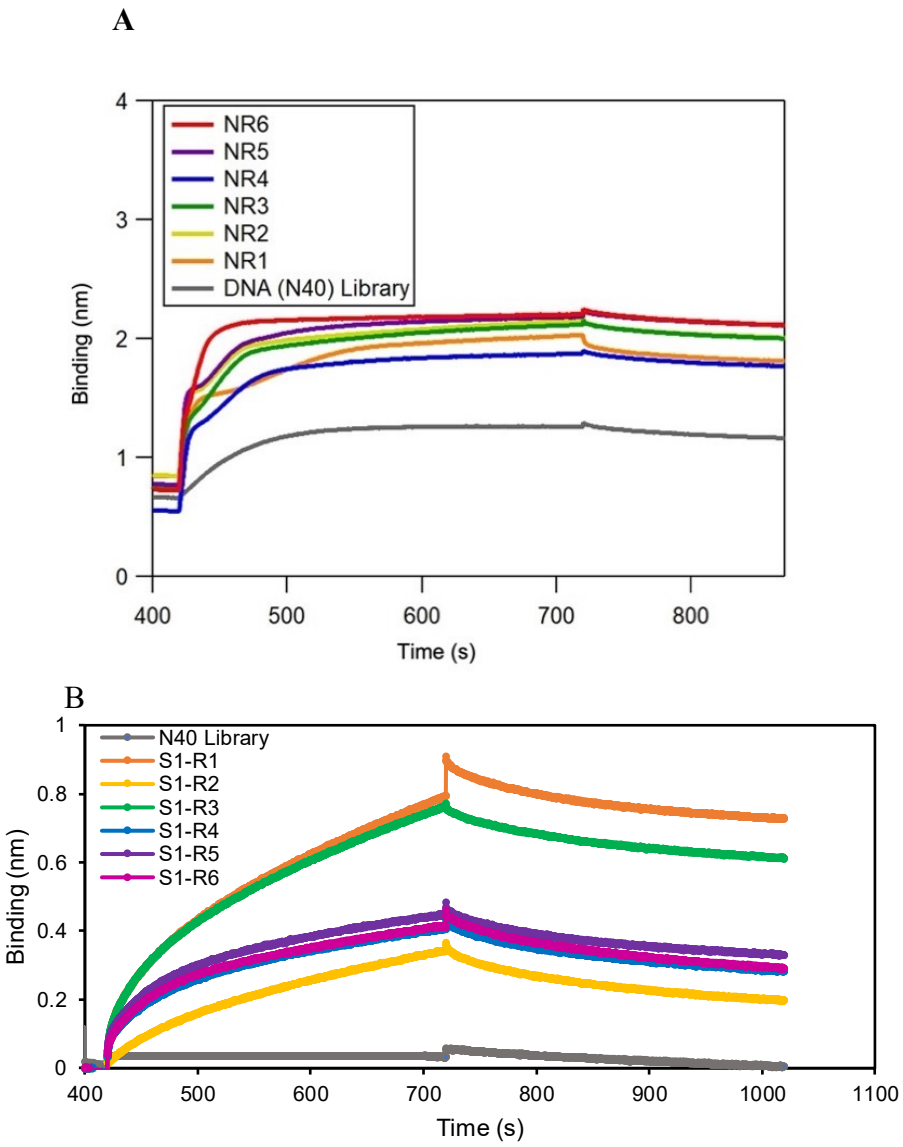


Figure S2. BLI affinity tests. BLI sensorgrams showing the association and dissociation curves of the six enriched pools obtained from the aptamer selection targeting (A) the N protein and (B) the S1 protein of SARS-CoV-2.

Table S2. The truncated aptamer sequences and their binding parameters interacting with His-tagged N protein immobilized on the Ni-NTA biosensors.

Aptamer	K_D (nM)	k_a (1/M·s)	k_d (1/s)
tNSP5	13 ± 2.6	8.3 x 10^5	1.1 x 10^{-2}
tNSP9	22 ± 1.8	4.8 x 10^5	1.1 x 10^{-2}
tNSP10	8.6 ± 1.2	7.4 x 10^5	6.4 x 10^{-3}
tNSP12	6.1 ± 2.4	6.1 x 10^5	4.0 x 10^{-3}
A48	6.5 ± 1.0	7.8 x 10^5	5.1 x 10^{-3}
A58	3.1 ± 1.0	3.6 x 10^5	1.1 x 10^{-3}

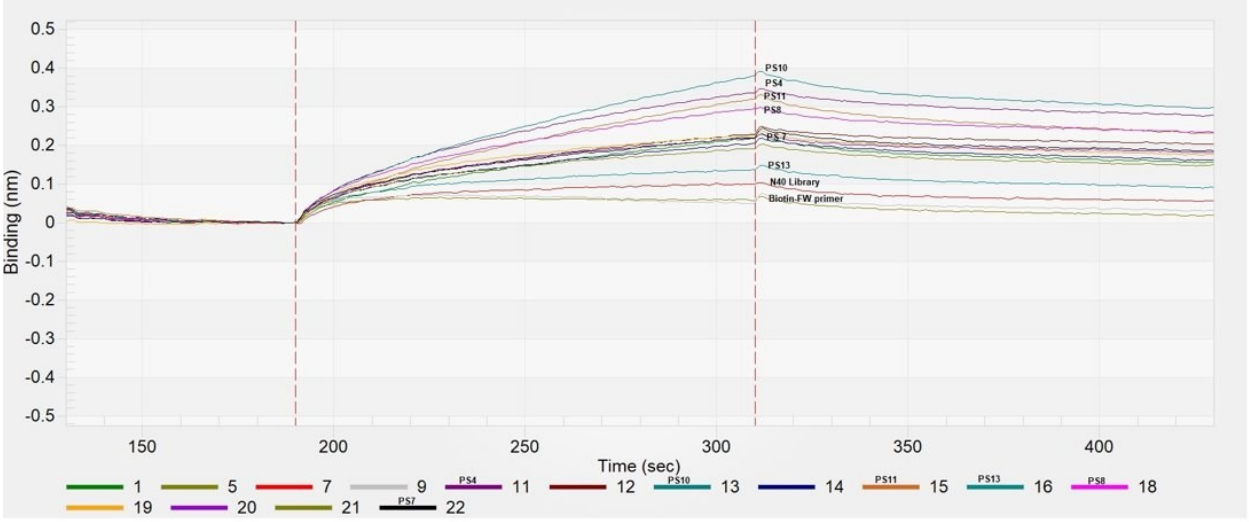


Figure S3. Aptamer-S1 affinity screening. The S1-binding enriched DNA sequences were hybridized with a biotin-labeled FW primer and immobilized on the streptavidin biosensors. BLI sensorgrams display the binding shift of the association curve representing the interaction of an individual aptamer with the S1 protein was compared among fourteen sequences.

.

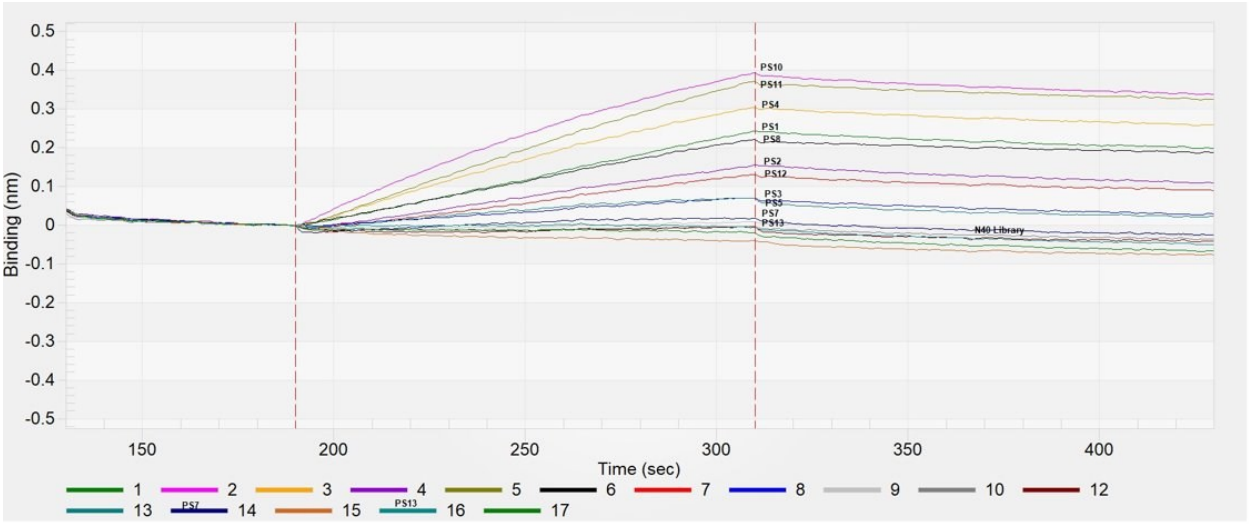


Figure S4. Aptamer-S1 affinity screening. The S1-binding enriched DNA sequences were hybridized with a biotin-labeled RW primer and immobilized on the streptavidin biosensors. BLI sensorgrams display the binding shift of the association curve representing the interaction of an individual aptamer with the S1 protein compared among fourteen sequences.

