BIOPHYSICAL AND BIOCHEMICAL CHARACTERIZATION OF AN HBV cccDNA G-QUADRUPLEX TARGETING THERAPEUTIC

GERARDO KARARI BALDERAS FIGUEROA

Bachelor of Science, University of Lethbridge, 2017

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

In

BIOCHEMISTRY

Department of Biochemistry and Chemistry
Faculty of Arts and Science
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

BIOPHYSICAL AND BIOCHEMICAL CHARACTERIZATION OF A NOVEL HBV cccDNA G-QUADRUPLEX TARGETING THERAPEUTIC

GERARDO K. BALDERAS FIGUEROA B. Sc.

Date of Defense: August 8th 2023

Dr. Trushar Patel Thesis Supervisor	Associate Professor	Ph. D.
Dr. Tony Russell Thesis Examination Committee Member	Associate Professor	Ph. D.
Dr. Athan Zovoilis Thesis Examination Committee Member	Associate Professor	Ph. D.
Dr. Jean-Denys Hamel Thesis Examination Committee Chair	Assistant Professor	Ph. D.

ABSTRACT

Hepatitis B virus (HBV) has chronically infected 296 million people, and it is the leading cause of cirrhosis and hepatocellular carcinoma (HCC). Current treatments target post-transcriptional steps of the viral cycle, which leaves HBV's covalently closed circular DNA (cccDNA) nuclear reservoirs unaffected. This allows patients to relapse if therapy ceases. We have previously identified a G-quadruplex (G4) forming region in the HBV preCore promoter involved in viral replication. Here, we develop recombinant single-domain antibodies (sdAbs) that can target this G4-forming region. The sdAbs are purified by chromatography, and their interaction with HBV G4 is characterized by biochemical binding assays as well as biophysical techniques. We conclude that sdAbs can be highly structure and sequence specific towards HBV G4. Overall, my work develops a pipeline for G4-targeting that can be used in future HBV cccDNA studies as well as be implemented in the study of other G4s involved in disease.

ACKNOWLEDGMENTS

I want to extend my gratitude to all the members of the biochemistry program I had the opportunity to work with, learn from, and draw inspiration from throughout my program. First and foremost, thank you, Dr. Trushar Patel, for allowing me to be part of your laboratory, for mentoring me for the past few years, but more importantly, for being a role model by being involved in a wide range of activities outside research. Thank you, Dr. Maulik Badmalia, for giving me the tools necessary to complete my thesis work and for always being willing to help me navigate graduate school. Thank you, Simmone D'souza, for being my thesis "twin," for being easy to work with, and for inspiring me to maintain a work-life balance. I want to thank all of the undergraduate students I had the opportunity to work with and mentor. However, a special thank you to Zachary Robinson and my sister Sara Balderas whose work became integral for the completion of my thesis work and whose work ethic inspires me daily. I want also to thank all the past and present BIPOC and LGBTQ+ scientists who have paved the way, extended a guiding hand in my life and highlighted the importance of visibility in the field.

Thank you, Dr. Tony Russell and Dr. Anthony Zovoilis, for being part of my committee.

I want to thank my family for supporting and encouraging my scientific endeavors. However, I want to specially thank my veterinarian mother for planting a scientific curiosity seed in me at a young age and always encouraging me to understand the natural world.

Lastly, thank you, my circus, ex-mormon, queer, and outreach (cultural or scientific) friends, for being excellent support systems and helping me unplug from research when necessary.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BLI	Bio-Layer Interferometry
BME	β - Mercaptoethanol
CAD	Canadian Dollar
cccDNA	covalently closed circular DNA
CD Spec	Circular Dichroism Spectrophotometry
CDR	Complementarity-Determining Region
CPP	Cell-Penetrating Peptide
CV	Column Volume
DLS	Dynamic Light Scattering
ETV	Entecavir
FLAG	DYKDDDDK (peptide sequence)
FT	Flow-Through
G4	G-Quadruplex
GFP	Green Fluorescent Protein
GST	Glutathione S-Transferase
HA	Hemagglutinin
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
His	Histidine
HSPG	Heparan Sulfate Proteoglycan
HT	Horizontal Transmission

IDP	Intrinsic Disorder Predictor
IFN-α	Interferon α
IgG	Immunoglobulin G
IMAC	Immobilized Metal Affinity Chromatography
ITC	Isothermal Titration Calorimetry
L	Linker
LMV	Lamivudine
MALS	Multiple Angle Light Scattering
MBP	Mean Disorder Prediction
MST	Microscale Thermophoresis
NA	Nucleoside Analogue
Ni-NTA	Nickel-nitrilotriacetic acid
NLS	Nuclear Localization Sequence
NTCP	Sodium Taurocholate Co-transporting Polypeptide
NUCs	Nucleos(t)ide analogue therapy
PEG	Pegylated
pgRNA	pregenomic RNA
PTM	Post-Translational Modifications
Pol	Polymerase
preC G4	SP1 preCore promoter G4
RBD	Receptor-Binding Domain
rcDNA	relaxed circular DNA
Rg	Radius of Gyration

Rh	Hydrodynamic Radius
RI	Refractive Index
RL	Rigid Linker
RNAi	RNA interference
RT	Reverse Transcription
SAXS	Small Angle X-ray Scattering
scFv	Single Chain Variable Fragment
sdAb	single domain antibody
SEC	Size Exclusion Chromatography
TDF	Tenofovir Disoproxil
VHH	Variable Heavy domain of Heavy chain
VNAR	Variable domain of New Antigen Receptor
VT	Vertical Transmission
W1	Wash 1
W2	Wash 2
WHO	World Health Organization
WT	Wild Type

CHAPTER 1: INTRODUCTION

1.1 CURRENT NATIONAL AND GLOBAL IMPACT OF HBV

Hepatitis B virus (HBV) is a leading cause of liver cirrhosis and hepatocellular carcinoma (HCC) [1] where 15-40% of chronic cases progress in this manner [2]. The WHO has estimated that in 2019, this hepatotropic virus was involved in 820,000 deaths worldwide. Extrapolating from current studies, it can be estimated that the 10-year cumulative cost of treatment will be between $6,128-$38,968 (CAD) per chronically infected Canadian patient [3]. In 2019, The public health agency of Canada reported 3,790 chronic cases in the country [4]. Based on these estimates, the total 10-year cost of Canadian chronic infection would be between $23 and $148 million. New strategies are needed to curb the costs of chronic infection in Canadian and worldwide healthcare.

Interestingly, the rate of acute versus chronic infections in Canada is 0.48 versus 10.2 per 100,000 people, where only 178 cases of acute infection were detected in 2019 [4]. Acute infection is mainly a product of adult horizontal transmission (HT) in the case of sexual or bodily fluid transmission [5]. In contrast, it has been observed in endemic regions that chronic infection is derived through vertical transmission (VT), where exposure to HBV occurs at an early age [5]. Current HBV vaccines are effective in infection prevention (98-100% protection) [1]. The WHO recommends vaccination within 24 hours of birth, but it is estimated that only 43% of the world population is vaccinated at birth [1]. Although in Canada, early prevention at or before two months is common in most provinces [6-9], the rate of infection per 100 000 people has not changed greatly where it stays relatively the same from 2010 (12.3) to 2019 (10.2) [4].At the

national level, the number of chronic cases in Canada still represents a significant economic burden.

On a global scale, there has been a rise in chronic infections worldwide from ~250 million worldwide cases in 2017 to 296 million cases in 2019 [1, 10]. The current vaccination efforts and treatment options alone have not been effective at eradicating the spread of HBV infection. Additionally, the ongoing costs of treating chronic cases make the disease a significant global and national health concern. New therapeutic strategies aiming to lower the cost of treatment, lower the spread of the infection and seek to cure the infection are needed.

1.2 HBV VIRAL CYCLE AND CURRENT THERAPEUTIC TREATMENT

The transmission of HBV takes two main routes, horizontal and vertical transmission. Horizontal transmission occurs through exposure to bodily fluids, mainly through contact with saliva, non-intact skin, and mucosa with blood-stained secretions [11]. It is widely accepted that horizontal transmission between family members is highly exacerbated in communities with poor socioeconomic status and hygiene as well as in highly interactive families [12]. Vertical transmission refers to infection from fluid exchange in the umbilical cord during labor or six months post-natal [12]. It has a high efficiency ranging from 70 to 90% in HBeA-positive mothers [13], and it is estimated that 90% of babies born to seropositive mothers become chronically infected [11]. The earlier an exposure occurs, the more likely it is to evolve into a chronic infection [14].

HBV is comprised of nine main genotypes of the Hepadnaviridae family [15] which exhibit different responses to interferon (IFN) and demonstrate differences in disease progression [16]. The ~3.2kb genome of all HBV genotypes has four overlapping reading frames necessary to encode for core (C), polymerase (P), Surface (S), and X proteins [15, 17, 18]. The infectious

particle's outer layer is composed of a lipid layer and three viral surface antigens derived from the same gene (large, middle and small surface antigen, Fig. 1.1) [17]. Inside the envelope of the virion, the nucleocapsid is formed by core protein and encapsulates the viral polymerase and viral genome DNA (relaxed circular DNA, rcDNA) (Fig. 1.1) [17].

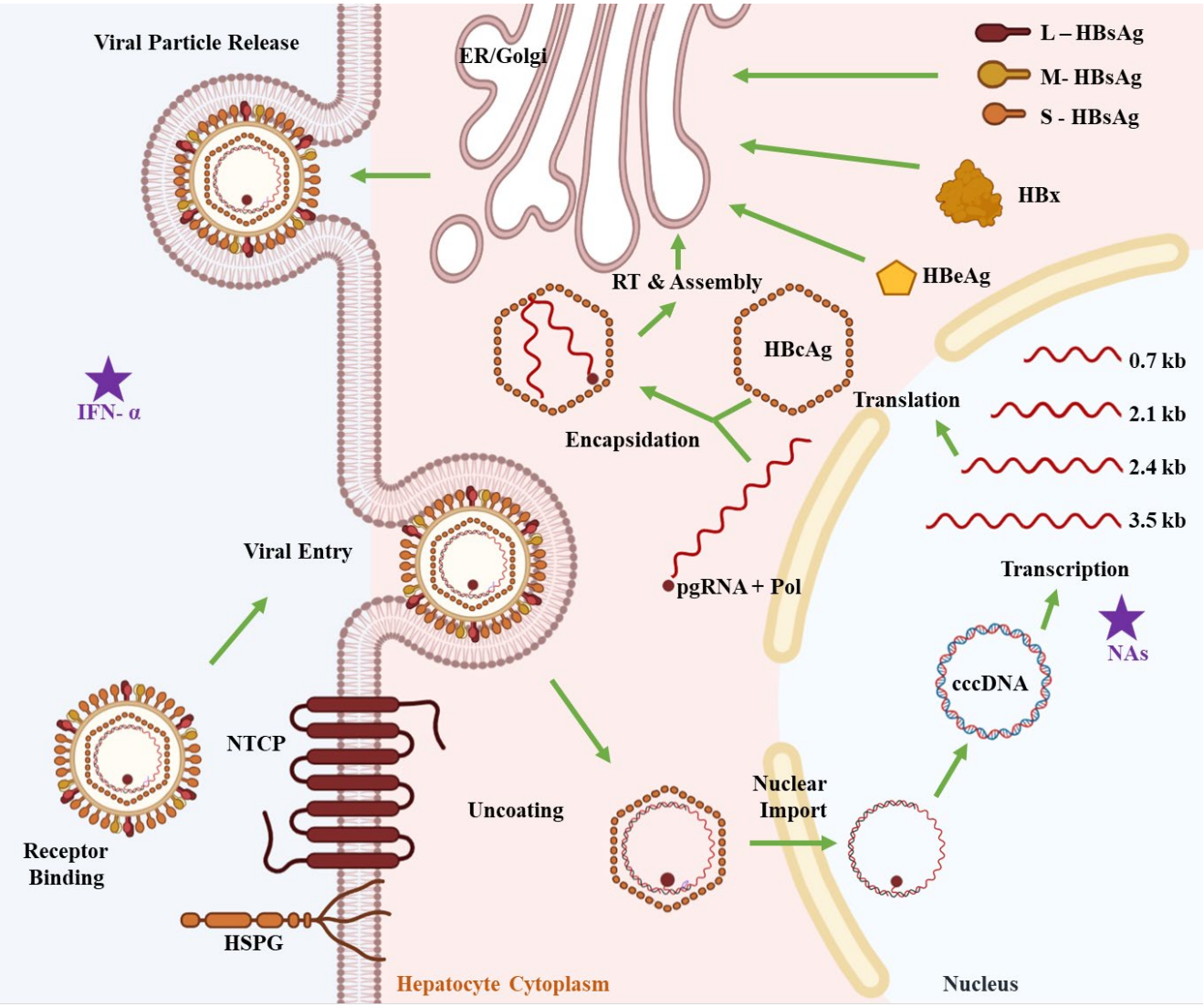

Figure 1.1 Simplified HBV viral cycle. The purple stars indicate the sites where both of the current therapeutic strategies interact to interrupt the viral cycle. "RT" stands for reverse transcription.

Viral entry into a hepatocyte is first initiated by the interaction of HBV's surface protein with heparansulfate proteoglycans (such as glypican 5) [17] and then by the interaction of the surface protein with sodium taurocholate co-transporting polypeptides (NTCP) which are found explicitly in hepatocytes for bile salt uptake [17, 19, 20]. Once in the cytosol, the nucleocapsid has been observed to depend on microtubule trafficking for its transportation to the nucleus [20]. The core protein then interacts with nuclear pore complexes where the viral relaxed circular DNA (rcDNA) is released into the nucleus [20].

Once in the nucleus, the Pol and RNA primers are removed from the rcDNA and DNA polymerases, ligases, and topoisomerases synthesize the complementary strand of DNA [17]. The rcDNA then adopts a compact and highly stable covalently closed circular DNA (cccDNA) conformation [19]. The cccDNA serves as a template for pregenomic RNA (pgRNA). Like mRNA, pgRNA is 5' capped and 3' polyadenylated [21]. These properties allow pgRNA to have functional versatility by being used for translation or encapsidation [19, 21]. Lastly, rcDNA can be derived from pgRNA by reverse transcription within the viral particle in the cytoplasm [19].

The typical "beads-on-a-string" conformation characteristic of nucleoproteins of the chromosomes has been observed in Cryo-EM studying cccDNA and, based on its small size, it has been referred to as a "minichromosome" [22]. Like conventional nucleosomes, each of the 18 nucleosomes in the cccDNA is formed by viral DNA, histones, and non-histone proteins [22, 23]. The levels of cccDNA in the nucleus are maintained throughout infection due to the stability of the minichromosomes, as well as through replenishment by intracellular amplification [24]. Furthermore, the ability of the viral DNA to form these minichromosomes allows it to camouflage in the cell and to use the host's transcriptional machinery for viral replication [25].

.

The persistence of episomal cccDNA reservoirs in hepatocytes has been noted to be the culprit of patient chronicity and relapse upon therapy discontinuation [21, 26, 27]. Chronic HBV infection therapies currently include nucleoside or nucleotide analogs (NAs) and interferon-α (IFN- α) [26]. Both therapies inhibit liver-related death by decreasing HBV DNA levels and promoting seroconversion [28]. NAs are designed to prevent reverse transcriptase activity of the HBV polymerase (Fig. 1.1) [26]. There are three main groups of nucleosides used against HBV, which include entecavir (ETV; deoxyguanosine analog), tenofovir disoproxil fumarate (TDF; acyclic phosphonate), and lamivudine (LMV; deoxy cytidine analog)) [29]. The antiviral pathway of IFN- α begins by interacting with interferon receptors in hepatocytes which activate Janus kinase-1 that then phosphorylates and dimerizes signal transducer and activator of transcription 1 (STAT1) [28]. STAT1 is responsible for inducing interferon-stimulated genes and the expression of antiviral proteins [28]. Peg-IFN-α represents an alternative treatment option in chronic cases due to their extended half-life and long-lasting activity [28]. However, it has been reported that Peg-IFN-α has various secondary effects and is effective in only 20% of patients [30].

NUCs and interferons are effective at controlling the progression of chronic HBV infection. However, the secondary effects and dependency of life-long treatment highlight a need to find alternative treatment options that can decrease patient costs, improve patient quality of life, and work towards a functional cure. Current therapeutics fail to target the culprit of chronicity in HBV infection, cccDNA, thereby therapeutic candidates looking to offer a functional cure must target this step in viral cycle [31].

1.3 cccDNA DISRUPTOR: G-QUADRUPLEXES AS NOVEL HBV THERAPEUTIC TARGETS

Genotypes A-H of HBV have >8% variance in their sequences [32], and finding regions that are conserved across the genotypes is critical to design a single therapeutic against all genotypes. Gene promoters play an essential role in regular gene expression. Mutations in promoter regions can lead to atypical gene expression, which has been linked to autism spectrum disorder [33], cancer [34, 35], and Parkinson's disease [36], amongst others in humans. Interestingly, the Human Sp1 protein (host transcription factor) has been documented to interact with some promoter regions in HBV cccDNA (Fig. 1.2) [32]. Previous work has determined that Sp1 interactions are critical for viral replication [37]. Based on this principle, sequence analysis of the X and core promoters in the eight different HBV genotypes has shown a conserved G-rich region in the preCore promoter region [38]. G-rich regions in promoters can contribute to transcriptional pausing, which aids in successful gene transcription [39]. These G-rich regions can potentially form G-quadruplex (G4) motifs [39]. At the atomic level, G-quadruplexes are guanine-rich DNA or RNA sequences that form stable quartets through Hoogsteen hydrogen bonding and are stabilized by cations [40]. This G-rich region can form highly ordered parallel G-quadruplex (G4) structures (preC-G4), which the Human Sp1 protein is known to bind to [32]. Furthermore, recent studies have determined that some mutations in the preC-G4 sequence are critical for typical HBcAg production [32].

Overall, the various reports of G4s present in viral genomes and cancers are growing [41, 42]. Their relatively recent discovery and relevance in viral replication, transcription and translation highlight their potential to be therapeutic targets [43]. Developing preC-G4 targeting therapeutics

can represent the first steps towards resolving the inner-nuclear cccDNA reservoirs in chronic HBV infection. Lastly, the development of precise anti-G4 therapeutic technology has the potential to expand into other viral infections and cancers where G4s play a crucial role in disease.

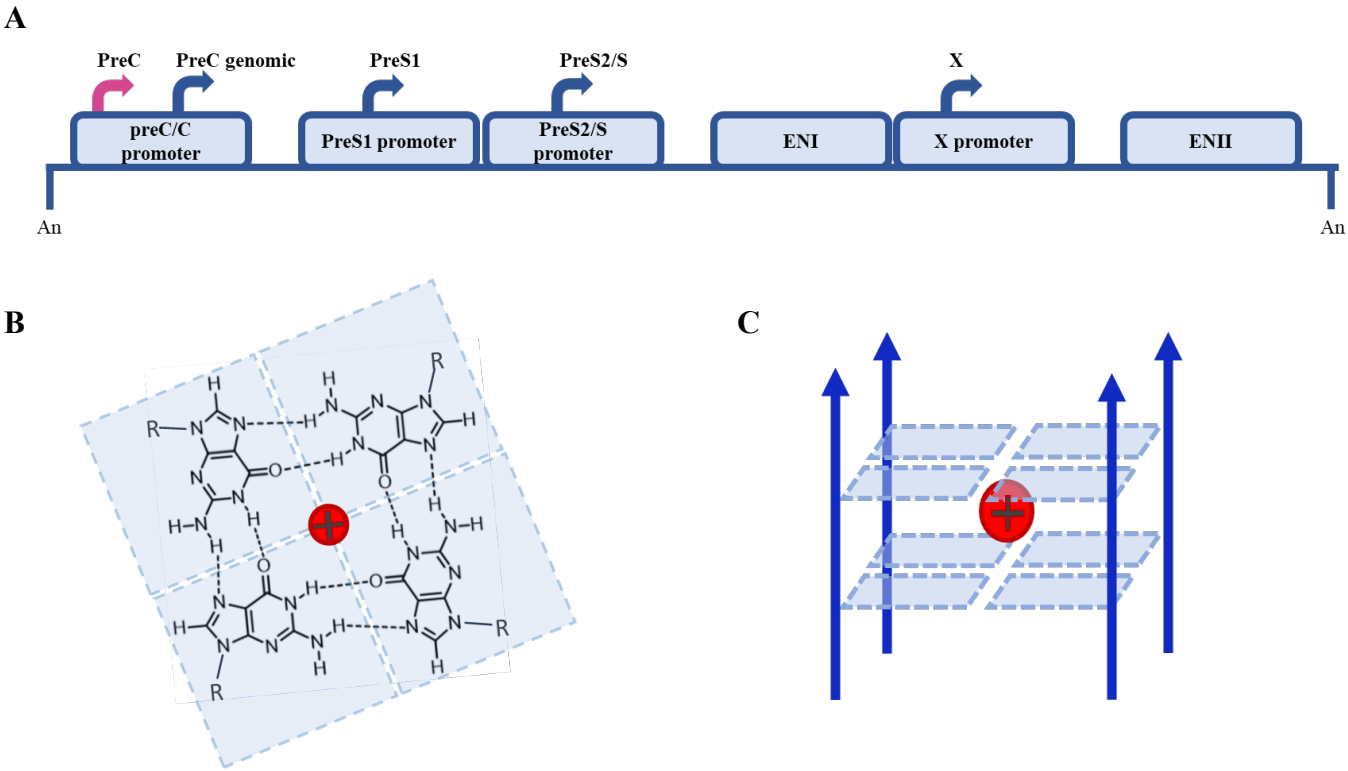

Figure 1.2. Promoter regions in HBV genome and preC G4. (A) Linearized representation of the HBV genome showing the promoter regions of the proteins that make up the viral particle (blue boxes) as well as the key sites where human interaction protein Sp1 interacts with the promoters (arrows). The pink arrow represents the preC G4 region. (B) Transverse view of a G4 showing the four guanine bases forming a quartet through Hoogsteen hydrogen bonds (dotted bonds). A cation in the middle of the quartet stabilizes the overall negative charges. (C) G-quadruplex side-

view of a parallel G4 structure where the blue arrows represent the 5’ to 3’ direction of the nucleic acid chain/chains.

1.4 DEVELOPMENT OF THERAPEUTIC CANDIDATE

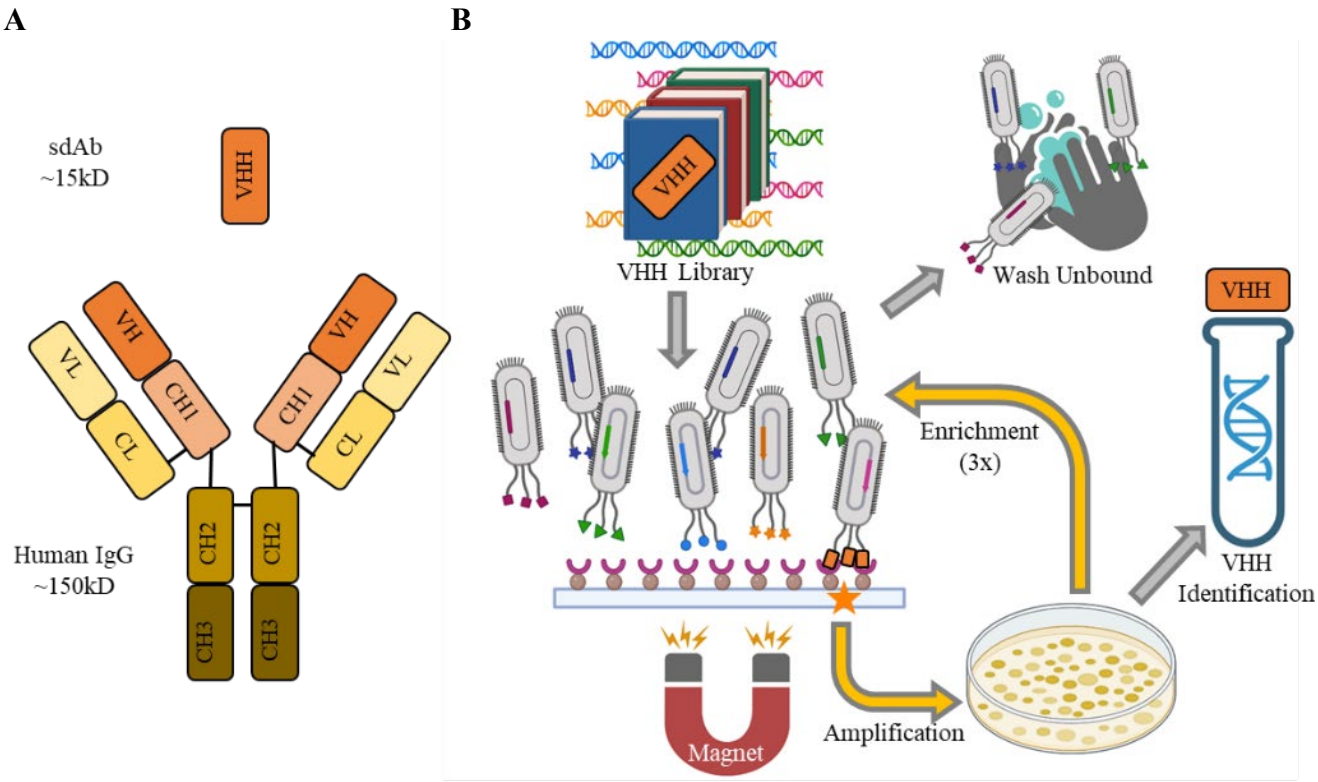

Figure 1.3 Identification of potential anti-G4 therapeutic. (A) Comparisons between the human Immunoglobulin G (IgG) and single-domain antibodies (sdAb) showing homologous structures (in orange). (B) Using a synthetic VHH library, phage display technology exposes potential candidates to the target of interest. Unbound phages are washed off and bound phages are amplified by *E. coli* infection. Amplified phages are used again to enrich the the potential binders. After three enrichment rounds, the VHHs (sdAbs) are identified.

The first therapeutic monoclonal antibody, muromonab, was approved in 1986 by the US Food and Drug Administration to treat acute transplant rejection [44]. Advancements in the field have

allowed therapeutic antibodies to be used in viral infection, cancer, autoimmunity, and inflammatory diseases [44]. Antibody therapy is advantageous due to its high specificity, but antibodies are often expensive and challenging to produce [45]. Progress has been made in the production and expression of recombinant antibodies in simpler systems like *E. coli* to target issues such as post-translational modifications (PTMs) [46]. However, antibody size still limits its ability to penetrate tissue and its pharmacokinetics [47, 48].

Single domain antibodies (sdAbs) offer an alternative technology for therapeutic development, discovered in camelids (VHH) and sharks (VNAR), they are homologous to the heavy chain region of IgG antibodies (Fig 1.3) [49]. Like traditional antibodies, sdAbs have a high affinity to their antigen [50]. Additionally, sdAbs have high physicochemical stability, small size, high tissue penetration level, and low immunogenicity [50]. Humanized sdAbs have also been developed where non-human VHH sequences are replaced with VH sequences to decrease immunogenicity [51]. Other therapeutic development research has already used sdAbs as targeting agents as is the case in ALX-0171 for respiratory syncytial virus treatment [52], sb23 for SARS-CoV-2 [53], alphaEGFR for glioblastoma [45], amongst others [49, 53, 54]. The technology's versatility to have the potential to target a variety of different antigens makes them an attractive tool for novel drug discovery efforts across a broad spectrum of diseases.

Although immunization for sdAb development is a practical approach, it takes up to 12 weeks for sdAb identification, whereas screening libraries through phage-display (Fig 1.3C) shortens the wait down to 1-2 weeks [49, 54, 55]. Furthermore, sdAbs do not require complex folding machinery or PTMs, which makes them suitable for expression in bacterial systems such as *E.*

coli [50]. These sdAb nuances provide therapeutic advantages that result in a cheaper and more efficient alternative to contemporary antibodies able to maintain the antibody's hallmark of antigen specificity and affinity [45].