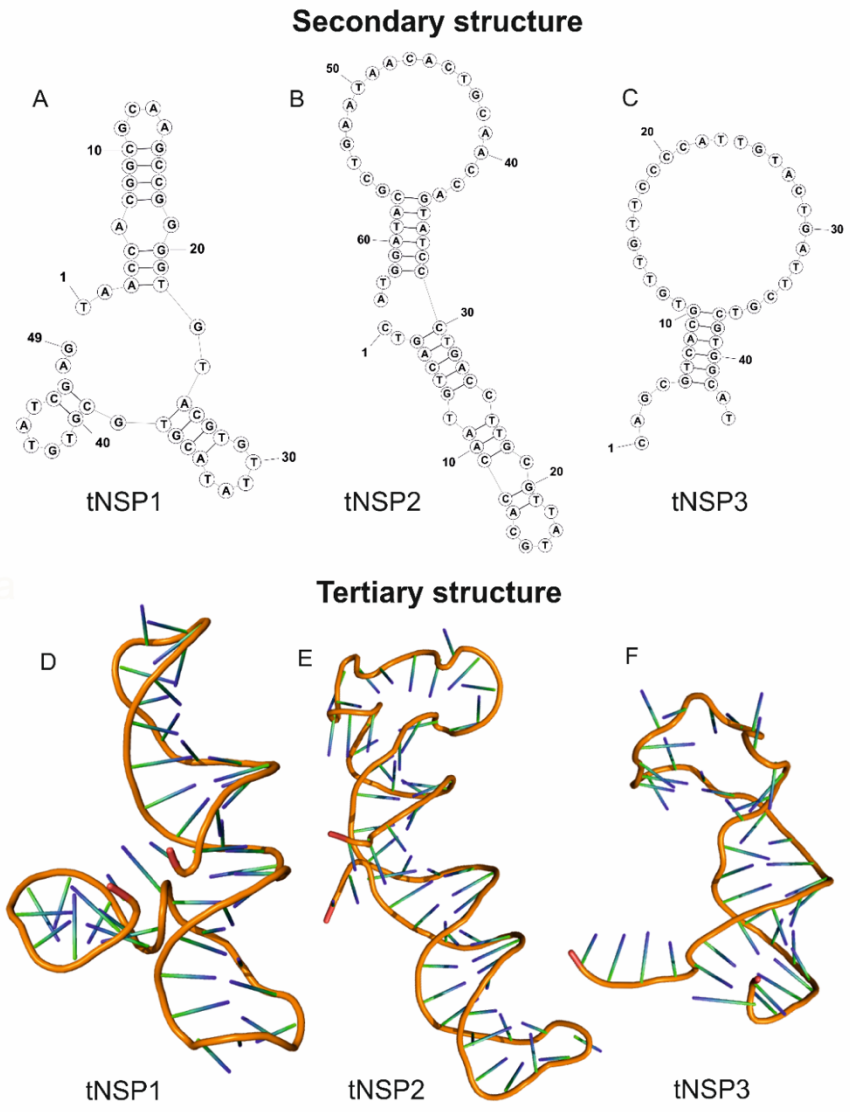


Figure S5. The structural prediction of truncated aptamers. Secondary structures (A-C) of tNSP1, tNSP2, and tNSP3 aptamers predicted by the mFold web server.[68] Tertiary structures (D-E) obtained from MD simulations (200 ns).

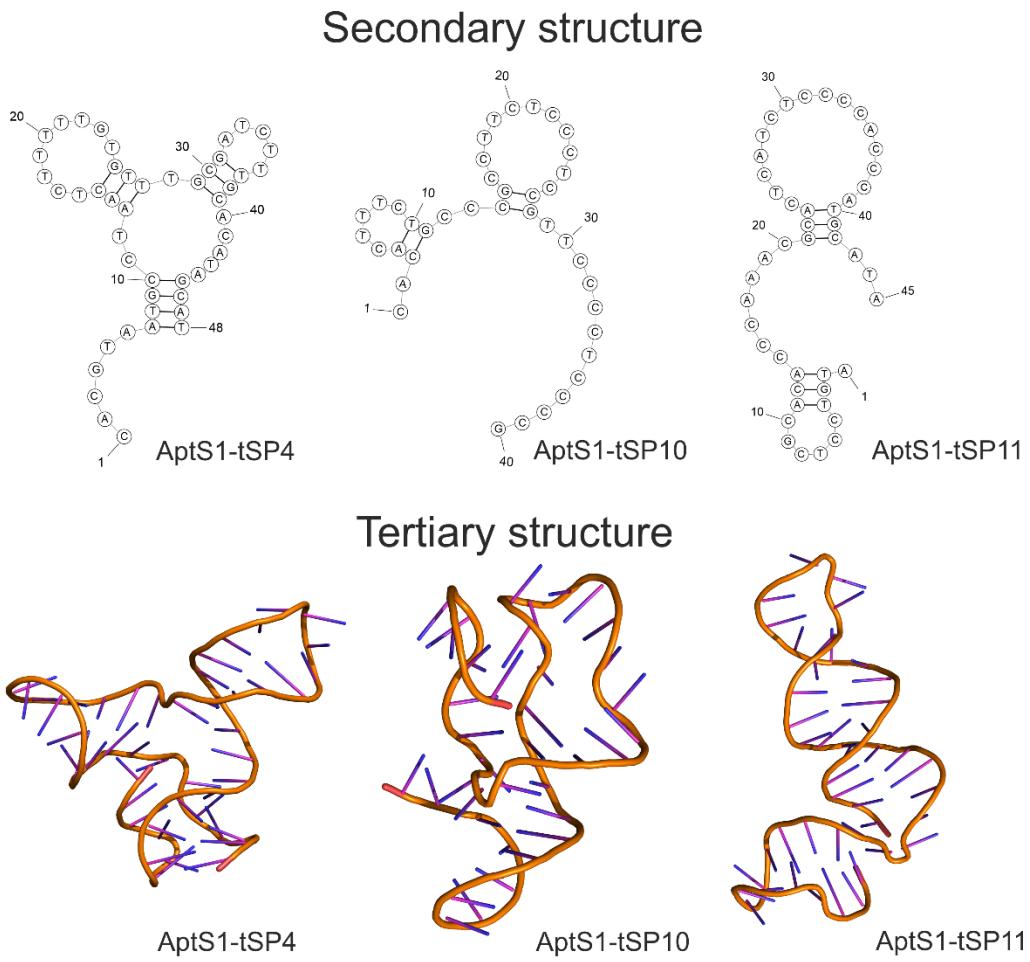


Figure S6. The structural prediction of truncated aptamers targeting S1 subunit. Secondary structures (A-C) of AptS1-tSP4, AptS1-tSP10, and AptS1-tSP11 aptamers predicted by the mFold web server.[68] Tertiary structures (D-E) obtained from MD simulations (200 ns).

Figure S7. Multiple sequence alignment of SARS-CoV-2 vs MERS-CoV Nucleocapsid. The amino acid sequence of the N protein of SARS-CoV-2 was aligned and compared with the N protein of MERS-CoV using Clustal Omega web server.

Figure S8. Multiple sequence alignment of SARS-CoV-2 vs MERS-CoV S1 subunit of spike. The amino acid sequences of the S1 protein of SARS-CoV-2 were aligned and compared with the S1 protein of MERS-CoV using Clustal Omega web server.

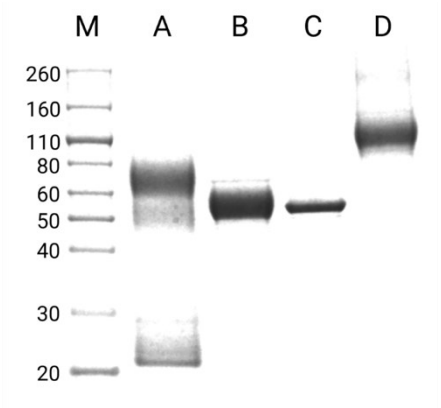


Figure S9. SDS-PAGE gel image with four proteins. M – molecular marker; A – nucleocapsid of MERS-CoV; B – nucleocapsid of SARS-CoV-2 obtained from ACROBiosystems; C – nucleocapsid of SARS-CoV-2 expressed in BL-21 *E. coli*. D. S1 subunit of the spike protein of SARS-CoV-2 obtained from ACROBiosystems.

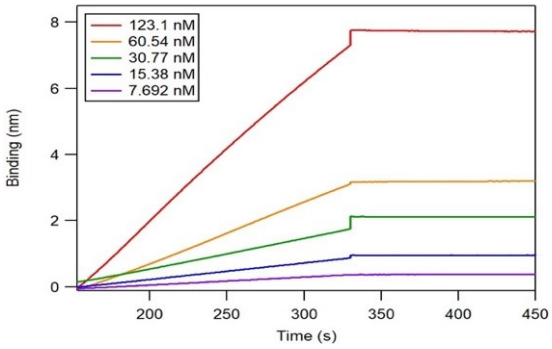


Figure S10. BLI sensorgrams of tNSP3 aptamer. The binding assay determines the K_D of tNSP3 aptamer binding to our in-house expressed SARS-CoV-2 nucleocapsid protein at 7.69, 15.38, 30.76, 61.52, and 123.04 nM.

Table S3. Summary of aptamer sequences targeting the N protein obtained from published literature.