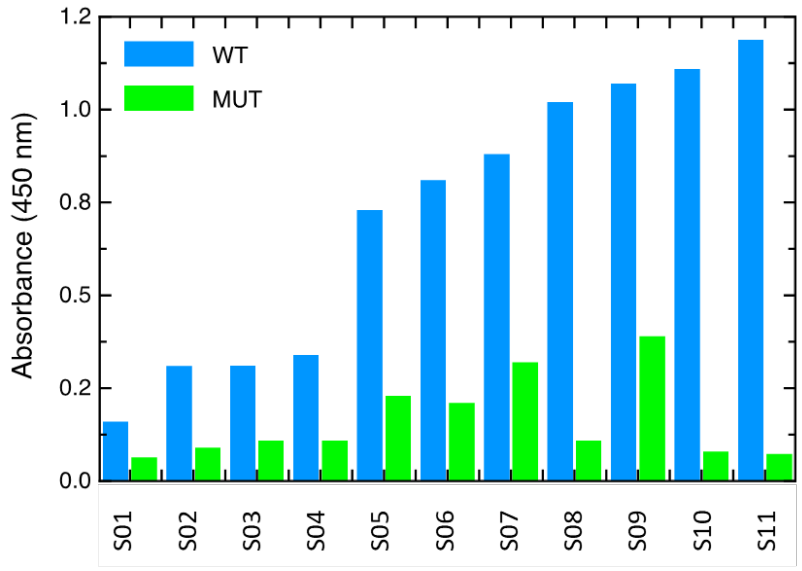

Figure 1.4 ELISA of the 11 preC G4-binding sdAb candidates identified through phage display technology, where higher absorbance level is indicative of a stronger interaction.

Tissue-specific delivery continues to be an obstacle in the success of therapeutic development [56-58]. However, peptide tags have been used to increase the protein performance or equip proteins with new biochemical properties [59]. In regards to therapeutic development, cell-penetrating peptides [60] and nuclear localization signal (NLS) sequences [61] have been developed over the past few decades to appropriate protein delivery into tissues and within-cell localization. This simple approach has already been shown to have succeeded with other sdAbs to localize them into the cell nucleus [62]. Using a similar approach, sdAbs can be designed to contain hepatocyte-specific peptide markers that seek to improve nuclear-delivery in future studies.

1.5 PROJECT OBJECTIVES

Eleven different preC-G4 binding sdAbs have been identified (S01-S11) from a humanized synthetic library (Fig 1.4). The first portion of this project focuses on testing various conditions to optimize sdAb expression by *E. coli* and purification through chromatography techniques. Next, binding studies at the biochemical level will help us determine which sdAb is the most appropriate candidate to proceed with therapeutic development. Interaction studies will also assess whether peptide-tagged sdAbs perform differently and whether this is an appropriate delivery strategy in our case. The interaction of the most promising candidates will also be studied at a biophysical level where the stoichiometry and presence of the complex will be addressed. These studies will also have the goal to propose a 3D model of the sdAb-preC G4 interaction. Lastly, strategies to predict peptide-tagged sdAb performance based on intrinsic disorder and homology are explored.

Overall, one of the goals of this project is to lay the groundwork necessary to eliminate the gap in HBV therapeutics where no cccDNA-targeting therapeutic has been approved for public use. Additionally, this work shows the ability of sdAbs to discriminate between sequences and DNA secondary structure. These findings show the potential of sdAb therapeutics in targeting key G4 regions essential in disease progression or chronicity outside HBV infection.

CHAPTER 2: GENERATION OF SINGLE DOMAIN ANTIBODIES AND G-QUADRUPLEXES

2.1 INTRODUCTION

Production of therapeutic proteins require refinement steps that seek to improve expression, increase yields, improve homogeneity, prevent aggregation, or improve the protein's physicochemical and pharmacokinetic properties [63-65]. Tracking and improving the analytes' purity and homogeneity allows them to perform consistently and avoids misleading results that culminate in failed downstream applications (such as clinical trials) [66]. The presence of harmful contaminants, such as lipopolysaccharides, in recombinant proteins derived from bacteria has been known to induce a cytokine storm as an immune response [67]. Different approaches of recombinant protein expression and purification strategies have been used looking to optimize the protein of interest.

Technology for recombinant protein production has evolved widely with an expansion of vectors available and expression host organisms ranging from simple prokaryotic organisms such as *E. coli* [68], to human cells [69] and even venturing into cell-free systems [70]. It is essential to match the expression vector and system to the requirements of therapeutic proteins that will ensure proper protein function. Eukaryotic protein expression often requires post-translational modifications (PTMs) that are yet to be understood or difficult to replicate in non-eukaryotic expression systems [71-73]. Additionally PTMs, such as N-glycosylation, can play a significant role in monoclonal antibody therapy and affect performance [74]. However, sdAbs do not have an Fc domain, so an N-linked oligosaccharide is unnecessary in the recombinant production of sdAbs [75]. This property places sdAbs in an advantageous position where they qualify to be

expressed in bacteria, which are systems known to be high-yielding, cost-effective, fast-growing, simple to culture and simple to induce protein expression [75].

Chromatography is a widely used purification technique where immobilized metal-ion affinity chromatography (IMAC), a common first step in tagged protein purification, dates back to 1975 [76]. Affinity chromatography is a powerful technology that allows tagged protein in lysate to bind to the immobilized resin while untagged protein, along with other molecules, flow through the column [77]. A variety of tags have been developed to solve different protein purification needs, where some of the tags include glutathione S-transferase (GST), histidine (His), Maltose-binding protein (MBP), and FLAG tags [77, 78]. Once unwanted molecules are not present in the column, a releasing agent is added, which competitively binds to the resin and allows for the protein of interest to be released [77]. Through this approach, the protein of interest can reach high levels of purity due to the high specificity of the resins used. However, Ni-NTA and SDS-PAGEs provide limited information regarding the types and sizes of species present in a sample, where monomers, dimers, and other oligomers may be present. Other refining techniques, such as size exclusion chromatography (SEC) are required to avoid having analyte samples composed of monomers, dimers, etc [66].

The recent SARS-Cov-2 pandemic and the sudden emergence of monkeypox in the past two years have brought infectious disease research into the spotlight. This newly developed attention has revealed the need to create faster therapeutic development approaches. Studies looking at sdAbs have used various purification approaches, and a well-established purification protocol is still needed. In the following work, *E. coli* is used as an expression system to produce

recombinant sdAbs. It also employs chromatography techniques such as Ni-NTA and SEC to achieve a homogenous protein sample that can be used in studies looking to characterize sdAb biochemical and biophysical properties. Through this work, an in-house sdAb expression and purification system is established and is used across many sdAbs with minor protocol modifications. Lastly, preC-G4 is also folded and purified to test with sdAbs in biochemical and biophysical characterization studies.

2.1 METHODS

2.1.1 Identification of G4-binding sdAbs

Biotin-labelled preC G4 in wildtype (WT) and mutants were given to Hybrigenics Services SAS (Hybrigenics Services sas, 3-5 Impasse Reille, 75014 Paris, France, www.hybrigenics-services.com). There, a synthetic hsd2Ab VHH library with $3x10^9$ clones was expressed on the surface of M13 phages. Biotin-labelled G4s were bound to streptavidin magnetic beads (Dynabeads ® M-280 Streptavidin, Life Technologies) with a concentration of biotinylated antigen of 50 nM for the first round and 10 nM for the second and third round. The hs2Ab Phage Display library was incubated with a G-quadruplex mutant (known to disrupt G4-structrue) to remove non-specific binders. Unbound VHHs expressed in *E. coli* supernatant were incubated with WT preC G4-biotin beads. The screening buffer was 20 mM HEPES, 100 mM KCL and 1mm EDTA. The resulting *E. coli*-produced phages were used in ELISAs with HRP-conjugated anti-M13 antibody (GE Healthcare) and a colorimetric substrate (TMB, TetraMethylBenzidine, Thermo Fischer). All positive hits were considered preC G4 binders and selected for sequence analysis. The sdAbs with the strongest affinity and seemingly higher specificity (from pull-

down) were selected for bacterial expression. The sdAb 01 was chosen as a negative control for biochemical characterization studies based on the phage-ELISA results.

2.1.2 Recombinant sdAb Expression and Purification

Recombinant sdAbs were produced by transforming *E. coli* Lemo 21 cells with pHEN2 vectors (as provided by Hybrigenics Services), pET22b+, or pET28a+ through heat shock. pET22b+ was used for modified sdAbs, and pET28a+ was used for SAXS studies and S10-3. The transformed cells were grown in 50 ml LB + 50 µg/ml chloramphenicol (Cam) + 100 µg/ml ampicillin (Amp; pHEN2 and pET22b+) or + 50 µg/ml kanamycin (Kan; for pET28a+) incubated for 16 hrs at 37 ºC and 225 rpm. The grown cultures were aliquoted into 1 L of LB with their respective antibiotics. This secondary culture was grown at 37 ºC and shaken at 225 rpm until OD_{600} reached 0.8 -1.2. Overexpression was induced by adding 1 mM IPTG and resuming growth at 16 °C for 16 hours with shaking at 225 rpm. Cells were harvested by centrifugation at 5000 g. The cell pellet (6-8 g/L) was resuspended in 5 mL of Lysis/binding buffer (50 mM Tris-Cl, 500 mM NaCl, 0.2% Tween-20, 5 % Glycerol, 2 mM PMSF, 5 mM β-mercaptoethanol (BME), 10 mM Imidazole) per every gram of cell pellet. The cells were then chemically lysed by adding 0.1 mg/mL of lysozyme, 1 mM PMSF, 1 mg/mL DNase, and 12.5 µg sodium deoxycholate per gram of cell pellet. The chemically lysing cells were incubated for 20 mins at 4 ºC. These were then physically lysed by sonication on ice at 10s on and 15s off intervals. Sonication took place at 60 % Amplitude for 25 cycles (13 cycles, 2-minute rest on ice, 12 cycles). The lysate was clarified by centrifugation at 30,000 g for 45 minutes. The supernatant was then filtered through a 0.45 µ filter.

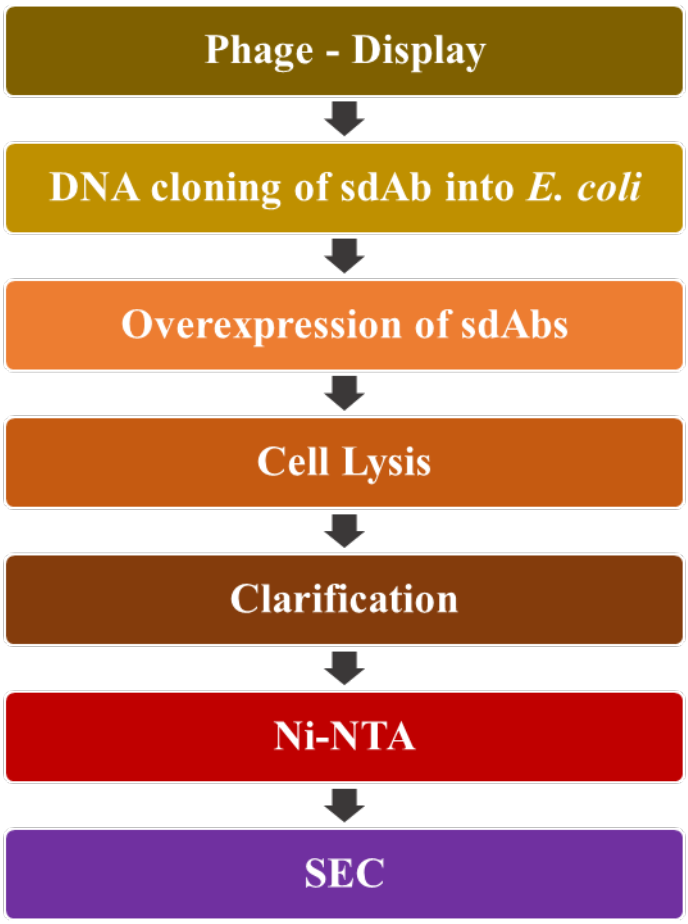

Figure 2.1. Simplified schematic of the strategy used for sdAb expression and purification. Once the G4-interacting sdAbs were identified, their gene sequence was introduced into bacterial vectors. *E. coli* cells were transformed with these vectors and used for the overexpression of the sdAbs. Access to the sdAbs and other cellular proteins was achieved by chemical and physical (sonication) lysis, and large cellular debris was separated by centrifugation. Lastly, the chromatography steps used to purify the sdAbs were employed by first using a nickel-affinity approach and then a size exclusion approach.

The sdAb was purified by immobilized metal affinity chromatography (IMAC) with a 1 mL HisTrap HP crude column (Cytiva ©) in an ÄKTA™ start system. The sample was loaded on a

pre-equilibrated (lysis buffer) column at a 1 mL/min flow rate. The lysis buffer and the elution buffer (50 mM Tris-Cl, 500 mM NaCl, 0.2 % Tween-20, 5 % Glycerol, 2 mM PMSF, 5 mM BME, 250 mM Imidazole) were used at different ratios in different steps of IMAC. The purification steps were divided into the binding step (sdAb binding to resin), wash step 1 (10 CVs of 4.2% elution buffer), then followed by wash step 2 (20CVs of 16% of elution buffer for sdAbs or 18% for sdAb-peptide tags), and gradient elution steps (20 CVs of 16-100% elution buffer). The collected fractions were analyzed through 15% SDS-PAGE, and fractions containing the sdAb were pooled and concentrated to 2-4 mg/mL. Tween-20 concentrations (0.2, 0.5, 0.75, 1, and 2 %) were also tested in the lysis and elution buffer for sdAb-peptide tag purifications to decrease non-specific binding. The concentrated sdAbs were lastly purified by size exclusion chromatography (SEC) with a Superdex 75 10/300 GL (Cytiva ©) column at a 0.45 ml/min flow rate in 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4) with 5 mM BME. A diagram of the proposed workflow from sdAb identification to SEC purification is seen in Figure 2.1.