Target Protein	DNA Aptamer	Sequence (5' → 3')	References
SARS-CoV-2 Nucleocapsid	A15 (58 nt)	GTTCGTGGTGTGCTGGATGTTCATGCTGGCA AAATTCCTTAGGGGCACCGTTACTT TGACACATCCAGCAGCACGA	[71].
	A48 (58 nt)	GTTCGTGGTGTGCTGGATGTCGCTTACGACA ATATTCCTTAGGGGCACCGCTACATTGACAC ATCCAGCAGCACGA	
	A58 (58 nt)	GTTCGTGGTGTGCTGGATGTCACCGGATTGTC GGACATCGGATTGTCTGAGTCATATGACACA TCCAGCAGCACGA	
	A61 (58 nt)	GTTCGTGGTGTGCTGGATGTTGACCTTTACAG ATCGGATTCTGTGGGGCGTTAAACTGACACA TCCAGCAGCACGA	
	SARS-CoV-2 NP (76nt)	ATCCAGAGTGACGCAGCAAACCCAAGCAAA CTACCTCTATACCCTTCGACCTTCATCATGGA CACGGTGGCTTAGT	[155]
	tNSP1	TAACCACGGCGCAAGCCGGGGTGTACGT GTTATACGTGCGTGTACGAG	This study
	tNSP2	CTGACTGTAACCACGTATTGCGTTCCAGT CCCTATGACCAACGTCCAATAAGTCGCAT AGGTA	
	tNSP3	CAGCGTCACGTGTTGTTCCCCATTGTACT GATTCGTCGTGGCAT	

A

```
A58             GTTCGTGGTGTGCTGGATGTCACCGGATTGTCGGAC-----ATCGGATTGTCTGAGTCAT    55
A61             GTTCGTGGTGTGCTGGATGTTGACCTTTACAGATCG-----GATTCTGTGGGGCGTTAAA    55
A15             GTTCGTGGTGTGCTGGATGTTCATGCTGGCAAAATT-----CCTTAGGGGCACCGTTACT    55
A48             GTTCGTGGTGTGCTGGATGTCGCTTACGACAATATT-----CCTTAGGGGCACCGCTACA    55
tNSP2           ------------CTGACTGTAACCACGTATTGCGTTCCAGTCCCTATGA-----------    37
tNSP1           --------------TAACCACGGCGCAAGCCGGGGT-----------GT-ACGTGTTATA    34
SARS-CoV-2      ---------ATCCAGAGTGACGCAGCAAACCCAAGCAAACTACCTCTAT-ACCCTTCGAC    50
tNSP3           ------------CAGCGTCACGTGTTG--------------------TT-CCCCATTGTA    27


A58             ATGACACATCCAGCAGCACGA-------    76
A61             CTGACACATCCAGCAGCACGA-------    76
A15             TTGACACATCCAGCAGCACGA-------    76
A48             TTGACACATCCAGCAGCACGA-------    76
tNSP2           --CCAACGTCCAATAAGTCGCATAGGTA    63
tNSP1           --CGTGCGTGTACGAG------------    48
SARS-CoV-2      CTTCATCATGGACACGGTGGCTTAGT--    76
tNSP3           CTGATTCGT------CGTGGCAT-----    44
                      * *
```

B

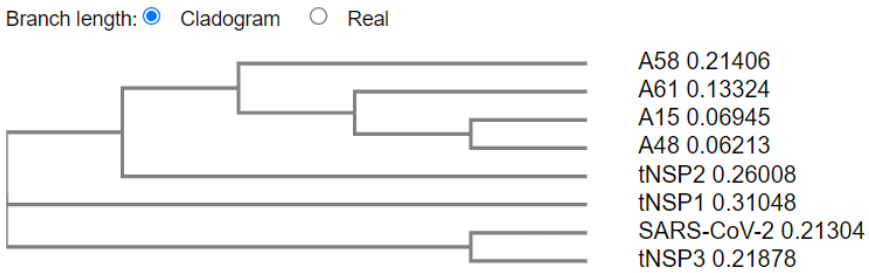


Figure S11. Comparison of DNA sequences of published aptamers targeting nucleocapsid of SARS-CoV-2. (A) Alignment of aptamer sequences that were published by other literature and our study shows the differences in the sequences of nucleotides for the N-binding aptamers and (B) Phylogenetic tree shows correlation of each aptamer.

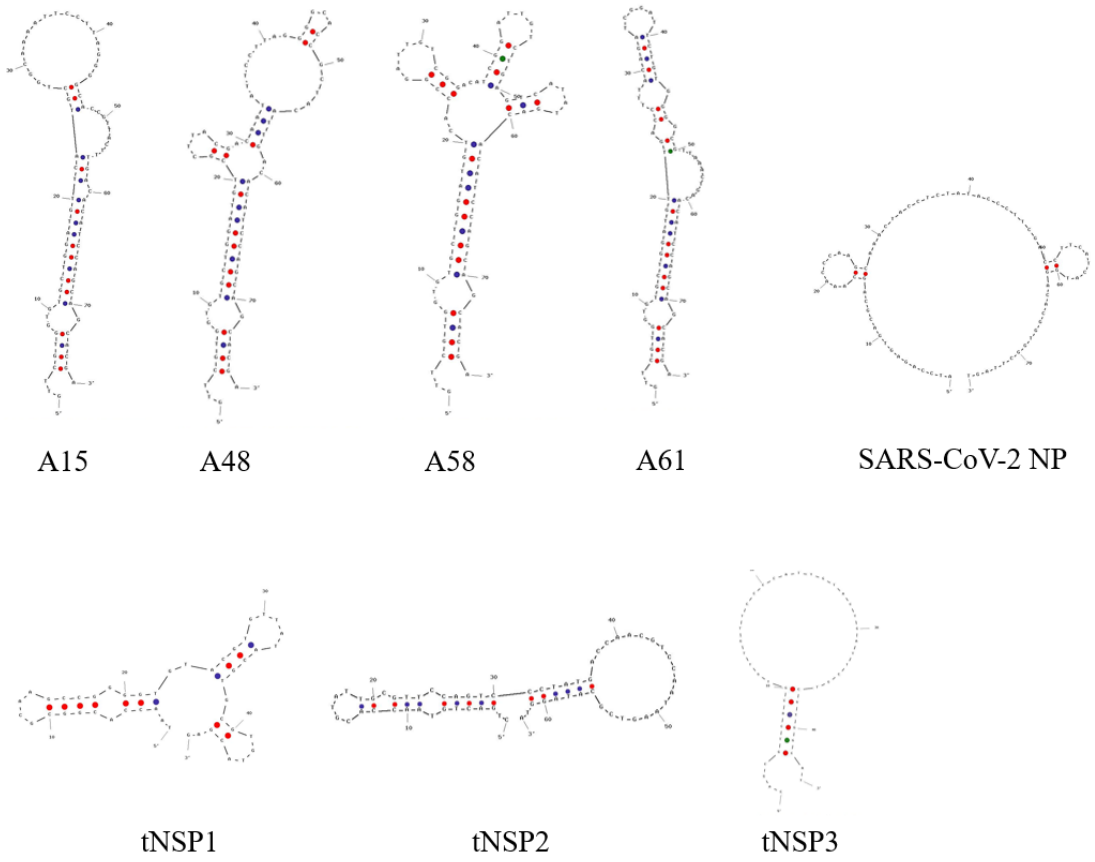


Figure S12. Comparison of secondary structures of DNA aptamers targeting nucleocapsid of SARS-CoV-2. The secondary structures were retrieved from MFold under the same condition with Na^{+} and Mg^{2+}. The A15, A48, A58, A61 and SARS-CoV-NP aptamers were published by other studies.[71, 155] Our study published the tNSP1, tNSP2 and tNSP3 aptamers.

Table S4. Summary of aptamer sequences targeting the S1 protein obtained from published literature.

Target Protein	DNA Aptamer	Sequence (5' → 3')	Ref.
SARS-CoV-2 S1 subunit	MSA1T (39 nt)	TTCCGGTTAATTTATGCTCTACCCGT CCACCTACCGGAA	[64]
	MSA5T (39 nt)	ACGGGTTTGGCGTCGGGCCTGGCG GGGGGATAGTGCGGT	
	SNAP1.50 (50 nt)	CGCGGTCATTGTGCATCCTGACTGACCCTAA GGTGCGAACATCGCCCGCG	[66]
	SNAP1.66 (66 nt)	CGCTTCTTCGCGGTCATTGTGCATCCTGACTG ACCCTAAGGTGCGAACATCGCCCGCGTAAGT CCG	
	nCoV-S1-Apt1	CCGCAGGCAGCTGCCATTAGT CTCTATCCGTGACGGTATG	[126]
	XN-268s (24 nt)	GGGGTG GGGTAGTGGTATGGAGCG	[125]
	S1-tSP4 (48 nt)	CACGTAATGCCTAACTCTTTTTGTGTTTG CGATCTTTGCACATAGCAT	This study
	S1-tSP10 (40 nt)	CACACTTTCTGCCCGCCTTCTCCCTCCGTT CCCCTCCCCG	
	S1-tSP11 (45 nt)	ATGTCCTCGCACACCCAAACGCACTCATC TCCCCACCCATGCATA	

A

```
SANAP1.50         --------CGCGGTCATTGTGCATCCTGACTGACCCTAAGGTGCGAACATCGCCCG-CG-  50
SNAP1.66          CGCTTCTTCGCGGTCATTGTGCATCCTGACTGACCCTAAGGTGCGAACATCGCCCG-CGT  59
S1-tSP11          --------------------ATGTCCTCGCACACCCAAACGCACTCATCTCCCCACCCAT  40
MSA5T             -------ACGGGTTTGGCGTCGGGCCTGGCGGGGGGATAG------TGCGGT--------  39
XN-268s           ---------GGGG-TG----------GGGTAGTGGTATGG------AGCG----------  24
S1-tSP4           ---------------CACGTAATGCCTAACTCTTTTTGTGTTTGCGATCTTTGCACATAG  45
S1-tSP10          ----------------------CACACTTTCTGCCCGCCTTCTCCCTCCGTTCCCCTCC  37
MSA1T             --------------------------TTCCGGTTAATTTATGCTCTACCCGTCCACCTAC  34
nCoV-S1-Apt1      ---------------------CCGCAGGCAGCTGCCATTAGTCTCTATCCGTGACGGTAT  39


SANAP1.50         -------       50
SNAP1.66          AAGTCCG       66
S1-tSP11          GCATA--       45
MSA5T             -------       39
XN-268s           -------       24
S1-tSP4           CAT----       48
S1-tSP10          CCG----       40
MSA1T             CGGAA--       39
nCoV-S1-Apt1      G------       40
```

B

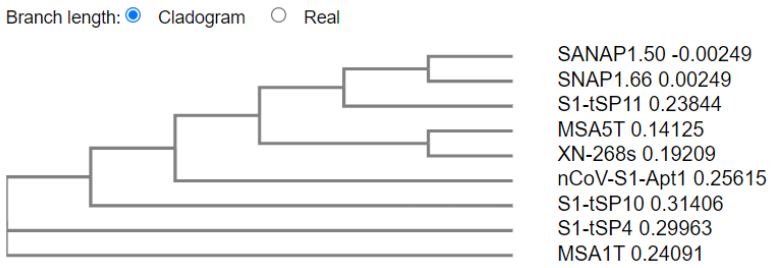


Figure S13. Comparison of DNA sequences of published aptamers targeting S1 subunit of SARS-CoV-2 spike protein. (A) Alignment of aptamer sequences that were published by other literature and our study shows the differences in the sequences of nucleotides for the S1-binding aptamers and (B) Phylogenetic tree shows correlation of each aptamer.

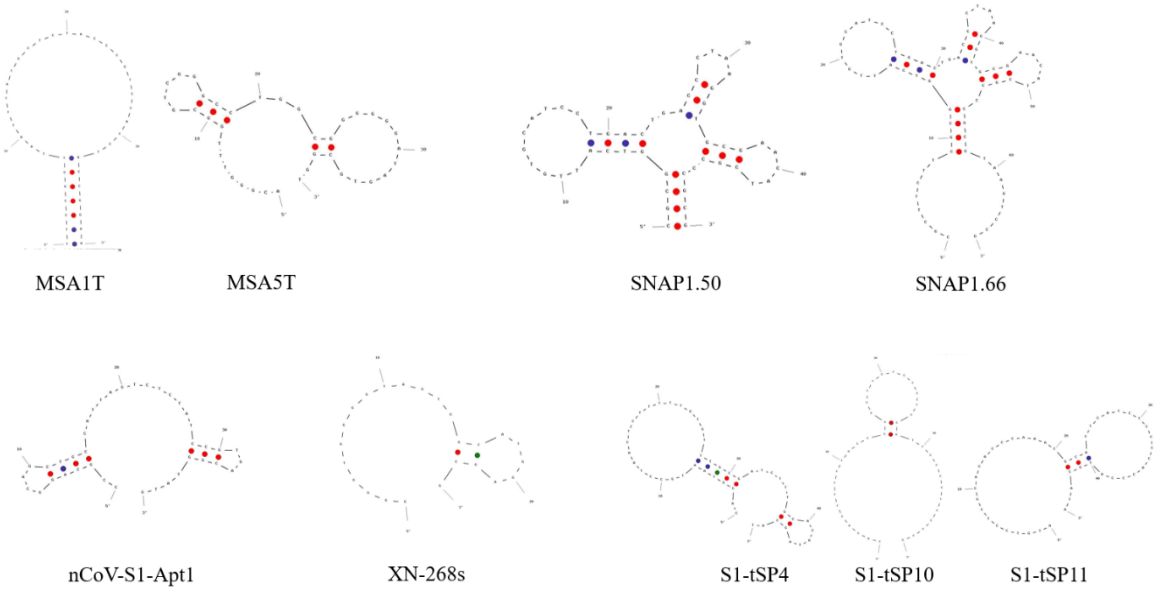


Figure S14. Comparison of secondary structures of DNA aptamers targeting S1 subunit of SARS-CoV-2 spike protein. The MSA1T, MSA5T, SNAP1.50, SNAP1.66, nCoV-S1-Apt1, XN-268s were published by other studies.[64,66,125,126] The S1-tSP4, S1-tSP10 and S1-tSP11 aptamers were reported by our study.

Table S5. Hydrogen bonds in the N protein-tNSP3 complexes. The number in parentheses indicates the conformer number. Amino acids of AA10 peptide are in bold.

N protein/NSP3 (1)		N protein/NSP3 (2)		N protein/NSP3 (1)		N protein/NSP3 (2)	
Binding site AA21		**Binding site AA21**		**Binding site AA10**		**Binding site AA10**	
Residue	Nucleotide	Residue	Nucleotide	Residue	Nucleotide	Residue	Nucleotide
GLN 9	G 35	GLN 9	G 35	LYS 338	C 4	ASN 126	T 23
ARG 10	G 35	ARG 10	G 35	LYS 361	C 37	LYS 127	T 23
ARG 14	C 34	ASN 29	C 1	GLU 367	G 30	ARG 276	T 23
ASN 29	T 6	**ASN 77**	T 32	**ARG 385**	T 17	ARG 293	C 19
ASN 29	A 2	**SER 79**	T 23	**GLN 384**	G 15	LYS 338	C 4
GLU 31	C 1	THR 14	C 21	**GLN 384**	C 18	LYS 361	C 37
GLU 31	A 2	LYS	C 19	GLN 386	T 14	GLU 367	G 30
SER 37	C 1	143	C 20	LYS 387	T 16	LYS 373	T 29
ASN 77	T 32	LYS	G 25	LYS 387	T 14	LYS 375	T 33
SER 79	A 22	143	A 27	GLN 408	C 4	**GLN 384**	G 15
SER 79	T 23	GLN	T 29	GLN 408	T 6	**GLN 384**	C 18
LYS 143	C 20	160	G 3			**GLN 384**	C 19
GLN 160	G 25	ARG				**ARG 385**	T 17
GLN 163	T 24	189				**ARG 385**	C 18
SER 193	T 29	SER				GLN 386	T 14
SER 201	T 29	193				GLN 386	T 16
ARG 203	G 30	LYS				LYS 387	T 16
ARG 203	T 29	237				GLN 408	C 4
LYS 237	G 3						
LYS 256	A 2						

Table S6. Apparent K_D determination of selected aptamers obtained from symmetric-emulsion PCR SELEX.

Aptamer	K_D (nM)	k_a (1/Ms)	k_d (1/s)
SymN1	6.00	8.04×10^5	4.83×10^{-3}
SymN3	9.84	7.09×10^5	6.98×10^{-2}
SymN5	19.0	7.83×10^5	1.49×10^{-2}
SymN6	17.9	6.71×10^5	1.20×10^{-2}
SymN8	5.99	1.02×10^6	6.10×10^{-3}
SymN9	5.62	1.10×10^6	6.19×10^{-3}
SymN12	11.6	5.32×10^5	6.18×10^{-3}
SymN14	25.6	5.62×10^5	1.44×10^{-2}

Method S1. Data analysis for retrieving aptamer sequences and secondary structures

- NGS data analysis on Command Prompt using FASTAptamer

1. Download Strawberry Perl
2. Download FASTAptamer from https://burkelab.missouri.edu/fastaptamer.html (a modular collection of scripts that can extract the information from the dataset)
3. Use Command Prompt to write the script for data analysis
4. Make sure to locate the.fastq files in the folder named FASTAptamer without any subfolders (Note: this will be easy to write the input code and output code later)
5. Once open the Command prompt, the starter script will be initially written as

Microsoft Windows [Version 10.0.19042.1348]

C:\Users\Lab>

Then, type “cd Documents\FASTAptamer” to locate where FASTAptamer is.

- How to write the script

C:\Users\Lab>cd Documents\FASTAptamer

6. Start with FASTAptamer count to convert a .fastq file to a .fasta file by typing “>perl fastaptamer_count” Followed by the input file “-i NR1_R1_001.fastq”, and output file “-o NR1_R1.fasta”

7. FASTAptamer_count will also provide the most abundant sequence of the pool.

- See how to write the script below.

C:\Users\Lab\Documents\FASTAptamer>perl fastaptamer_count -i NR1_R1_001.fastq -o NR1_R1. fasta (Click Enter)

8. Once the .fastq file was converted to a .fasta file, FASTAptamer_cluster will be used to cluster families of the sequences. Type “>perl fastaptamer_cluster” Followed by the input file “-i NR1_R1. fasta”, and output file “-o NR1_count. fasta”

- How to write the script

C:\Users\Lab\Documents\FASTAptamer>perl fastaptamer_cluster -i NR2_R1. fasta -o NR2_cluster.fasta -d 5 -f 4 (Click Enter)

9. If the abundance of the sequences is mainly focused, running the raw data (.fasta) through FASTAptamer_count and/or followed by FASTAptamer_cluster should be sufficient.

- FASTAptamer_Enrichment

10. FASTAptamer_enrich can be used to calculate fold-enrichment for an individual sequence across the populations. If you would like to enrich the sequences by observing which sequence appears in every round of selection, FASTAptamer_enrich can be used to run the fasta files.

11. Only three files (max) obtained from either FASTAptamer_Count or FASTAptamer_Cluster can be added as an input file and processed on FASTApatmer_enrich.

12. Beginning with telling the command prompt that FASTAptamer_Enrich is used for this work by typing >perl fastaptamer_enrich and designate the input (-x, -y, -z) and output (-o) files by locating the folder containing the input (.fasta) files and directing the folder where the output file will be located.

- How to write the script

C:\Users\Lab>cd Documents\FASTAptamer>perl fastaptamer_enrich -x S1_Count_fasta/S1-R1_cluster.fasta -y S1_Count_fasta/S1-R2_cluster.fasta -o S1_Count.fasta/S1_vs_S2_enrich.tsv (Click Enter)

13. However, if all the. fasta files are located in FASTAptamer folder, simply designate the input as a cluster.fasta without locating where the folder containing that input file is.