2.2.3 Development of the quadruplexes and their preparation

A sequence previously confirmed to form quadruplexes [32] was used to study binding to the G4-forming region within the preCore promoter. BLASTN ® software was used to find potential 23-mers that may interact with the sdAb candidates, and the 23-mer preC G4 sequence was aligned. The Human RefSeqGene sequence database was searched, and the search algorithm was chosen to search for somewhat similar sequences (BLASTn). The outcome parameters searched for a maximum of 500 sequences. The following sequences were selected to study the sequence-specificity of sdAbs in the interaction studies: NG_007666.1 (Ras1), NG_008801.2

(CACNA1C), and NG_016564.2 (FAM9A). Ras1 was selected since it contains a G4 sequence within a promoter site known to interact with human Sp1 protein [79]. CACNA1C had the highest E-value and identity percentage (82.6%, 19 matching residues out of 23) to preC G4, and FAM9A was selected based on its lower identity percentage (65.2%, 15 matching residues out of 23). Based on these results, four 23-mer oligos were designed and confirmed to contain G4 structures with QGRS mapper software [80]. The oligos were commercially acquired from AlphaDNA (Montreal QC, Canada) as follows:

1. Sp1 preCore (preC G4): 5'-CTGGGAGGAGCTGGGGGAGGAGA-3'
2. Ras1: 5'-CGGGCGCGGGGCGGGGCGTGCGC-3'
3. CACNA1C: 5'-TTTAGAGGAGCTGGGGGAGGAGA-3'
4. FAM9a: 5'-CTGGAAGGAGCTCGGGGGCGAGG-3'

The folding of the oligomers into G4 structures was followed as per our last protocol [32], where 5 μM of oligomers were dissolved into the quadruplex buffer (20 mM HEPES, pH 7.5, 100 mM KCl, and 1 mM EDTA). Concentrations were estimated by calculating the extinction coefficients at 260 nm (Derived from IDT OligoAnalyzer tool) for Sp1 (237 400 M^{-1}-cm^{-1}), Ras1 (207 200 M-1-cm-1), CACNA1C (204 900 M-1-cm-1), and FAM9a (226 800 M^{-1}-cm^{-1}). 1 M KCl was added to the four oligomers in the G4 buffer and was heated to 95 ºC for 10 minutes. G4 formation occurred by passively cooling the G4s to room temperature. The G4s were concentrated to 500 μL using a three kDa cut-off Millipore's Amicon® Ultra-15 centrifugal filter units. The G4s were then purified by SEC using a superdex 75 10/300 GL (Cytiva ©) column at 0.45 ml/min in G4 buffer. The peaks observed in the chromatogram were pooled and analyzed by 6 % Urea PAGE and 6% Native PAGE. The samples that exhibited single bands in PAGE checks were used for subsequent experiments.

2.3 RESULTS

2.3.1 sdAb selection and validation

Phage display identified 90 VHHs binding to G-quadruplex WT after three rounds. Non-adsorbed Phage ELISA showed that of these, 22 were specific to G-quadruplex. Following sequencing analysis, 11 different sdAbs were identified, as observed in Fig 1.4. Of these, the top five were selected to be assessed for expression based on the high absorbance shown when interacting with WT-G4 versus Mut-G4 in phage ELISA.

2.3.2 Expression and Purification of sdAbs in pHEN2

S01, and S07-S11 showed single colonies growth in LB plates following transformation. However, S08 and S09 were not able to be grown in liquid LB or Terrific broth, so no further purification steps were followed for these two sdAbs. Overexpression of the recombinant sdAbs was successful for S01, S07, S10, and S11, as shown by SDS-PAGE following Ni-NTA affinity chromatography, where the bands around 20 kDa (shown in Fig 2.2) were the most prominent in contrast to others, including those shown in washing steps. Fig 2.2 shows a higher contrast in S10's band which is indicative of a higher yield than other sdAbs. S07 had the lowest yield based on SDS-PAGE band intensity. Interestingly, the sdAb elution peak resulted in an absorbance of ~900 mAU for sdAbs expressed in pET28a+ vectors, and ~100 mAU for those expressed in pHEN2 vectors.

The refining step of SEC allowed for further purification of the sdAbs, where the amount of protein visible larger than ~30 kDa in SDS-PAGE decreased when contrasted to Ni-NTA profiles (Fig 2.4D). The sharpness of S10's elution peak at 11.5 mL and the minimal quantity of other peaks throughout the remaining SEC chromatogram indicate the relative purity and quality

of the Ni-NTA purification (Fig 2.2A). Similar to S10 elution profile, S01 and BG4 had a single defined peak above 10 mL in its chromatogram was present. S11 showed the highest absorbance between 7.5 to 10 mL elution volume, but the fractions from volumes of less than 10 mL have little to no visible S11 content suggesting those peaks are mainly a product of the elution of unwanted protein (Fig 2.4E).

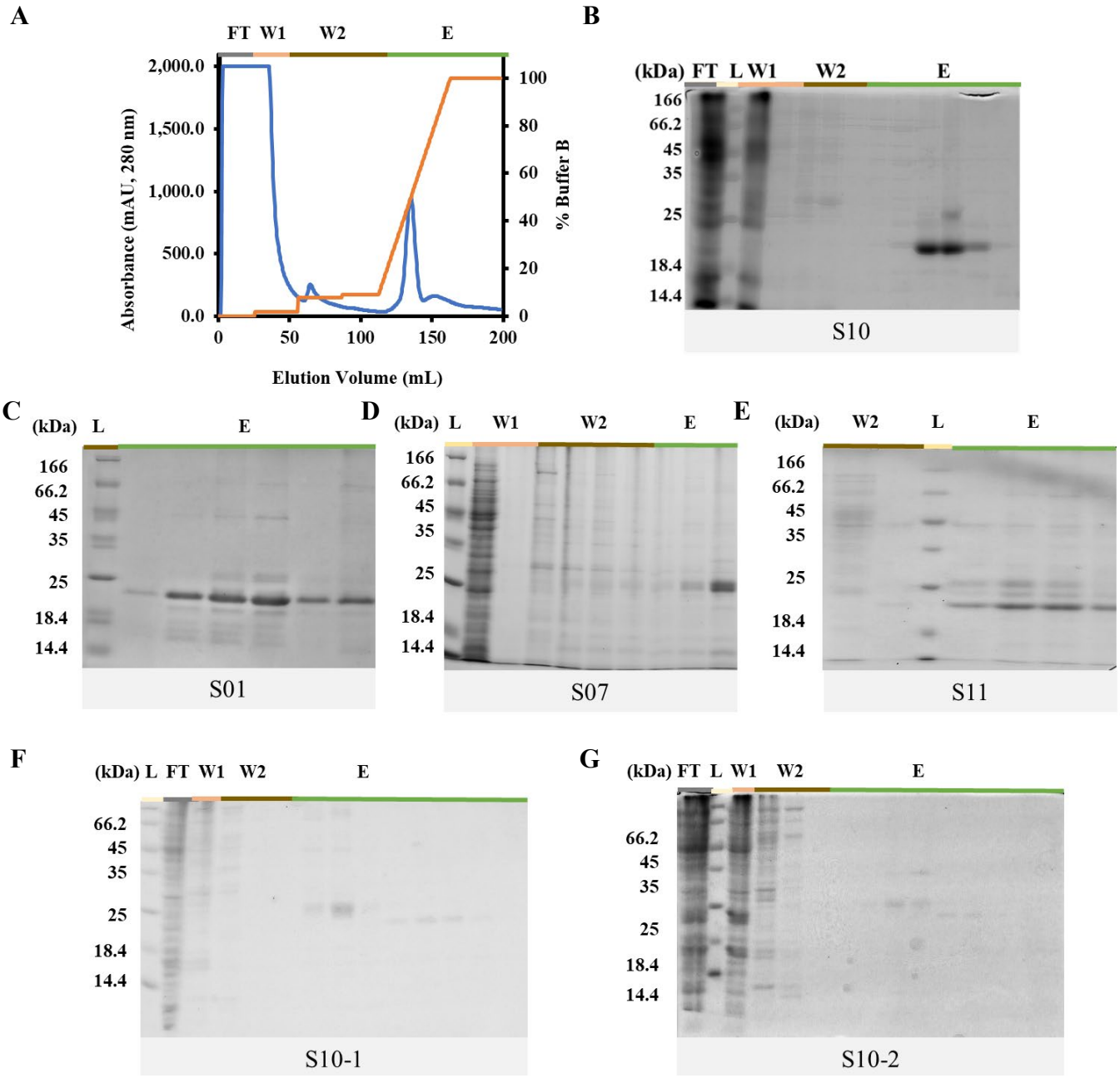

Figure 2.2. Elution profiles of sdAbs from Ni-NTA affinity chromatography purifications. (A) Representative example of a Ni-NTA chromatogram of an sdAb purification (S10 chromatogram shown). The flow through is represented as “FT”, the first wash “W1”, the second wash “W2”, and the elution is shown by “E”. The blue lined data shows the absorbance detected at 280 nm whereas the orange shows the incremental use of buffer “B” which indicates an increase in imidazole concentration, where 0% is 10 mM and 100% is 250 mM. (B-G) 15% SDS-PAGEs of sdAbs as observed following Coomassie staining. The same abbreviations and color coding are used as in (A). “L” represents the protein ladder. The presence of (B) S10, (C) S01, (D) S07, (E) S11, (F) S10-1, (G) and S10-2 can be confirmed by the presence of strong bands between 18.4 and 25 kDa.

A

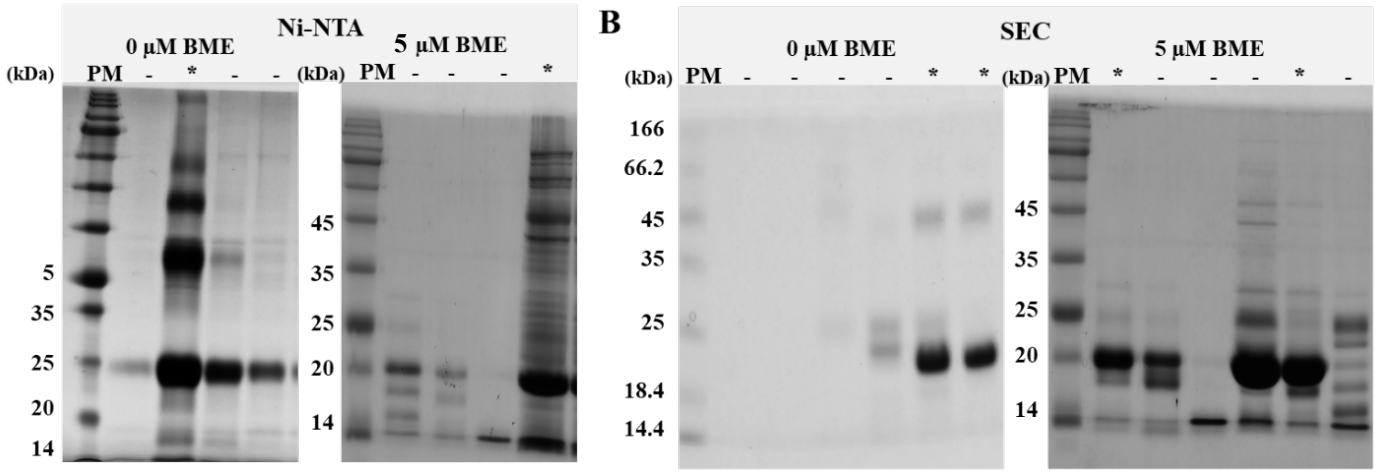

Figure 2.3. Chromatography fractions profiles showing oligomerization of S10 as seen in 15% SDS-PAGE when purified in A) Ni-NTA buffers and B) SEC buffers with and without BME. “*” represents the fractions found at the apex of the elution peak. The monomeric molecular weight according to the ladder is expected to be ~20 kDa and the dimer molecular weight is expected to be ~40 kDa.

The western blot analysis performed on S10 in pHEN2 confirmed the presence of the sdAb (Fig 2.4). Interestingly, it was observed that S10 requires fresh BME to purify monomeric sdAbs (Fig 2.3), for the lack of BME in chromatography buffers promotes the formation of dimerized S10 species. It is important to note that the observed dimerization was not able to be corrected by adding BME after the Ni-NTA purification, so this component must be present since the lysis step. The SDS-PAGE of concentrated sdAb post-SEC (Fig. 2.4G) shows a relatively faint presence of other analytes compared to the analyte of interest following storage at -20 ^{o}C. However, based on the band intensity of the sample, the main analyte present remains being S10 (Fig 2.4G).

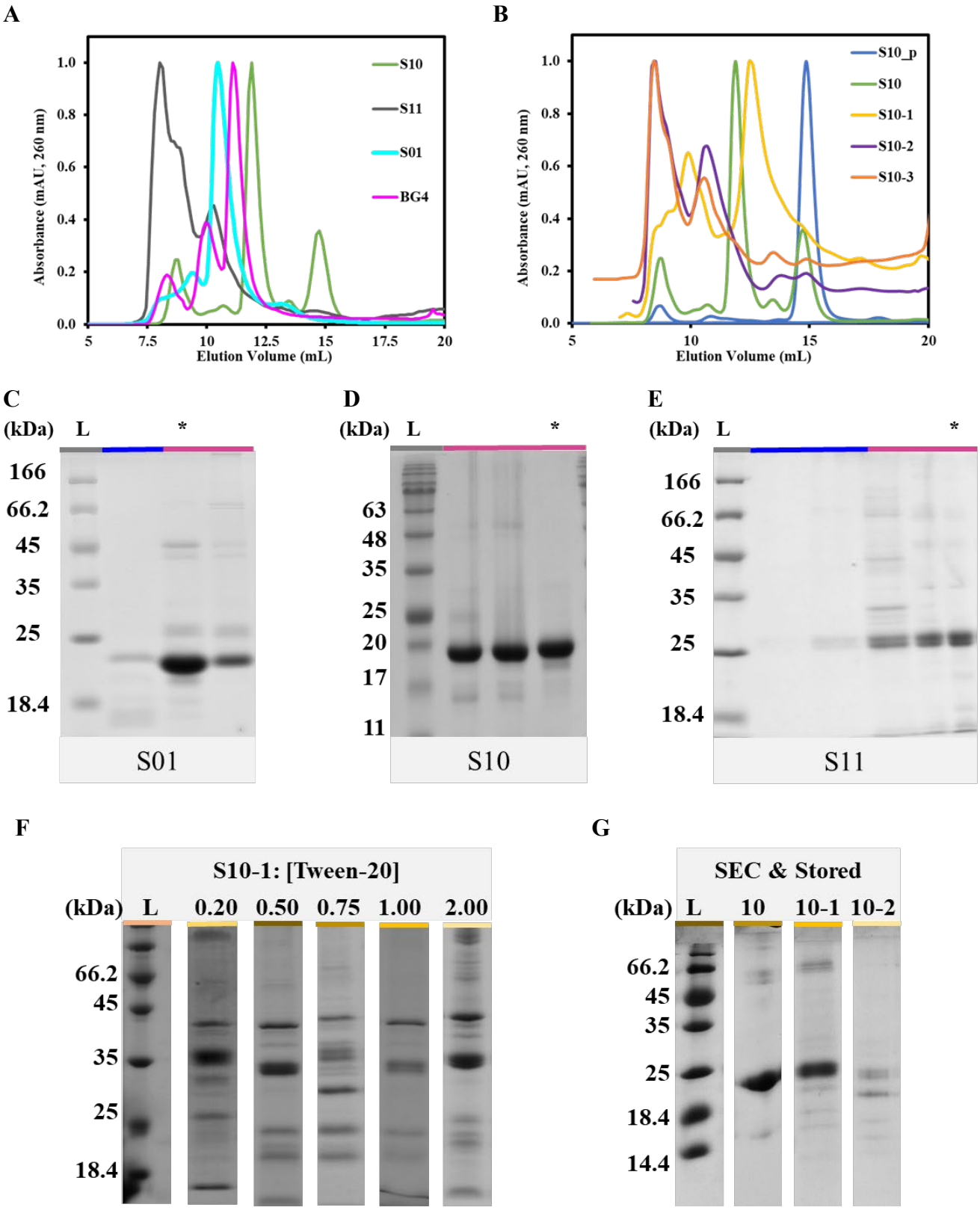
A
Absorbance (mAU, 260 nm)
Elution Volume (mL)
S10
S11
S01
BG4
B
Absorbance (mAU, 260 nm)
Elution Volume (mL)
S10_p
S10
S10-1
S10-2
S10-3
C
(kDa)
L
*
166
66.2
45
35
25
18.4
S01
D
(kDa)
L
*
63
48
35
25
20
17
11
S10
E
(kDa)
L
*
166
66.2
45
35
25
18.4
S11
F
S10-1: [Tween-20]
(kDa)
L
0.20
0.50
0.75
1.00
2.00
66.2
45
35
25
18.4
G
SEC & Stored
(kDa)
L
10
10-1
10-2
66.2
45
35
25
18.4
14.4