- How to write the script

C:\Users\Lab>cd Documents\FASTAptamer>perl fastaptamer_enrich -x NR1_cluster.fasta -y NR2_cluster.fasta -z NR3_cluster.fasta -o NR1_vs_NR2_ vs_NR3_enrich.tsv (Click Enter)

14. After running the FASTAptamer_enrich, the data will be retrieved in the format of .tsv which can be opened on the excel spreadsheet.

15. Once open the enriched data, filter the data by first looking at the column named "Length" to a certain nucleotide length as preferred. For example, our DNA N40 library has 80 nucleotides so the sequences with the length of 80-83 nucleotides will be considered.

16. Second, check the fold enrichment which is supposed to show the values obtained for all three pools as y/x, z/y and z/x

17. Third, the fold enrichment of ± 1.5 or 2 will be considered. [NOTICE] This criterion can be optional. If the data does not meet this condition and the fold enrichment is below 2, the data can still be considered with the condition in Step 16 by checking the fold enrichment of the three enriched pools.

- How to choose the sequences

18. First, when the sequences are conditionally retrieved from step 16 and 17, check the construction of the aptamers by looking at the position and sequences of forward and reverse primers.

19. Then, check where the random region is. The random region should be about 40 – 43 in length and located in between the forward/reverse primer binding sites.

- Alignment of the selected DNA sequences

20. To observe the similarity and differences, align all the sequences obtained from step 19 on Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) by using a tool named Multiple Sequence Alignment.

21. Choose the data set as DNA in the drop-down menu

22. Copy and paste the sequences obtained from step 19. For each sequence, label the name by starting with > followed by the name.

23. Choose the output format as ClustalW with character counts. The default setting should work.

24. Then, click submit.

25. Compare the similarity and diversity of the sequences by looking at a phylogram or a phylogenic tree. If some aptamer sequences have the value of neighboring distance as zero, basically they have the same nucleotide sequence so they will be counted as one sequence.

26. Summarize all the sequences obtained and order the aptamer clones on IDT (https://www.idtdna.com/).

- Prediction of secondary structures using RNAstructure web server

27. Check secondary structures of each selected sequence on https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html

28. Copy and paste the sequence into the input data box

29. Change the temperature to 277.15 K which is equal to 4°C as we folded the DNA in the aptamer selection. (Other parameters can be set up by default)

30. However, if the secondary structure cannot be used to consider which aptamer clones will be further checked the binding affinity and if there are not more than 15 clones, all sequences obtained from step 26 can be ordered on iDT (https://www.idtdna.com/)

References

1. Li, G., and De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat. Rev. Drug. Discov.*19(3)*, 149-150.
2. World Health Organization (2022). WHO Coronavirus (COVID-19) Dashboard. https://Covid19.Who.Int/
3. Cleri, F., Lensink, M.F., and Blossey, R. (2021). DNA Aptamers Block the Receptor Binding Domain at the Spike Protein of SARS-CoV-2. Front. Mol. Biosci. *8*,713003.
4. Ke Z., Oton J., Qu K., Cortese M., Zila V., McKeane L., Nakane T., Zivanov J., Neufeldt C.J., Cerikan B., Lu J.M., Peukes J., Xiong X., Kräusslich H.G., Scheres S.H.W., Bartenschlager R., and Briggs J.A.G. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature. *588(7838)*, 498-502.
5. Yang, H., and Rao, Z. (2021). Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat. Rev. Microbiol. *19*, 685–700.
6. Chilamakuri, R. and Agarwal, S. (2021). COVID-19: Characteristics and therapeutics. Cells. *10*, 206.
7. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. *181*, 281–292.
8. Taleghani N., and Taghipour F. (2021). Diagnosis of COVID-19 for controlling the pandemic: A review of the state-of-the-art. Biosens. Bioelectron. *174*, 112830.
9. Udugama, B., Kadhiresan, P., Kozlowski, H.N., Malekjahani, A., Osborne, M., Li, V.Y.C., Chen, H., Mubareka, S., Gubbay, J.B., and Chan, W.C.W. (2020). Diagnosing COVID-19: the disease and tools for detection. ACS Nano *14*, 3822–3835.
10. Yüce, M., Filiztekin, E., and Özkaya, K.G. (2021). COVID-19 diagnosis —a review of current methods. Biosens. Bioelectron. *172*, 112752.
11. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.
12. Smyrlaki, I., Ekman, M., Lentini, A., de Sousa, N.R., Papanicolaou, N., Vondracek, M., Aarum, J., Safari, H., Muradrasoli, S., Rothfuchs, A.G., et al. (2020). Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR. Nat. Commun. *11*, 4812.
13. Dong, L., Zhou, J., Niu, C., Wang, Q., Pan, Y., Sheng, S., Wang, X., Zhang, Y., Yang, J., Liu, M., et al. (2021). Highly accurate and sensitive diagnostic detection of SARS-CoV-2 by digital PCR. Talanta *224*, 121726.

14. Vogels, C.B.F., Brito, A.F., Wyllie, A.L., Fauver, J.R., Ott, I.M., Kalinich, C.C., Petrone, M.E., Casanovas-Massana, A., Muenker, M.C., Moore, A.J., et al. (2020). Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. Nat. Microbiol. *5*, 1299–1305.

15. U.S. Food and Drug Administration (2022). In Vitro Diagnostics EUAs - Antigen Diagnostic Tests for SARS-CoV-2. https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas-antigen-diagnostic-tests-sars-cov-2.

16. Machado, B., Hodel, K., Barbosa-Júnior, V., Soares, M., and Badaró, R. (2020). The main molecular and serological methods for diagnosing COVID-19: an overview based on the literature. Viruses *13*, 40.

17. Wang, D., He, S., Wang, X., Yan, Y., Liu, J., Wu, S., Liu, S., Lei, Y., Chen, M., Li, L., et al. (2020). Rapid lateral flow immunoassay for the fluorescence detection of SARS-CoV-2 RNA. Nat. Biomed. Eng. *4*, 1150–1158.

18. Liu, W., Liu, L., Kou, G., Zheng, Y., Ding, Y., Ni, W., Wang, Q., Tan, L., Wu, W., Tang, S., et al. (2020). Evaluation of nucleocapsid and spike protein-based enzyme-linked immunosorbent assays for detecting antibodies against SARS-CoV-2. J. Clin. Microbiol. *58*, e00461-20.

19. Chen, Z., Zhang, Z., Zhai, X., Li, Y., Lin, L., Zhao, H., Bian, L., Li, P., Yu, L., Wu, Y., et al. (2020). Rapid and sensitive detection of anti-SARS-CoV-2 IgG, using lanthanide-doped nanoparticles-based lateral flow immunoassay. Anal. Chem. *92*, 7226–7231.

20. Peeling, R.W., Heymann, D.L., Teo, Y.Y., and Garcia, P.J. (2022). Diagnostics for COVID-19: moving from pandemic response to control. Lancet. *399(10326)*, 757-768.

21. Yan, C., Cui, J., Huang, L., Du, B., Chen, L., Xuel, G., Li, S., Zhang, W., Zhao, L., Sun, Y., et al. (2020). Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop–mediated isothermal amplification assay. Clin. Microbiol. Infect. *26*, 773-779.

22. Filchakova O., Dossym D., Ilyas A., Kuanysheva T., Abdizhamil A., and Bukasov R. (2022). Review of COVID-19 testing and diagnostic methods. Talanta. *244*, 123409.

23. Hwang, Y.-C., Lu, R.-M., Su, S.-C., Chiang, P.-Y., Ko, S.-H., Ke, F.-Y., Liang, K.-H., Hsieh, T.-Y., and Wu, H.-C. (2022). Monoclonal antibodies for COVID-19 therapy and SARS-CoV-2 detection. J. Biomed. Sci. *29*, 1.

24. Yakoh, A., Pimpitak, U., Rengpipat, S., Hirankarn, N., Chailapakul, O., and Chaiyo, S. (2021). Paper-based electrochemical biosensor for diagnosing COVID-19: Detection of SARS-CoV-2 antibodies and antigen. Biosens. Bioelectron. *176*, 112912.

25. Pan, Y., Zhang, D., Yang, P., Poon, L.L.M., and Wang, Q. (2020). Viral load of SARS-CoV-2 in clinical samples. Lancet Infect. Dis. *20 (4)*, 411–412.

26. Augustine, R., Das, S., Hasan, A., Abhilash, S., Abdul Salam, S., Augustine, P., Dalvi, Y.B., Varghese, R., Primavera, R., Yassine, H.M., Thakor, A.S., and Kevadiya, B.D. (2020) Rapid antibody-based COVID-19 mass surveillance: relevance, challenges, and prospects in a pandemic and post-pandemic world. J. Clin. Med. *9(10)*,3372.

27. Benzigar, M.R., Bhattacharjee, R., Baharfar, M., and Liu, G. (2021). Current methods for diagnosis of human coronaviruses: pros and cons. Anal. Bioanal. Chem. *413(9)*, 2311-2330.

28. Cui, F., and Zhou, H.S. (2020). Diagnostic methods and potential portable biosensors for coronavirus disease 2019. Biosens. Bioelectron. *165*, 112349.

29. Hill, A., Beitelshees, M., and Pfeifer, B. A. (2021). Vaccine Delivery and Immune Response Basics. Methods Mol. Biol. *2183*, 1–8.

30. Ma, H., Zeng, W., He, H., Zhao, D., Jiang, D., Zhou, P., Cheng, L., Li, Y., Ma, X., and Jin, T. (2020). Serum IgA, IgM, and IgG responses in COVID-19. Cell. Mol. Immunol. *17(7)*, 773-775.

31. Fernandes, Q., Inchakalody, V.P., Merhi, M., Mestiri, S., Taib, N., Moustafa Abo El-Ella, D., Bedhiafi, T., Raza, A., Al-Zaidan, L., et al. (2022). Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. Ann. Med. *54(1)*, 524-540.