Figure 2.4. Size exclusion chromatography of sdAbs. (A-B) SEC chromatograms of sdAbs as seen from Superdex 75 10/300 GL, where the absorbance shown has been normalized to the maximum absorbance of each individual SEC sample. (A) Chromatograms of different types of antibodies. (B) Chromatograms of different modified versions of S10. (C-G) 15% SDS-PAGEs of sdAbs post- SEC. (C-E) The "*" represent the elution fractions chosen for MST and other experiments. Lanes labelled in blue represent fractions collected below 10 mL, those labelled pink represent the sdAb peak fractions. (C) S01. (D) S10. (E) S11. (F) Representative fractions of S10-1 following Ni-NTA chromatography after modifying the percent of Tween-20 concentration in the chromatography buffers. The numbers above represent each concentration tested (from 0.2% - 2.0%). (G) S10, S10-1, and S10-2 following a week-long storage at -20 ^{o}C. (F-G) The complete gels of each fraction are shown as supplemental material in appendix A.

2.3.3 Expression and Purification of Peptide-Tagged S10

S10-1, S10-2 and S10-3 were also successfully purified and overexpressed (Fig 2.3G) similar to S10. S10-1 and S10-2 had similar yields following Ni-NTA purification as seen in SDS-PAGEs (Fig 2.2F-G). However, in preparation for SEC, a greater amount of lost S10-2 was observed in comparison to S10-1. As observed in Fig 2.3G, this trend was also observable following SEC and storage in -20 ^{o}C, where S10-2 shows fainter sdAb bands than S10-1 regardless of being stored at comparable concentrations. S10-1, S10-2, and S10-3 displayed a greater amount of sample eluted between 7.5-10 mL in comparison to S10 (Fig 2.3B), but just like S11, these fractions contained little to no sdAb.

Following Ni-NTA, S10-1 specifically depicted a higher presence of nonspecific binders. Figure 2.3G shows that the purification obtained at 0.2 % tween-20 contained more non-specific binders

after Ni-NTA than S10 did in the same chromatography step. At other concentrations of tween-20, such as 0.5% and 1.0%, the non-specific binders decreased. However, purifications using 1.0 % Tween-20 resulted in the aggregation or loss of the S10-1 sample following a couple of hours post-purification. Thereby, 0.5% Tween-20 demonstrated the most success at decreasing the number of non-specific binders in the elution of S10-1. Lastly, the identity of S10-1 was confirmed alongside S10 through western blots (Fig 2.4). Interestingly, a small amount of degradation can be observed in the S10-1 fraction that is consistent with the size of unmodified S10.

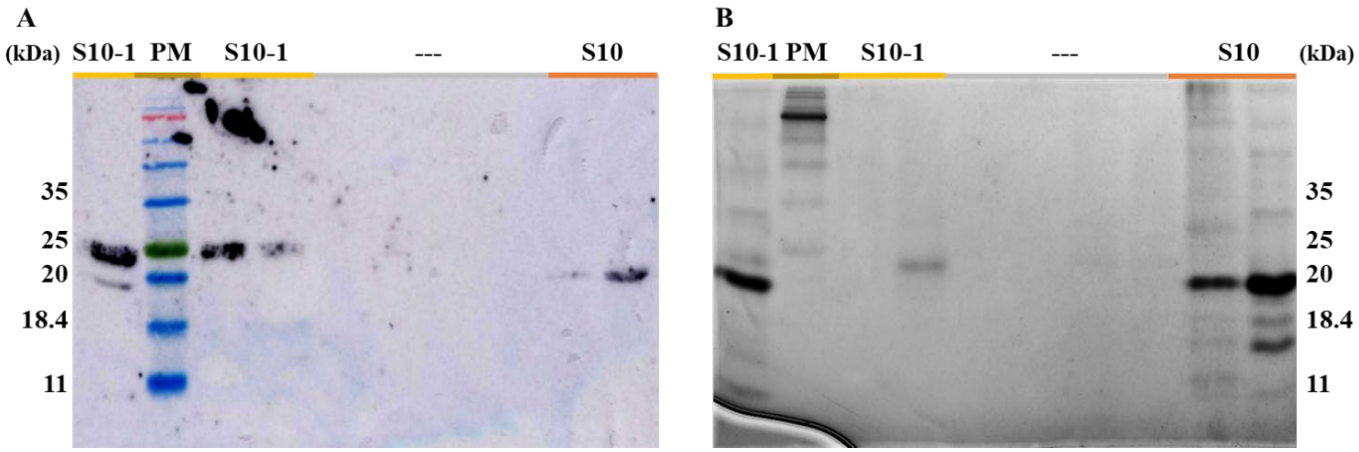

Figure 2.5 Western Blot of S10 and S10-1 elution fractions as resolved from a 15% SDS-PAGE. A) Final image resolved from western blot, where the black bands represent presence of His-tagged protein. B) Ponceau stain displaying all protein content in the nitrocellulose paper before blocking.

2.3.4 Purification and Folding of G4s

As expected from commercially-acquired oligos, the purity levels of G4 analytes showed that a single species was observed in a native and a denaturing PAGE (Fig. 2.6 B-C). The

chromatograms of each oligo (Fig 2.6A) showed the presence of well-formed single peaks in each analyte, apart from FAM9a. Both peaks shown in the SEC chromatogram of FAM9a were individually analyzed in CD spec.

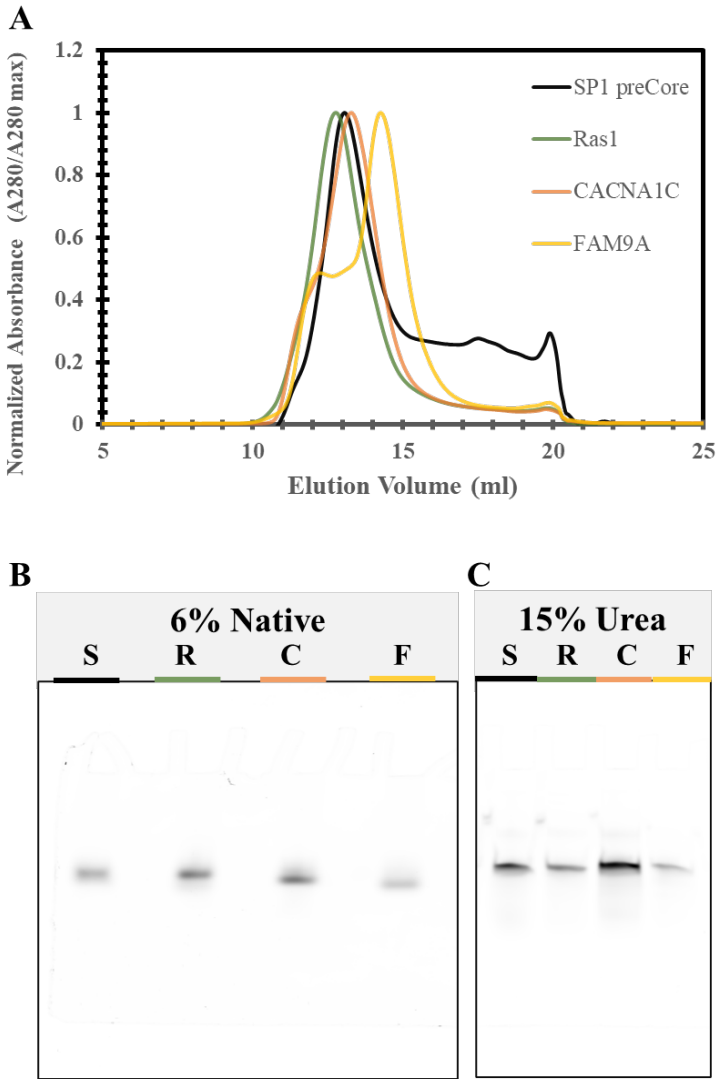

Figure 2.6. G4-forming sequences following G4-folding protocol and SEC. (A) SEC chromatogram profiles of four nucleic acids following G4 folding. The four oligos were visualized in PAGEs to observe their homogeneity in (B) native conditions (6% acrylamide) and in (C) denaturing conditions (8M Urea, 15% acrylamide).

2.4 DISCUSSION

Other groups have developed purification protocols that successfully purified recombinant sdAbs from *E. coli* [81, 82]. However, many of these research groups don't fully elaborate on their production methods [83, 84]. Other groups have included production approaches such as buffer exchanging sdAbs into different pHs [85], or that rely on refolding after denaturing protocols [81, 86], adopt difficult to access in-house pipelines [87], or don't test the sdAb binding ability post-purification [81]. An expression and purification protocol that does not rely on the destabilization of contaminants or unfolding of the sdAbs can facilitate its implementation to other sdAbs without additional activity and structural characterization.

In this work, we identified 11 different sdAbs that exhibited some level of interaction and specificity for preC G4 utilizing phage display technology. ELISAs revealed that some of these sdAbs had some affinity to the mutant G4, and all sdAbs varied in their affinity to preC G4. Within the primary binding affinity results from ELISA, two candidates, S10 and S11 were predicted to perform better than other sdAbs with lower preC G4 binding affinity or with higher mutant pre C G4 affinity. Low target specificity and unintended off-target effects that may initiate patient complications is one of the main reasons for therapeutic failure [88-90]. The ELISA results guided the set up of binding studies (Ch. 3) where S01 was used as a negative control due to its low affinity too preC G4 and where the binding to off-target sequences and quadruplexes was crucial to be determined.

As has been widely studied before, sdAbs conserve a large portion of their sequence apart from their CDRs [91, 92]. The successfully Ni-NTA purified sdAbs share similar elution profiles where comparable levels of non-specific binders are evident (Fig 2.2). The purification approach taken is a similar chromatography pipeline as other sdAbs involved in therapeutic candidacy

[53]. Ni-NTA purification took place in Tris-base, and then the buffer was exchanged into a 1x PBS buffer at the SEC step. This approach also worked for peptide-tagged sdAbs (such as S10-1 and S10-2). However, S10-1 and S10-2, the sdAbs with the most peptide sequences tagging the sdAb, did show the need for a higher imidazole concentration in the wash and a higher tween-20 concentration overall. In a similar manner as HBV enters the hepatocyte by NTCP interactions [93], S10-1 and S10-2 have NTCP interacting peptide sequences as well as CPPs and NLS sequences. It is thought that some of these sequences may be responsible for the association of non-specific binders to the sdAbs resulting in more stringent purification methods. The addition of tween-20, as seen in other Ni-NTA technologies [94], effectively decreases the non-specific binding in Ni-NTA, which has been shown to limit the amount of protein-protein interfaces available for interactions to occur [95]. Lastly, it is suspected that in the absence of BME, sdAbs may form dimers, as observed in Fig 2.5. S10 contains two residues that may be able to participate in disulfide bridge formation [96]. Including BME in buffers to formulate reducing conditions resulted in a mainly monomeric sample.

Previous work has determined the importance of the G4 structure in the SP1 pre-core region for the HBV viral cycle [32]. Thereby, this region was selected as the therapeutic target for developing recombinant sdAbs. However, as previously mentioned off-target effects are often associated with detrimental patient health responses. Thereby, the production and testing of other G4s can allow for future assessment of potential cross-reaction effects that the sdAb candidates may have. Thereby, Ras1, CACNA1C, and FAM9a sequences were chosen to undergo G4 folding. The three different sequences are naturally occurring in the human genome and thereby serve as appropriate off-targets. As expected, the previously published G4-folding protocol [32] provided seemingly homogenous G4 species as gathered from the urea and native PAGES in Fig

2.6. Though the chromatogram of FAM9a varies the most in its elution profile from other G4s, its PAGE profile remains uniform and comparable to other G4s. Though not surprising based on the % query coverage preC G4 has on Fam9A (65%), the shoulder and apex of the dilution profile peaks should undergo quality checks, such as CD spec [97-100], to better understand its structural properties.

2.5 CONCLUSION

In conclusion, the quadruplex-folding and purification protocols, as previously described [32], produced homogeneous samples as observed in native and urea PAGEs. The resulting G4s will be tested against sdAbs to assess their specificity and potential cross-reactivity in the following chapter. Additionally, the approach taken for expression and purification has resulted in an easily reproducible protocol that provides sdAbs with a high purity while maintaining stable conditions and preventing unfolding. As observed, the protocol can easily be extrapolated to express and purify sdAbs other than S10, as only minor adjustments may be needed when additional peptide tags are used (S10-1 and S10-2). These additional adjustments are needed in peptide-tagged protein purification to decrease the non-specific binding observed in the elution fractions. It is hypothesized that this increase in non-specificity observed in Ni-NTA affinity chromatography of tagged proteins may be a product of interaction between the sdAb-tag and the bacterial proteins that may share a high level of homology to human proteins. However, I have shown that increasing surfactant concentration, in this case Tween-20, decreases the presence of non-specific binders in the elution fractions. Lastly, this protocol showed consistent results across many sdAbs and can be used in the future for the expression and purification of other sdAbs being studied for therapeutic development.