32. Budd, J., Miller, B.S., Weckman, N.E., Cherkaoui, D., Huang, D., Decruz, A.T., Fongwen, N., Han, G.-R., Broto, M., Estcourt, C., et al. (2023). Lateral flow test engineering and lessons learned from COVID-19. Nat. Rev. Bioeng. *1*, 13–31.

33. Parolo, C., Sena-Torralba, A., Bergua, J.F., Calucho, E., Fuentes-Chust, C., Hu, L., Rivas, L., Álvarez-Diduk. R., Nguyen, E.P., Cinti, S., et al. (2020). Tutorial: design and fabrication of nanoparticle-based lateral-flow immunoassays. Nat. Protoc. *15(12)*, 378,8-3816.

34. Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, Y., Li, S., Sun, R., Wang, Y., Hu, B., Chen, W., et al. (2020). Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. J. Med. Virol. *92(9)*, 1518-1524.

35. Luo, J., Brakel, A., Krizsan, A., Ludwig, T., Mötzing, M., Volke, D., Lakowa, N., Grünewald, T., Lehmann, C., Wolf, J., et al. (2022). Sensitive and specific serological ELISA for the detection of SARS-CoV-2 infections. Virol. J.*19(1)*, 50.

36. Krähling, V., Halwe, S., Rohde, C., Becker, D., Berghöfer, S., Dahlke, C., Eickmann, M., Ercanoglu, M.S., Gieselmann, L., et al. (2021). Development and characterization of an indirect ELISA to detect SARS-CoV-2 spike protein-specific antibodies. J. Immunol. Methods. *490*, 112958.

37. Kasetsirikul, S., Umer, M., Soda, N., Sreejith, K.R., Shiddiky, M.J.A., and Nguyen, N.T. (2020). Detection of the SARS-CoV-2 humanized antibody with paper-based ELISA. Analyst. *145(23)*, 7680-7686.

38. Chen, X.F., Zhao, X. and Yang, Z. (2022). Aptasensors for the detection of

infectious pathogens: design strategies and point-of-care testing. Microchim. Acta. *189*, 443.

39. Carter, L.J., Garner, L.V., Smoot, J.W., Li, Y., Zhou, Q., Saveson, C.J., Sasso, J.M., Gregg, A.C., Soares, D.J., Beskid, T.R., Jervey, S.R., and Liu, C. (2020). Assay Techniques and Test Development for COVID-19 Diagnosis. ACS Cent. Sci. *6(5)*, 591-605.

40. Liu, G., and Rusling, J.F. (2021). COVID-19 antibody tests and their limitations. ACS Sensors. *6*, 593–612.

41. Li, L., Xu, S., Yan, H., Li, X., Yazd, H.S., Li, X., Huang, T., Cui, C., Jiang, J., and Tan, W. (2021). Nucleic acid aptamers for molecular diagnostics and therapeutics: advances and perspectives. Angew. Chemie Int. Ed. *60*, 2221–2231.

42. Park, K.S. (2018). Nucleic acid aptamer-based methods for diagnosis of infections. Biosens. Bioelectron. *102*, 179-188.

43. Tuerk, C., and Gold, L. (1990). Systematic Evolution of Ligands by Exponential Enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*(4968), 505–510.

44. Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

45. Li, L., Xu, S., Yan, H., Li, X., Yazd, H.S., Li, X., Huang, T., Cui, C., Jiang, J., and Tan, W. (2021). Nucleic acid aptamers for molecular diagnostics and therapeutics: advances and perspectives. Angew. Chemie Int. Ed. *60*, 2221–2231.

46. Poolsup, S., and Kim, C.-Y. (2017). Therapeutic applications of synthetic nucleic acid aptamers. Curr. Opin. Biotechnol. *48*, 180–186.

48. Shum, K.T. and Tanner, J.A. (2008). Differential inhibitory activities and stabilization of DNA Aptamers against the SARS Coronavirus Helicase. ChemBioChem. *9*, 3037-3045.

49. Kim, G., Kim, J., Kim, S.M., Kato, T., Yoon, J., Noh, S., Park, E.Y., Park, C., Lee, T., and Choi, J.-W. (2022). Fabrication of MERS-nanovesicle biosensor composed of multi-functional DNA aptamer/graphene-MoS 2 nanocomposites based on electrochemical and surface-enhanced Raman spectroscopy. Sens. Actuators B Chem. *352*, 2.

50. Pfeiffer, F., Rosenthal, M., Siegl, J., Ewers, J., and Mayer, G. (2017). Customized nucleic acid libraries for enhanced aptamer selection and performance. Curr. Opin. Biotechnol. *48*, 111−118.

51. Wang, J., Gong, Q., Maheshwari, N., Eisenstein, M., Arcila, M. L., Kosik, K. S., and Soh, H. T. (2014). Particle display: a quantitative screening method for generating high-affinity aptamers. Angew. Chem. *126*, 4896−4901.

52. Gordon, C. K. L., Eisenstein, M., and Soh, H. T. (2018). Direct selection strategy for isolating aptamers with pH-sensitive binding activity. ACS Sens. *3*, 2574−2580

53. MacPherson, I. S., Temme, J. S., and Krauss, I. J. (2017). DNA display of folded RNA libraries enabling RNA-SELEX without reverse transcription. Chem. Commun. *53(19)*, 2878–2881.

54. Hoon, S., Zhou, B., Janda, K. D., Brenner, S., and Scolnick, J. (2011). Aptamer selection by high-throughput sequencing and informatic analysis. Biotechniques. *51(6)*, 413–416.

55. Riley, K. R., Gagliano, J., Xiao, J., Libby, K., Saito, S., Yu, G., Cubicciotti, R., Macosko, J., Colyer, C. L., Guthold, M., and Bonin, K. (2015). Combining capillary electrophoresis and next-generation sequencing for aptamer selection. Anal. Bioanal. Chem. *407(6)*, 1527–1532.

56. Stoltenburg, R., Schubert, T. and Strehlitz, B. (2015). In vitro Selection and Interaction Studies of a DNA Aptamer Targeting Protein A. PloS one, *10(7)*, e0134403.

57. Sun, X., Xie, L., Qiu, S., Li, H., Zhou, Y., Zhang, H., Zhang, Y., Zhang, L., Xie, T., and Chen, Y., et al. (2022). Elucidation of CKAP4-remodeled cell mechanics in driving metastasis of bladder cancer through aptamer-based target discovery. Proc. Natl. Acad. Sci. U.S.A., *119(16)*, e2110500119.

58. Tan, S. Y., Acquah, C., Sidhu, A., Ongkudon, C. M., Yon, L. S., and Danquah, M. K. (2016). SELEX modifications and bioanalytical techniques for aptamer-target binding characterization. Crit. Rev. Anal. Chem. *46(6)*, 521–537.

59. Gade, P. S., Sonkar, R. M., and Bhatt, P. (2022). Graphene oxide-mediated fluorescence turn-on GO-FAM-FRET aptasensor for detection of sterigmatocystin. Anal. Methods. *14(39)*, 3890–3897.

60. Shigdar, S., Schrand, B., Giangrande, P. H., and de Franciscis, V. (2021). Aptamers: cutting edge of cancer therapies. Mol. Ther. *29(8)*, 2396–2411.

61. Damborsky, P., Svitel, J. and Katrlik, J. (2016). Optical biosensor. Essays. Biochem. *60*, 91–100.

62. Chen, C., and Wang, J. (2020). Optical biosensors: an exhaustive and comprehensive review. Analyst. *145*, 1605.

63. Zou, X., Wu, J., Gu, J., Shen, L., and Mao, L. (2019). Application of aptamers in virus detection and antiviral therapy. Front. Microbiol. *10*, 1462.

64. Li, J. (2021) Diverse high-affinity DNA aptamers for wild-type and B.1.1.7 SARS-CoV-2 spike proteins from a pre-structured DNA library. Nucleic Acids Res. *49(13)*, 7267–7279.

65. Gupta, A., Anand, A., Jain, N., Goswami, S., Anantharaj, A., Patil, S., Singh, R., Kumar, A., Shrivastava, T., Bhatnagar, S., et al. (2021). A novel G-quadruplex aptamer-based spike trimeric antigen test for the detection of SARS-CoV-2. Mol. Ther. Nucleic Acids. *26*, 321–332.

66. Kacherovsky, N., Yang, L. F., Dang, H. V., Cheng, E. L., Cardle, I. I., Walls, A. C., McCallum, M., Sellers, D. L., DiMaio, F., Salipante, S. J., et al. (2021). Discovery

and Characterization of Spike N-Terminal Domain-Binding Aptamers for Rapid SARS-CoV-2 Detection. Angew. Chem. Int. Ed. *60*, 21211–21215.

67. Aithal, S., Mishriki, S., Gupta, R., Sahu, R. P., Botos, G., Tanvir, S., Hanson, R. W., and Puri, I. K. (2022). SARS-CoV-2 detection with aptamer-functionalized gold nanoparticles. Talanta. *236*, 122841.

68. Svobodova, M., Skouridou, V., Jauset-Rubio, M., Viéitez, I., Fernández-Villar, A., Cabrera Alvargonzalez, J. J., Poveda, E., Bofill, C. B., Sans, T., Bashammakh, A., et al. (2021). Aptamer sandwich assay for the detection of SARS-CoV-2 spike protein antigen. ACS Omega. *6*, 35657−35666.

69. Deng, J., Tian, F., Liu, C., Liu, Y., Zhao, S., Fu, T., Sun, J., and Tan, W. (2021). Rapid one-step detection of viral particles using an aptamer-based thermophoretic assay. J. Am. Chem. Soc. *143(19)*, 7261−7266.