CHAPTER 3: BIOCHEMICAL CHARACTERIZATION OF sdAb AND G4 INTERACTIONS

3.1 INTRODUCTION

The movement towards gene therapeutics has increased in the past decades, but challenges involving off-target effects still exist today, as is the case for CRISPR [101], RNAi [102], and other anti-sense technologies [103]. Off-target effects can result in abnormal phenotype expression [104] as a result of genomic instability caused by mutation [105] or other changes in gene expression [104]. Thereby, there is a need for other non-gene editing avenues that can inhibit gene expression to avoid mutation. However, off-target gene expression changes can still occur, thereby target specificity testing is a crucial step in therapeutic development.

The affinity of a molecule (ligand) to its target can be measured by calculating the dissociation constant (K_D) [106]. This measure directly represents the ligand concentration required to bind to 50% of the target population [107]. Thereby, a stronger ligand-target interaction results in a lower K_D, for a lower ligand concentration is required to interact with 50% of the target population. Through this approach, the K_D value can aid in discriminating therapeutic candidates based on their ability to have a strong interaction with their intended target and a weak or no interactions with potential off targets.

There are multiple techniques that have been developed to estimate the affinity constant value for drug-target interactions [106]. Microscale Thermophoresis (MST) is often used to determine the binding affinity of a molecule to a target. MST relies on changes in Brownian motion in solution due to the formation of complexes or lack thereof [108]. One of the analytes is fluorescently labelled and kept at a single concentration throughout sixteen 2-fold dilutions of the fluorescently unlabeled analyte, as seen in Fig 3.1 [109]. The apparatus is equipped with an

infrared laser that introduces thermal changes in the sample while simultaneously exciting the fluorescently labelled molecule to track any motion changes [32]. In this manner, changes in Brownian motion are indirectly plotted in a titration curve where one can determine the binding constant of the interaction [109]. MST allows testing of molecules in a wide range of buffer conditions and temperatures. It requires a relatively low sample concentration and can study dissociation events in the range of µM to pM [32]. These assets make it a suitable technique for studying therapeutic candidates' potential interactions.

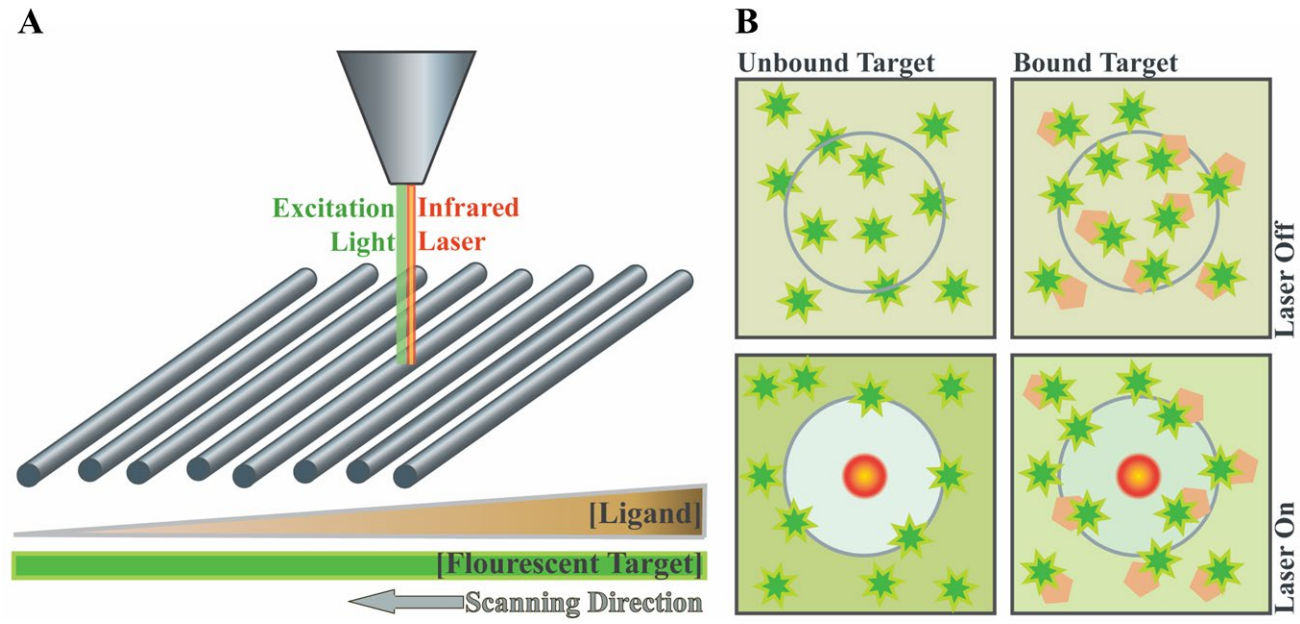

Figure 3.1. Technical set up and Fluorescence changes in Microscale Thermophoresis. A) The sixteen capillaries (8 shown) have a concentration of fluorescently labelled target analytes, which are titrated in 2-fold dilutions of ligand. The apparatus scans the capillaries from the sample with the highest to the lowest ligand concentration. The MST machine is equipped with an excitation light to observe changes in fluorescence detected. It also has an infrared laser that causes sudden temperature changes and introduces kinetic energy into the capillary system. B) Diagram showing two samples within a capillary when the target is bound and unbound. The circle in the

middle indicates the area within the range of fluorescence detection. The green stars represent the labelled target analyte, while the beige pentagons represent the unlabeled ligands. The bottom red circles represent the lasers. Both unbound and bound targets show the same amount of fluorescence. Once the infrared laser is on, the analytes migrate away from the detection area and the infrared laser. Due to the size of the complex formed in the bound target group, the overall fluorescence detected is higher than that from the unbound target.

The preC G4 region, necessary for core protein production [32], can be fluorescently labelled to serve as a target for MST and help to determine differences in binding strength amongst different therapeutic candidates. In this manner, the most potent binding candidate can offer the most extended permanence time and thereby form a stable complex with preC G4 that can interrupt transcription of core protein indefinitely. Currently, other sdAbs therapeutic candidates have had their binding kinetics characterized (affinity at the nanomolar range) and used to predict their therapeutic success [53, 110, 111]. Though these studies support the potential of sdAbs to be strong therapeutic contenders, their off-target effects are not commonly studied at the biochemical level and targeting nucleic acids for therapeutic purposes has not been shown. The presence and importance of G4s in humans have been extensively researched [112, 113], and the ability of sdAbs to bind non-specifically to G4s has been demonstrated before [62]. The study of potential off-target effects in preC G4 binding therapeutics is critical for the success of therapeutic development.

One of the main obstacles in achieving an HBV functional or sterilizing cure is the difficulty of reaching the cccDNA within the nucleus [114-116]. There are a variety of therapeutic delivery strategies that exist today [117-119]. Interestingly, sdAbs are already known to pass through cell membranes easily, so adopting tissue-specific and nuclear-penetrating abilities can improve the

drug's effectiveness when tested at the cellular level and beyond. The adoption of fused peptide tags for cellular penetration remains a simple way to acquire tissue and organelle specificity without a need to include additional molecules for encapsulation [119, 120]. However, adding peptides or domains to a protein may result in steric hindrances that can affect the protein's function [121]. Therefore, studying whether these modifications impact sdAb function can provide insight into future therapeutic delivery approaches.

This study tests various sdAb candidates against preC G4 to determine which sdAb is the strongest binding candidate. Additionally, the nature of the specificity of the best sdAb candidate is studied to determine if affinity to G4s is specific to the sequence and structure of preC G4. Furthermore, the tissue-targeting strategy of using peptide tags on sdAbs will be assessed to observe whether adding these tags affects sdAb function or affinity to preC G4. Overall, this work will provide insight into the feasibility of using sdAbs as G4-targeting therapeutics.

3.2 METHODS

3.2.1 Microscale Thermophoresis (MST) Interaction Set-up

Binding studies were performed in MST to assess the binding strength and specificity of the sdAb to the target G4 and other G4s. Commercially available, generic G-4 binding BG4 scFv [113, 122] and sdAbs tested were expressed and purified as mentioned in Ch. 2. As indicated in previous work [32, 109], Fluorescein isothiocyanate-labelled G4s were maintained constant at 200 nM in MST G4 buffer (G4 buffer + 0.1 % Tween-20). However, to study the interaction of S10 with non-G4 preC oligo (preC G4*), the oligo underwent folding and purification, as noted before but using LiCl instead of KCl as a salt. Similarly, the oligo was maintained at 200nM in MST G4* buffer for this interaction using LiCl instead of KCl. The sdAbs and BG4 scFv were

diluted in a 2-fold manner throughout 16 capillaries where the highest concentration of the binding molecule was at least 12 μM. The titrations were run in the Monolith NT. 115 with standard capillaries (Nanotemper Technologies, San Francisco, CA). The G4 and binding molecule titrations were incubated at room temperature for 16 hours. The excitation settings were set at medium but occasionally high to increase signal: noise ratio in cases non-bound and bound fluorescence had high noise. The laser power was set to be adjusted automatically which the software consistenyly determined it to be 20% or 40% power.

3.2.2 Analysis of MST Data

Each interaction was analyzed in triplicate with MO Affinity Analysis software v2.1.3 (Nanotemper Technologies, San Francisco, CA). This software allowed the raw MST fluorescence data to be plotted to determine a dissociation constant (KD) for every analyte pairing. The MST on-times used for KD determination were chosen to be below 5.0s, as recommended by the provider, to avoid potential structural destabilization of the analytes due to localized increases in temperature. The chosen range also was determined by choosing a range in which the model's chi-square was the lowest possible. The binding curves acquired from MO Affinity Analysis software were set to "Fraction Bound" to compare interactions against other ligands or targets with different fluorescent magnitudes and changes. The binding affinity (K_D) was calculated by the software using eq 1, where F(c) is the fraction bound at a ligand (sdAb) concentration c, unbound is the F_{norm} signal of the isolated target (DNA oligo), bound is the F_{norm} signal of the complex, and the target is the final concentration of the target in the specific assay.

Eq. 1: $F(c) = \text{Unbound} + (\text{Bound} - \text{Unbound})\ (c + c_{target} + K_D - ((c + c_{target} + K_D)^2)^{1/2} - 4cc_{target})\ (2c_{target})^{-1}$

3.2.3 SEC Complex Analysis

Based on the initial interaction studies testing S10 and S11, the strongest binding sdAb was chosen to confirm the strength of the interaction by subjecting the analytes to SEC. Following the expression and purification of both analytes, as previously explained, S10 was incubated overnight with preC G4 for 12 hours at a 1:2 concentration ratio in 50% 1x PBS and 50% G4 buffer. The samples were then concentrated to 4 mg/mL and loaded into Superdex 75 10/300 GL (Cytiva ©) column at 0.45ml/min in 1x PBS. The presence of a complex was confirmed in two manners. First, the complex was to demonstrate the ability to elute earlier than S10 and preC G4 alone. Second, the peaks were run in 15 % SDS-PAGE for 60 minutes at 200 V, where the resulting gel was first stained with SYBR-safe for 10 minutes and pictured with Amersham Typhoon Gel and Blot Imaging System under settings “Cy2,” which excites at 489 nm and emission is read at 506 nm. The gel was then dehydrated with 50% ethanol + 10% acetic acid and rehydrated and stained with 5% ethanol, 7.5% acetic acid and Coomassie G-250. The stained gel was imaged with Amersham Typhoon Gel and Blot Imaging System using the “densitometry” tools.

3.2.4 Selection Criteria of sdAbs and Peptide Tag Fusion

The selection of the sdAb most suitable for the next phase of therapeutic testing was as follows. In the first round of interaction studies, the therapeutic candidacy of each sdAb tested was revoked based on poor binding affinity to preC G4. Next, SEC was used to observe the physical manifestation of complex formation as well as the stability of the complex. Then, the next round focused on testing potential changes in the interaction strength of peptide-tagged sdAb candidates. S10-1 and S10-2 were created as shown in Table 3.3 to include tissue-specific tags

(CPP, CPP2, CPP3) and nuclear localization signals (NLS). Subsequently, S10-3 was created with mainly an NLS sequence to reduce potential functional changes that are products of the addition of multiple peptide tags. All the binding studies with peptide tagged sdAbs were executed as mentioned in section 3.2.1 and 3.2.2.

3.3 RESULTS

3.3.1 Interactions of S10 with Nucleic Acids

All the MST assays showed no evidence of photobleaching or aggregation and passed other quality checks presented by MO control software. All of the interaction studies (Table 3.1, 3.2) have signal-to-noise (S: N) ratios above the threshold (5) considered by Nanotemper to provide desirable data indicative of interactions occurring, and apart from CACNA1C's assay, all fall above the S: N ratio (12) considered to be indicative of an excellent assay indicative of interactions. The positive control, scFv-BG4, interacted with a KD of 5.671 ± 0.161 μM when interacting with the target quadruplex, preC G4. The known weakest binding sdAb, S01, showed no binding activity with preC G4, and thereby no K_D or binding curve was modelled or calculated by the software.

As observed in Fig. 3.2, of the two sdAbs tested against preC G4, S10 had the highest affinity (~20-fold stronger) in MST. S10 was tested against other human G4s, and its interaction with the viral G4 is at least 10x stronger than its interaction with the human G4s tested. As a summary, the strength in the binding of S10 to different sequences of G4s can be summarized as follows: Sp1 > CACNA1C > Ras1 > FAM9a (Strongest interaction to weakest). Interestingly, the MST traces from the S10 – preC G4* interaction with LiCl G4* buffer showed a difference in dissociation constant of ~80-fold. This result suggested S10's interaction preference for G4-

folded preC G4. The dissociation constants (K_D) of FAM9a and Ras1 were estimated by Nanotemper's Affinity Analysis software with data sets that did not reach saturation during their interaction studies. Following 1 week, 1 month and 6 months of -20 ºC storage, S10 was able to maintain its performance and interact with preC G4 within a nanomolar range (Table 3.2).