70. Lee, J. M., Kim, C. R., Kim, S., Min, J., Lee, M. H., and Lee, S. (2021). Mix-and-read, one-minute SARS-CoV-2 diagnostic assay: development of PIFE-based aptasensor. Chem. Commun. (Camb). *57(79)*, 10222–10225.

71. Zhang, L., Fang, X., Liu, X., Ou, H., Zhang, H., Wang, J., Li, Q., Cheng, H., Zhang, W., and Luo, Z. (2020). Discovery of sandwich type COVID-19 nucleocapsid protein DNA aptamers. Chem. Commun., *56(70)*, 10235–10238.

72. Stanborough, T., Given, F.M., Koch, B., Sheen, C.R., Stowers-Hull, A.B., Waterland, M.R., and Crittenden, D.L. (2021). Optical detection of CoV-SARS-2 viral proteins to sub-picomolar concentrations. ACS Omega. *6(9)*, 6404−6413.

73. Song, Y., Song, J., Wei, X., Huang, M., Sun, M., Zhu, L., Lin, B., Shen, H., Zhu, Z., and Yang, C. (2020). Discovery of aptamers targeting the receptor-binding domain of the SARS-CoV-2 spike glycoprotein. Anal. Chem. *92(14)*, 9895−9900.

74. Cennamo, N., Pasquardini, L., Arcadio, F., Lunelli, L., Vanzetti, L., Carafa, V., Altucci, L., and Zeni, L. (2021). SARS-CoV-2 spike protein detection through a plasmonic D-shaped plastic optical fiber aptasensor. Talanta. *233*, 122532.

75. Lewis, T., Giroux, E., Jovic, M., and Martic-Milne, S. (2021). Localized surface plasmon resonance aptasensor for selective detection of SARS-CoV-2 S1 protein. Analyst. *146(23)*, 7207.

76. Lee, T., Kim, G. H., Kim, S. M., Hong, K., Kim, Y., Park, C., Sohn, H., and Min, J. (2019). Label-free localized surface plasmon resonance biosensor composed of multi-functional DNA 3 way junction on hollow Au spike-like nanoparticles (HAuSN) for avian influenza virus detection. Colloids. Surf. B: Biointerfaces. *182*, 110341.

77. Cho, S.-J., Woo, H.-M., Kim, K.-S., Oh, J.-W., and Jeong, Y.-J. (2011). Novel system for detecting SARS coronavirus nucleocapsid protein using an ssDNA aptamer. J. Biosci. Bioeng. *112*, 535–540.

78. Zou, X., Wu, J., Gu, J., Shen, L., and Mao, L. (2019). Application of aptamers in virus detection and antiviral therapy. Front. Microbiol. *10*,1462.

79. Schmitz, A., Weber, A., Bayin, M., Breuers, S., Fieberg, V., Famulok, M., and Mayer, G. (2021). A SARS-CoV-2 spike binding DNA aptamer that inhibits pseudovirus infection by an RBD-independent mechanism. Angew. Chemie - Int. Ed. *60*, 10279–10285.

80. Liu, X., Wang, Y., Wu, J., Qi, J., Zeng, Z., Wan, Q., Chen, Z., Manandhar, P., Cavener, V.S., Boyle, N.R., et al. (2021). Neutralizing aptamers block S/RBD-ACE2 interactions and prevent host cell infection. Angew. Chemie, Int. Ed. *60*, 10273–10278.

81. Savastano, A., de Opakua, A.I., Rankovic, M., and Zweckstetter, M. (2020). Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Nat. Commun. *11*, 6041.

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

83. Che, X.-Y., Hao, W., Wang, Y., Di, B., Yin, K., Xu, Y.-C., Feng, C.-S., Wan, Z.-Y., Cheng, V.C.C., and Yuen, K.-Y. (2004). Nucleocapsid protein as early diagnostic marker for SARS. Emerg. Infect. Dis. *10*, 1947–1949.

84. Ge, C., Feng, J., Zhang, J., Hu, K., Wang, D., Zha, L., Hu, X., and Li, R. (2022). Aptamer/antibody sandwich method for digital detection of SARS-CoV-2 nucleocapsid protein. Talanta *236*, 122847.

85. Ramanathan, S., Gopinath, S.C.B., Ismail, Z.H., Arshad, M.K.M., and Poopalan, P. (2022). Aptasensing nucleocapsid protein on nanodiamond assembled gold interdigitated electrodes for impedimetric SARS-CoV-2 infectious disease assessment. Biosens. Bioelectron. *197*, 113735.

86. Chen, R., Kan, L., Duan, F., He, L., Wang, M., Cui, J., Zhang, Z., and Zhang, Z. (2021). Surface plasmon resonance aptasensor based on niobium carbide MXene quantum dots for nucleocapsid of SARS-CoV-2 detection. Mikrochim. acta. *188*(10), 316.

87. Berezovski, M. V, Lechmann, M., Musheev, M.U., Mak, T.W., and Krylov, S.N. (2008). Aptamer-facilitated biomarker discovery (AptaBiD). J. Am. Chem. Soc. *130*, 9137–9143.

88. Blind, M., and Blank, M. (2015). Aptamer selection technology and recent advances. Mol. Ther. Nucleic. Acids. *4(1)*, e223.

89. Avci-Adali, M., Paul, A., Wilhelm, N., Ziemer, G., and Wendel, H.P. (2010). Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation. Molecules. *15*, 1-11.

90. Wilson, R. (2011). Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads. Nucleic Acid Ther. *21(6)*, 437-40.

91. Luthra, R., Singh, R.R. and Patel, K.P. (2016). Clinical applications of PCR, Methods in Molecular Biology, vol 1392. New York: Humana Press.

92. Yeoh, T.S., Anna, A., Tang, T.H., and Citartan, M. (2022). Development of an optimization pipeline of asymmetric PCR towards the generation of DNA aptamers: a guide for beginners. World J. Microbiol. Biotechnol. *38(2)*, 31.

93. Fraser, L.A., Kinghorn, A.B., Tang, M.S.L., Cheung, Y.-W., Lim, B., Liang, S., Dirkzwager, R.M., and Tanner, J.A. (2015). Oligonucleotide functionalized microbeads: indispensable tools for high-throughput aptamer selection. Molecules. *20*, 21298-21312.

94. Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006). Amplification of complex gene libraries by emulsion PCR. Nat. Methods. *3(7)*, 545–550.

95. Shao, K., Ding, W., Wang, F., Li, H., Ma, D., and Wang, H. (2011). Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection. PLoS One. *6(9)*, e24910.

96. Witt, M., Phung, N. L., Stalke, A., Walter, J. G., Stahl, F., von Neuhoff, N., and Scheper, T. (2017). Comparing two conventional methods of emulsion PCR and optimizing of Tegosoft-based emulsion PCR. Eng. Life Sci. *17(8)*, 953–958.

97. Yufa, R., Krylova, S.M., Bruce, C., Bagg, E.A., Schofield, C.J., and Krylov, S.N. (2015). Emulsion PCR significantly improves nonequilibrium capillary electrophoresis of equilibrium mixtures-based aptamer selection: allowing for efficient and rapid selection of aptamer to unmodified ABH2 protein. Anal. Chem. *87(2)*, 1411-9.

98. Liu, H., Zhou, Y., Xu, Q., and Wong, S.M. (2020). Selection of DNA aptamers for subcellular localization of RBSDV P10 protein in the midgut of small brown planthoppers by emulsion PCR-based SELEX. Viruses. *12(11)*, 1239.

99. Siu, R. H. P., Liu, Y., Chan, K. H. Y., Ridzewski, C., Slaughter, L. S., and Wu, A. R. (2021). Optimization of on-bead emulsion polymerase chain reaction based on single particle analysis. Talanta. *221*, 121593.

100. Alam, K.K., Chang, J.L., and Burke, D.H. (2015). FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. Mol. Ther. Nucleic. Acids. *4*, e230.

101. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic. Acids. Res. *50(W1)*, W276-W279.

102. Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform. *11*, 129.

103. Tolnai, Z., Harkai, Á., Szeitner, Z., Scholz, É. N., Percze, K., Gyurkovics, A., and Mészáros, T. (2019). A simple modification increases specificity and efficiency of asymmetric PCR. Anal. Chim. Acta. *1047*, 225-230.

104. Yeoh, T.S., Anna, A., Tang, T.H., and Citartan, M. (2022). Development of an optimization pipeline of asymmetric PCR towards the generation of DNA aptamers: a guide for beginners. World J. Microbiol. Biotechnol. *38(2)*, 31.

105. Shao, K., Ding, W., Wang, F., Li, H., Ma, D., and Wang, H. (2011). Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection. PLoS One. *6(9)*, e24910.

106. Li, J., Zhang, Z., Gu, J., Stacey, H. D., Ang, J. C., Capretta, A., Filipe, C. D. M., Mossman, K. L., Balion, C., Salena, B. J., et al. (2021). Diverse high-affinity DNA aptamers for wild-type and B.1.1.7 SARS-CoV-2 spike proteins from a pre-structured DNA library. Nucleic Acids Res. *49(13)*, 7267–7279.

107. Verma, V., Gupta, A. and Chaudhary, V.K. (2020). Emulsion PCR made easy. Biotechniques. *69*, 65–69.

108. Kraytsberg, Y., and Khrapko, K. (2005). Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. Expert Rev. Mol. Diagn. *5*, 809–815.

109. Ouellet, E., Foley, J.H., Conway, E.M., and Haynes, C. (2015). Hi-Fi SELEX: A high-fidelity digital-PCR based therapeutic aptamer discovery platform. Biotechnol. Bioeng. *112*, 1506–1522.