SEC was performed to investigate the nature of the stability in S10-preC G4 interaction. Consistent with SEC elutions in respective to size differences, the analytes eluted from biggest to smallest as a complex (10.5 mL), S10 (11.5 mL), and preC G4 (13.5 mL) as shown in Fig 3.3A-B. As expected, only two peaks were observed in the complex profile since the S10 to preC G4 concentration was in a 1: 2 ratio. Superimposed SDS-PAGE images (Fig 3.3C) confirmed that the elution at 10.5 mL from the complex profile contained S10 (~20 kDa, Coomassie stain) and preC G4 (~7 kDa, SYBR safe).

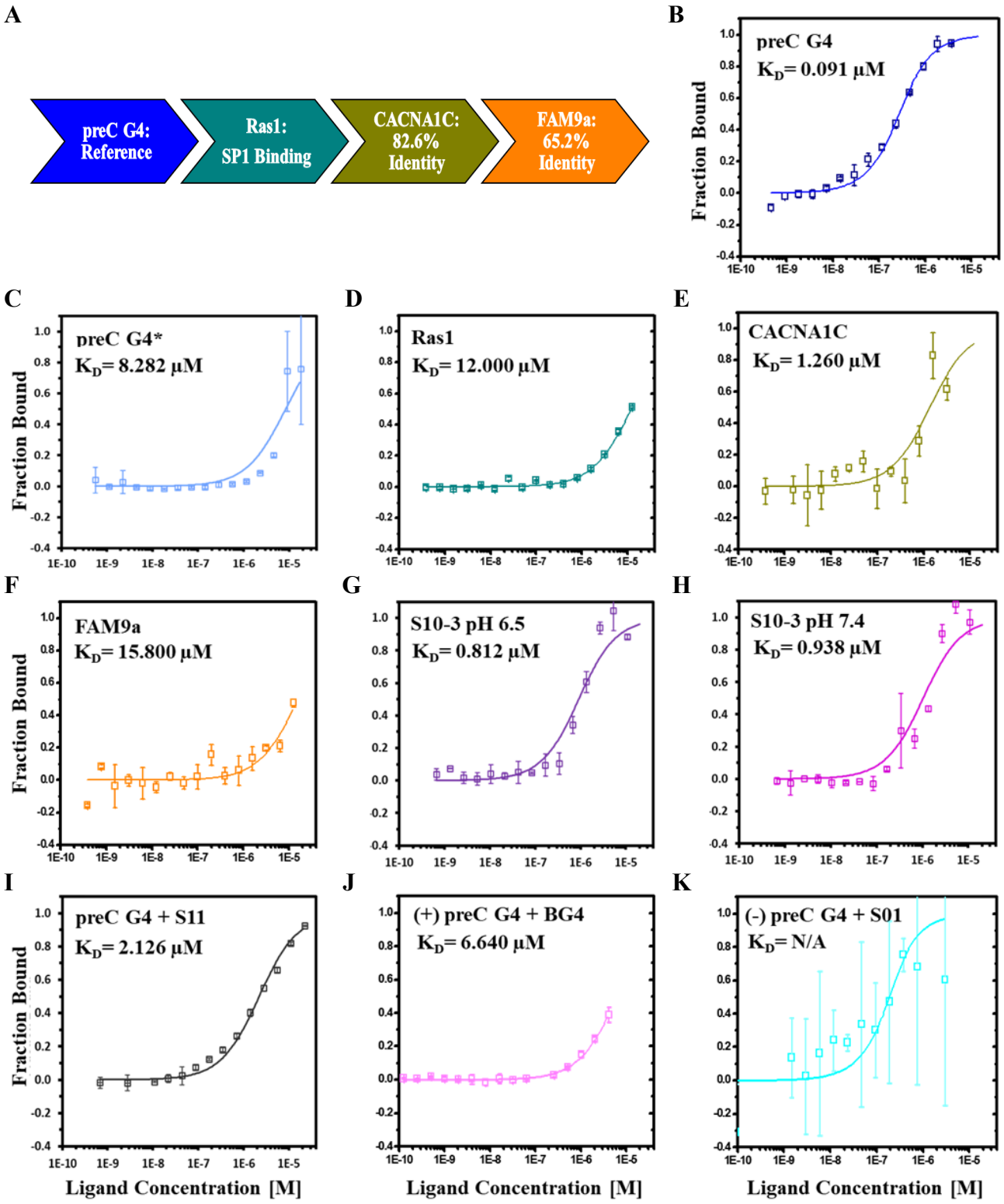

Figure 3.2. MST Interaction studies of sdAbs to G4s. A) The four G4-forming nucleic acid oligos alongside their criteria for being chosen to be tested in the sdAb sequence-specificity

assays. B-F) Interaction studies of S10 where it is tested against (B) preC G4, (C) the preC G4* (D) Ras 1 G4, (E) CACNA1C G4, and (F) FAM9a G4. Note that preC G4* was performed in LiCl buffer to avoid G4 formation. (G-H) Binding interactions of S10-3 with preC G4 in pH (G) 6.5 and (H) 7.4. (I) Binding interaction of S11 against preC G4. The positive control (J) and negative control (K) are shown. All the y-axis of the plots represents the fraction bound, and the x-axis show the Ligand concentration. (n=3).

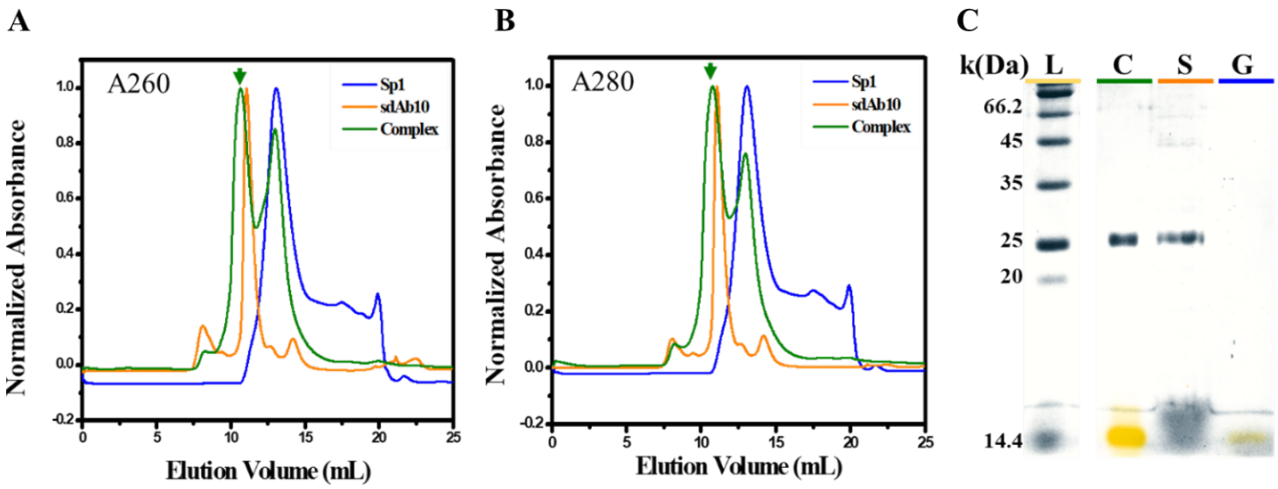

Figure 3.3. The size exclusion chromatography elution profiles of S10, preC G4, and S10-preC G4 as observed under absorbance at 260 nm (A) and 280 nm (B). The G4 (Sp1) has an elution peak of 13.5mL, S 10 has an elution peak of 11.5 mL, and the complex has an elution peak of 10.5 mL. C) Two overlayed images of a single SDS-PAGE stained with SYBR safe (yellow) and Coomassie stain (grey). L represents the ladder, C represents the complex, S represents the sdAb by itself, and G represents the G4 (Sp1).

Table 3.1. Binding affinity of the interactions between quadruplex binding candidates and DNA oligos as determined by MST. "*" represents a DNA oligo which did not undergo the G4 folding protocol. (n=3). Note that two S10-Sp1 preCore interactions are noted (one contrasted to LiCl-treated oligo and the other to different sequences of oligos) due to their batch differences in affinity. Each color to the right of the target matches its respective MST binding curve in Figures X,X.

;and	Target	Kd (uM)	Kd Confidence (uM)	Response Amplitude	Std. Error of Regression	Reduced χ^2	Signal:Noise	Ligand Starting Concentration (uM)	Excitation Power	MST Power	Hot Region (s)
10	Sp1_preCore	0.091	0.019	4.3	0.27	3.71	16.5	3.53	20	Medium	0.5-1.5
	Ras1	12.000	3.020	64	1.22	32.1	56.38	12.90	20	Medium	0.5-1.5
	CACNA1C	1.260	0.352	1.6	0.19	3.10	8.32	3.22	20	High	0.5-1.5
	Fam9a	15.800	19.700	5.5	0.40	24.4	15	12.90	20	Medium	0.5-1.5
	3' UTR Zika	57.400	3.490	56	1.03	6.73	55.8	29.90	20	Medium	0.5-1.5
S10	Sp1_preCore*	8.282	0.000	18	1.52	72.54	11.92	0.00	60	Medium	0.5-1.6
11	Sp1_preCore	2.200	0.148	26.76	0.85	1.73	31.72	22.50	20	Medium	0.5-1.5
1 (-)	Sp1_preCore	-	0.000	1	-	-	2.34	3.07	40	Medium	0.5-1.5
;cFv (+)	Sp1_preCore	6.640	0.189	31	0.29	1.68	109.00	4.16	40	High	0.5-1.5
0-1	Sp1_preCore	0.258	0.042	80	4.99	50.50	16.00	9.50	40	High	1.5-2.5
	Ras1	0.046	0.007	144	5.13	8.26	28.00	8.85	40	High	1.5-2.5
	CACNA1C	0.271	0.048	39	2.66	2.30	15.00	17.40	40	High	1.5-2.5
	Fam9a	1.570	0.233	35	3.32	208.60	10.00	1.74	40	Medium	1.5-2.5
11	Sp1_preCore	2.200	0.148	26.76	0.85	1.73	31.72	22.50	20	Medium	0.5-1.5
S10-3	Sp1_preCore (pH 6.5)	0.800	0.000	15.1	1.26	5.90	12.63	10.79	20	Medium	1.5-2.5
S10-3	Sp1_preCore (pH 7.4)	1.026	0.000	13.8	1.24	6.63	11.64	10.79	20	Medium	1.5-2.5

Table 3.2 Changes in binding affinity of S10 to the G4 in Sp1 preCore following different periods of time in storage at -20°C.

ɒrage Time	Kd (uM)	Kd Confidence (uM)	Response Amplitude	Std. Error of Regression	Reduced χ^2	Signal:Noise	Ligand Starting Concentration (uM)	Excitation Power	MST Power	Hot Region (s)
Day 1	0.091	0.019	4.3	0.27	3.71	16.5	3.53	20	Medium	0.5-1.5
1 Week	0.022	0.011	3.22	0.18	1.06	19.1	3.78	20	Medium	0.5-1.5
1 Month	0.027	0.013	3	0.25	1.92	11.9	2.90	20	Medium	0.5-1.5
6 Month	0.281	0.081	2.9	0.33	1.22	9.07	150000.00	20	Medium	0.5-1.5

3.3.2 Effects of Peptide Chain Additions in S10-preC G4 Interactions

Table 3.3. Design of the peptide tagged sdAbs. The peptide tag design planning for every sdAb considered is noted. S10-4 and S10-5 were only studied in the context of Ch. 4. "L" represents a linker chain, and "RL" represents a rigid linker chain. "CPP, CPP2, CPP3" represent three different cell-penetrating peptide sequences. "NLS" denotes the nuclear localization signal sequence. "GFP" represents a green fluorescent protein. MYC, HA, 3xFLAG, and 6xHis tags are also in some modified S10s.

sdAb	Sequence
S10	**sdAb - L - 6xHis – MYC**
S10-1	CPP – L - **sdAb - L** – CCP2 – RL – CCP3 - L - *NLS* - ~~TCS~~ - **6xHis**
S10-2	**sdAb** – L - CPP- L – CCP2 – RL – CCP3 – L – *NLS* - ~~TCS~~ - **6xHis**
S10-3	**HA** – L – *NLS* – L – **sdAb - 6xHis**
S10-4	**sdAb - 6xHis - 3xFLAG**
S10-5	**sdAb** - L - *NLS* - **GFP** - L - **3xFLAG**

The interaction studies looking to observe potential binding affinity changes due to peptide chain additions (Table 3.3) showed that S10's function might be affected depending on the peptide chain's physical and biochemical properties. S10-2, with peptide chain additions in the C-terminal of the protein, showed no binding ability. S10-1 showed a similar pattern to S10 in terms of specificity, although weaker binding generally. However, upon SDS-PAGE quality check analysis, the size of S10-1 resembled that of S10 and showed a small number of bands below 15 kDa was also observed. Furthermore, S10-1 interaction studies performed with fresh protein under the same conditions as S10 did not exhibit reproducible binding interactions. Interestingly, S10-3 presented the binding interactions with preC G4 to be the most successful in terms of peptide additions and the most comparable with S10 (Fig 3.2).

3.4 DISCUSSION

Traditional *in vivo* methods of drug screening tend to be costly and time-consuming. Standard drug permeability studies typically take 21 days to complete with frequent media changes, which quickly elevate the cost of the experiments [123]. Though this approach provides valuable insight into drug screens that could be extrapolated to *in vivo* studies, biochemical interaction studies can reduce the overall cost and time screens take. Previous work [53] has shown the advantage of using biochemical approaches such as MST and Bio-Layer Interferometry (BLI) to screen a wide selection of drugs for the best performers. However, the potential of these technologies can be further exploited to understand potential off-target interactions and to screen for changes in function following peptide fusion.

Due to their relevance in cancer and infectious diseases, a variety of small molecules, peptides, and other ligands are being developed to target specific G4s [124, 125]. One of the therapeutic peptides, α,ε-poly-l-lysine (α,ε-PLL), has been shown to have a K_D of 0.56±0.14 μM for Pu22 and 1.0 ±0.20 μM for Tel22 (a c-myc oncogene and a telomeric sequence, respectively) through surface plasmon resonance [126]. Truncated versions of the LL37 G4-binding domain resulted in a binding affinity (KD) for Pu27 G4 ranging from 8-60 μM [127]. Other ligands [127, 128] have also shown binding affinity with a KD of higher than 800 nM, with only a few having affinity within similar ranges as our own sdAb. However, work demonstrating the sequence and shape specificity of other therapeutic G4 candidates is limited.