110. Yu, H., Alkhamis, O., Canoura, J., Liu, Y., and Xiao, Y. (2021). Advances and challenges in small-molecule DNA aptamer isolation, characterization, and sensor development. Angew. Chem. Int. Ed. Engl. *60(31)*, 16800–16823.

111. Freyer, M.W., and Lewis, E.A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods. Cell. Bio. *84*, 79–113.

112. Johnson C. M. (2021). Isothermal titration calorimetry. Methods Mol. Bio. *2263*, 135–159.

113. Guo, X. (2012). Surface plasmon resonance based biosensor technique: a review. J. Biophotonics. *5(7)*, 483–501.

114. Kamat, V., and Rafique, A. (2017). Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. Anal. Biochem. *536*,16-31.

115. Concepcion, J., Witte, K., Wartchow, C., Choo, S., Yao, D., Persson, H., Wei, J., Li, P., Heidecker, B., Ma, W., Varma, R., et al. (2009). Label-free detection of biomolecular interactions using bio-layer interferometry for kinetic characterization. Comb. Chem. High Throughput Screen. *12(8)*, 791-800.

116. Petersen, R.L. (2017). Strategies using bio-layer interferometry biosensor technology for vaccine research and development. Biosensors. *7(4)*, 49.

117. Kamat, V., and Rafique, A. (2017). Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. Anal. Biochem. *536*, 16-31.

118. Kumaraswamy, S., and Tobias, R. (2015). Label-free kinetic analysis of an antibody-antigen interaction using biolayer interferometry. Methods Mol. Biol. *1278*, 165-82.

119. Tobias, R., Kumaraswamy, S., and Apiyo, D. (2021). Biomolecular binding kinetics assays on the Octet® platform. Sartorius Application Note.

120. Dhiman, A., Anand, A., Malhotra, A., Khan, E., Santra, V., Kumar, A., and Sharma, T.K. (2018). Rational truncation of aptamer for cross-species application to detect krait envenomation. Sci. Rep. *8*, 17795.

121. Ruscito, A., and DeRosa, M.C. (2016). Small-molecule binding aptamers: selection strategies, characterization, and applications. Front. Chem. *4*, 14.

122. Wang, X., Gao, X., He, J., Hu, X., Li, Y., Li, X., Fan, L., and Yu, H.-Z. (2019). Systematic truncating of aptamers to create high-performance graphene oxide (GO)-based aptasensors for the multiplex detection of mycotoxins. Analyst *144*, 3826–3835.

123. Wu, C., Qavi, A.J., Hachim, A., Kavian, N., Cole, A.R., Moyle, A.B., Wagner, N.D., Sweeney-Gibbons, J., Rohrs, H.W., Gross, M.L., et al. (2021). Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. IScience. *24*, 102681.

124. Gao, T., Gao, Y., Liu, X., Nie, Z., Sun, H., Lin, K., Peng, H., and Wang, S. (2021). Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. BMC Microbiol. *21*, 58.

125. Stocks, B. B., Thibeault, M. P., Schrag, J. D., and Melanson, J. E. (2022). Characterization of a SARS-CoV-2 spike protein reference material. Anal. Bioanal. Chem. *414*, 3561–3569 (2022)

126. Yang, G., Li, Z., Mohammed, I., Zhao, L., Wei, W., Xiao, H., Guo, W., Zhao, Y., Qu, F., and Huang, Y. (2021). Identification of SARS-CoV-2-against aptamer with high neutralization activity by blocking the RBD domain of spike protein 1. Sig. Transduct. Target. Ther. 6, 227.

127. Shi, L., Wang, L., Ma, X., Fang, X., Xiang, L., Yi, Y., Li, J., Luo, Z., and Li, G. (2021). Aptamer-functionalized nanochannels for one-step detection of SARS-CoV-2 in samples from COVID-19 patients. Anal. Chem. *93(49)*, 16646–16654.

128. Narayan, C., Kwon, J., Kim, C., Kim, S.J., and Jang, S.K. (2020). Virus-based SELEX (viro-SELEX) allows development of aptamers targeting knotty proteins. Analyst. *145(4)*, 1473–1482.

129. Rahman, M.S., Han, M.J., Kim, S.W., Kang, S.M., Kim, B.R., Kim, H., Lee, C.J., Noh, J.E., Kim, H., Lee, J.-O., et al. (2023). Structure-guided development of bivalent aptamers blocking SARS-CoV-2 infection. Molecules. *28*, 4645.

130. Lupu, L.-M., Wiegand, P., Holdschick, D., Mihoc, D., Maeser, S., Rawer, S., Völklein, F., Malek, E., Barka, F., Knauer, S., et al. (2021). Identification and affinity determination of protein-antibody and protein-aptamer epitopes by biosensor-mass spectrometry combination. Int. J. Mol. Sci. *22*, 12832.

131. Ion, L., and Petre, B.A. (2019). Immuno-affinity mass spectrometry: a novel approaches with biomedical relevance. Adv. Exp. Med. Biol. *1140*, 377-388.

132. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J. V, and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome. Res. *10*, 1794–1805.

133. Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top orbitrap. Mol. Cell. Proteomics. *11*, M111.013722.

134. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526.

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

136. Morozov, D., Mironov, V., Moryachkov, R. V, Shchugoreva, I.A., Artyushenko, P. V, Zamay, G.S., Kolovskaya, O.S., Zamay, T.N., Krat, A. V, Molodenskiy, D.S., et al. (2021). The role of SAXS and molecular simulations in 3D structure elucidation of a DNA aptamer against lung cancer. Mol. Ther. Nucleic Acids. *25*, 316–327.

137. Yan, Y., Tao, H., He, J., and Huang, S.-Y. (2020). The HDOCK server for integrated protein-protein docking. Nat. Protoc. *15*, 1829–1852.

138. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. *1–2*, 19–25.

139. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696–3713.

140. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

141. Heyer, L.J., Kruglyak, S., and Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. Genome Res. *9*, 1106–1115.

142. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

143. Owczarzy, R., Tataurov, A. V, Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G., Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O., et al. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res. *36*, W163-9.

144. Boniecki, M.J., Lach, G., Dawson, W.K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K.M., and Bujnicki, J.M. (2016). SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res. *44*, e63.

145. Sankararaman, S., Hamre, J. 3rd, Almsned, F., Aljouie, A., Bokhari, Y., Alawwad, M., Alomair, L., and Jafri, M.S. (2022). Active site prediction of phosphorylated SARS-CoV-2 N-Protein using molecular simulation. Inform. Med. Unlocked *29*, 100889.

146. Jackson, C. B., Farzan, M., Chen, B., and Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. Nat. Rev. Mol. Cell Biol. *23*, 3–20.

147. Zhang, J., Xiao, T., Cai, Y., and Chen, B. (2021). Structure of SARS-CoV-2 spike protein, Curr. Opin. Virol. *50*, 173-182.

148. Quaglia, F., Salladini, E., Carraro, M., Minervini, G., Tosatto, S. C. E., and Le Mercier, P. (2022). SARS-CoV-2 variants preferentially emerge at intrinsically disordered protein sites helping immune evasion. FEBS J. *289(14)*, 4240–4250.

149. Poolsup, S., Zaripov, E., Hüttmann, N., Minic, Z. Artyushenko, P.V., Shchugoreva, I.A., Tomilin, F.N., Kichkailo, A.S. and Berezovski. M.V. (2023). Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics. Mol. Ther. Nucleic Acids. *31*, 731-743.

150. Hernández-Rodríguez, J.F., Rojas, D., and Escarpa, A. (2021) Electrochemical sensing directions for next-generation healthcare: trends, challenges, and frontiers. Anal. Chem. *93 (1)*, 167-183.

151. Sen, D., and Lazenby, R.A. (2023) Selective aptamer modification of Au surfaces in a microelectrode sensor array for simultaneous detection of multiple analytes. Anal. Chem. *95(17)*, 6828-6835.

152. Yang, L.F., Kacherovsky, N., Liang, J., Salipante, S.J., and Pun, S.H. (2022). SCORe: SARS-CoV-2 Omicron variant RBD-binding DNA aptamer for multiplexed rapid detection and pseudovirus neutralization. Anal. Chem. *94 (37)*, 12683-12690.

153. Yang, L.F., Kacherovsky, N., Panpradist, N., Wan, R., Liang, J., Zhang, B., Salipante, S.J., Lutz, B.R., and Pun, S.H. (2022). Aptamer sandwich lateral flow assay (AptaFlow) for antibody-free SARS-CoV-2 detection. Anal. Chem. *94 (20)*, 7278-7285.

154. Li, J., Zhang, Z., Gu, J., Amini, R., Mansfield, A.G., Xia, J., White, D., Stacey, H.D., Ang, J.C., et al. (2022). Three on three: universal and high-affinity molecular recognition of the symmetric homotrimeric spike protein of SARS-CoV-2 with a symmetric homotrimeric aptamer. J. Am. Chem. Soc. *144 (51)*, 23465-23473.

155. Schaub, J.M., Chou, C.W., Kuo, H.C., Javanmardi, K., Hsieh, C.L., Goldsmith, J., DiVenere, A.M., Le, K.C., Wrapp, D., Byrne, P.O., et al. (2021). Expression and characterization of SARS-CoV-2 spike proteins. Nat. Protoc. *16*, 5339–5356.

156. Kang, J., Jang, H., Yeom, G., and Kim, M.G. (2021). Ultrasensitive detection platform of disease biomarkers based on recombinase polymerase amplification with H-sandwich aptamers. Anal. Chem. *93(2)*, 992–1000.

157. Quaglia, F., Salladini, E., Carraro, M., Minervini, G., Tosatto, S. C. E., and Le Mercier, P. (2022). SARS-CoV-2 variants preferentially emerge at intrinsically disordered protein sites helping immune evasion. FEBS J. *289(14)*, 4240–4250.