Previous work in drug discovery has detected sdAb candidates able to perform binding interactions with their respective targets with a binding strength within the nanomolar range and lower [53, 84, 110, 129], indicating that sdAbs may be a stronger-interacting alternative. The binding strength of S11 to preC G4 showed an interaction ~3x stronger than that of scFv BG4. In contrast to S11, S10 shows a ~73-fold difference in binding strength and is also ~20-fold stronger than S11. Not only did S10 perform notably better than S11, but as expected from other sdAbs, S10 showed a stronger interaction than other G4-targeting therapeutic candidates.

The development of SG4 has already proved the ability of sdAbs to be able to bind to G4s in general [62]. However, in the past few decades, the presence of G4s in various vital cellular processes [130] requires G4-targeting candidates to be highly sequence-specific to avoid potential damage to these pathways. The initial off-target interaction screening results have shown the ability of S10 to bind strongest to preC G4. Though S10's binding affinity to CACNA1C G4 was estimated to be stronger than that observed in BG4-preC G4, the K_D remains within the micromolar range and saturation in the binding curve in Fig 3.2 was not observed. Thereby, the true affinity strength may be higher. Additionally, LiCl-based buffers have been used before to inhibit G4-formation in G4-containing oligos [131-135]. The interaction studies executed using LiCl-based buffer resulted in a ~100-fold difference in binding strength, which highlights the importance of the G4 structure for the interaction with the therapeutic target sequence. Interestingly, to our knowledge, other G4-binding molecules have not been tested to assess whether structure-specificity plays a role in their targeting. Overall, since only S10-preC G4 interaction resulted in an interaction with a dissociation constant within a nanomolar range, it

can be determined that interactions with other G4s may be transient in nature. This is significant, for this prevents stable interactions with host G4s to be formed.

Worldwide distribution of Covid-19 therapeutics to low- and middle-income sites have been challenging in a rapidly evolving pandemic where access to inexpensive storage strategies is limited [136, 137]. HBV infections have also had a global effect and therapeutic distribution is expected to face similar challenges as other worldwide-spread infectious disease therapeutics. One of the limitations in therapeutic development is the ability for them to remain physiologically active for an extended period [138-141]. S10 has shown therapeutic relevance from its ability to maintain its performance following a six-month storage period (Table 3.2). Though a decrease in affinity is observed (Table 3.2) at six months up to a K_D of ~200 nM, there exists other strategies to increase the stability of protein and antibody therapeutics available. Some of these strategies include the addition of protease inhibitors, BSA, or cryoprotectants such as glycerol [142-146]. Further work focusing on increasing the sdAb stability is yet to be done, but protease inhibitors may be key to storage improvement since one of the main observable changes is the emergence of degradation bands in SDS-PAGE following storage of S10 (Fig2.3G).

For this therapeutic approach to efficiently neutralize any preC G4 activity, the complex formed by the interaction cannot be transient. SEC has been used before to study interactions between molecules to assess the transientness of their interactions, where transient interactions exhibit asymmetric peaks with long-tailing shoulder regions [147]. S10-preC G4 complex demonstrates

profiles consistent with stable interactions, as observed in Fig 3.3, where the peaks are well separated and show a normal distribution. SEC-MALS experiments can provide insight into the exact composition of the complex formed (stoichiometry, molecular weight) [147]. However, the SDS-PAGEs confirm that the earlier elution peak in the complexed sample contains both analytes, S10 and preC G4.

Lastly, we gained preliminary information on the suitability of peptide tag additions to sdAbs for improved drug delivery. Though longer coding peptide chains (Table 3.2) added to sdAbs, as is the case in S10-1 and S10-2, show a negative impact on the ability of S10 to target preC G4, S10-3 shows promising results. The K_D of the interaction between 10-3 and preC G4 approximates the high nanomolar range regardless of pH differences used to improve the binding ability of sdAbs [148]. The addition of peptide chains may have an effect on the ability to interact with preC G4. However, previous work has shown promise in fusing proteins and peptides to sdAbs, such as NLS sequences, to localize sdAbs in the cellular nuclei [62]. Further screening of other peptide tag additions to S10 may be needed to maximize the ability of S10 to function while acquiring the localization advantages that NLS and tissue specificity sequences provide.

3.5 CONCLUSION

In this study, it has been shown that interaction studies performed at the biochemical level allow for the preliminary screening and testing of drug candidates. From the initial candidates, S10 has shown the most promise for preC G4 targeting, where the strength in its binding to its target falls within the nanomolar range. Though expected from antibody-based therapeutics, it also shows a

stronger interaction than other G4-targetting therapeutics in development. Furthermore, S10 is the first report of an antibody-based ligand that is highly structure and sequence-specific when targeting G4s. This suggests a low risk of S10 potentially exhibiting off-target effects due to interactions with human G4s. SEC confirms that the nature of the complex formed between S10 and preC G4 is non-transient and indicates its potential to neutralize preC G4 function in infected cells. Furthermore, S10 has been shown to retain its function following long-term storage, but further testing may be advantageous to increase the storage time it can withstand while maintaining its function. Lastly, peptide-chain additions to S10 can affect its performance. However, the development of S10-3 and previous work in SG4 provide evidence that further screening may be needed to develop a combination of peptide chains that allow S10 to target preC G4 with a similar binding strength when unmodified.

CHAPTER 4: BIOPHYSICAL AND STRUCTURAL CHARACTERIZATION OF THE sdAb AND preC G4

4.1 INTRODUCTION

During the 1990s, X-ray crystallography structure-based drug design was successfully used for the first time, and since, the biophysical applications in the early phases of drug discovery have grown [149]. The application of these techniques extends to a variety of different goals in the field, which include the identification of binding sites by X-ray crystallography [150], thermodynamic determination by ITC [151-153], mass measurements by mass spectrometry [154], and tracking of large-scale structural changes upon ligand binding by SAXS [155], amongst others. Overall, the biophysical information gathered by these methods can guide how to proceed in the therapeutic design process, whether the next step involves improving target interactions, tissue delivery, cargo improvement, or other modifications [149, 156].

It is known that protein and nucleic acids can adopt a variety of different stable secondary or tertiary conformations due to local thermodynamic and kinetic local energy folding funnels [157]. High-specificity drugs are needed to avoid potential off-target effects and practical drug payoff [158]. Previous work has observed the ability of receptor binding domains in SARS-CoV-2 spike proteins to change between “down” and “up” conformations [53]. Since only “up” conformations participate in infection progression, the ability to distinguish between both structures can be vital for the proper development of a therapeutic candidate [53]. Though resorting to techniques such as Cryo-EM, SAXS, and X-ray crystallography can be time-

consuming and challenging to access in some cases, CD spec can offer quick, low-resolution structural data for both DNA and proteins [97, 159]. S10 has been designed to target transcriptionally active promotor sites, which require the region to exhibit its G4 structure[32]. Since CD spec has already been widely used to characterize the existence of G4s and assess the β-sheets content of proteins, the method can serve as a quick quality control test to verify that both the ligand and target are appropriately folded.

Structural data from these techniques ensure the quality of analytes tested and can provide the necessary information to take the following steps in drug development. Previous work [160] has been able to innovate antibody function through bioengineering new antibodies based on their biophysical characteristics and project goals. Some may include changing the therapeutic candidate's avidity, solubility, stability, increased tissue penetration and retention. FAB_C105, a therapeutic candidate for SARS-Cov-2, was developed to target the receptor binding (RBD) protein and was biophysically characterized through X-ray crystallography and Cryo-EM [161]. Interestingly, structural studies observed that the interacting surface area of FAB_C105 and another candidate (sb23) overlapped. However, only sb23 would be suitable for increasing avidity by coupling or fusion to other RBD-targeting candidates [53]. These biophysical studies proposed a future drug development pathway toward bivalent drugs to increase avidity and residency time [162]. Structural information further confirms the ligand-target interaction and can provide the information necessary to generate a direction for therapeutic development on a case-by-case basis.

The biochemical characterization of modified S10's interactions to preC G4 has shown that the addition of peptide tags affects the ability of S10 to perform as expected. Fusion tags can impact protein properties such as folding, solubility, and expression and may have different outcomes when expressed at the N or C terminus of the protein of interest [163]. Understanding the nature of these changes can be vital for further developments in tissue-specificity and nuclear-localization of S10. However, the limitations of biophysical techniques such as CryoEM, SAXS, and X-ray crystallization [149, 164] make them inefficient and expensive approaches to studying poorly performing therapeutic candidates. The development of computational tools such as AlphaFold has allowed us to predict the structure of various proteins, but limitations in predicting intrinsically disordered regions remain [165]. Tools such as homologous-structure servers and intrinsic disorder region predictors offer insight into the structural properties of a protein sequence based on known protein models [166], and sequence-based properties [167, 168]. Interestingly, the feedback from this readily available software has not yet been used as a tool to assess potential structural changes in the protein of interest that may arise from peptide tag additions.

Overall, the characterization of the structural properties of the interactions between the target and therapeutic candidate helps us elucidate opportunities for innovation and refinement in the therapeutic development strategy. In the following study, CD Spec is used as a quality control technique to show the proper folding of G4 targets and proper folding of S10. Next, the biophysical properties of the S10-preC G4 complex interaction are characterized, and a model depicting the complex formed is proposed. Lastly, homology-based modelling and intrinsic

disorder prediction software are used to study the potential effects of peptide tags on sdAb folding and function.

4.2 METHODS

4.2.1 Circular Dichroism (CD) Spectropolarimetry

The spectra for sdAbs and G4s were acquired on a calibrated Jasco J-815 spectropolarimeter (Jasco Inc, Easton, MD) at 220-320 nM (G4s) and 195-320 nM (sdAbs). A 1.0 mm cuvette was used at 32 s integration time. Each sample was analyzed in triplicate, and the read-out of their respective buffer was subtracted from the spectra reading. Protein secondary structure of the sdAb was assessed using K2D3 [169] and BESTSEL [159, 170-173] software available online. The sequence in BESTSEL was analyzed with a single-spectrum analysis tool.

4.2.2 Size Exclusion Chromatography – Multiple Angle Light Scattering – Dynamic Light Scattering (SEC-MALS-DLS)

The MALS-DLS experiments incorporate SEC using a Thermo Scientific™ - Vanquish™ Core HPLC System alongside a Shodex KW403-4F (Showa Denko America, Inc) column. The analytes were tested in 1x PBS buffer (137 nM NaCl, 2.7 mM KCl, 10 mM NA2HPO4, 1.8 mM KH2PO4, pH 7.4) at room temperature (20 ºC), and at a 0.5mL/min flow rate. The samples were injected at a volume of 50 µL and a concentration of ~1.5 mg/mL. The complex was first mixed in a 2:1 concentration ratio (S10: preC G4) and incubated overnight at 4 ºC before concentrating to 1.5mg/mL. S10 and S10-preC G4 were studied in Dawn ® (Wyatt Technology Corporation, Santa Barbara, CA, USA) multi-angle light scattering instrument. The equipment has 18 detecting angles, which use a 658 nm laser. Optilab ® refractometer (Wyatt Technology Corporation, Santa Barbara, CA, USA) was used to measure the solvent refractive index, defined

as 1.331(20OC) in both experiments. The dn/dc (refractive index increment) used for S10 – preC G4 complex analysis was 0.1775 mL/g (average of standard nucleic acid and protein dn/dc) and 0.1850 mL/g for the S10 analysis.

The data were processed and analyzed with μAstra v8.0.0.25 (Wyatt Technology Corporation, Santa Barbara, CA, USA). The analytes' molecular weight (Mw) was calculated with equation 1, where R(θ) is Rayleigh's ratio, K is the polymer constant, and c is the concentration of the solute.

Eq. 1: $2Mw=R(\theta)K*c$

The diffusion coefficient (Dτ) in the same scattering volume view was determined with integrated DLS measurements taken at 3-second intervals by the MALS detector. The hydrodynamic radius (RH) was calculated using the Stokes-Einstein equation (eq. 2) [174], where: kB is the Boltzmann coefficient (1.380 x 10-23 kg·m2·s-2K-1), T is the absolute temperature, η is the medium viscosity (1.025). The values were calculated using ASTRA 144v8.0.0.25.145.

Eq. 2: $Rh = (k_BT/6\pi\eta D\tau$

4.2.3 Small Angle X-ray Scattering (SAXS)

HPLC-SAXS data for the sdAb and sdAb-Sp1 G4 Complex was collected by the B21 BioSAXS beamline at the Diamond Light Source (Didcot, UK) synchrotron facility, as previously described [175]. Both samples were injected (50 µL) into an in-line Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC connected to a specialized flow cell at 450 µmol/L concentration and flowed through a buffer equilibrated Shodex KW403-4F column (Showa

Denko America, Inc) size exclusion column. S10 was run in 1x PBS buffer (pH 7.4), and S10 – preC G4 complex was run in MST buffer (50% 1x PBS, 50% G4 buffer, 0.5% Tween-20, pH 7.4). Each analyte sample was injected into 1x PBS pre-equilibrated Shodex KW403-4F column. X-rays were exposed to each frame for 3s, and an average of 10 frames per sample were integrated.

The peak regions were buffer subtracted and merged with CHROMIXS [176]. The Guinier approximation was used to obtain the radius of gyration (Rg) and to study the homogeneity of the merged data sets [177]. Dimensionless Kratky plots examined the analytes' foldedness[178]. The electron pair-distance distribution function plot (P(r) analysis done by GNOM) [179] was used to obtain an Rg and a maximum particle dimension (Dmax). Twenty models per analyte were generated from the P(r) plots using DAMMIN[180], and as described before, no enforced symmetry was used [181]. A single model representative of the 20 models was obtained by averaging and filtering them through DAMAVER [182]. The SAXS data analysis packages are included in ATSAS Suite (version 3.0.5).

4.2.4 Predictor of Natural Disordered Regions (PONDR) Software

The following six residues intrinsic disorder prediction software were tested to determine a potential aid in the process of modifications design: SPOT-Dis, ODiNPred, MFDp2, IUPred-long, prDOS, and PONDR-VLXT, RIDAO. The software was chosen based on previous publications discussing either improved accuracy of prediction or analysis speed [168, 183, 184]. The software can be trained to predict using a variety of approaches, so at least one software using each prediction approach was used: meta, evolutionary profile, and NMR-trained methods.

Based on previously identified G4-binding sdAb [62], a fourth modification, "-4," was designed to test in the software. Overall, the S10 residue sequence, modified counterparts, and previously known G4-binding sdAb with its peptide tag were analyzed. The three CDRs, as previously defined (26-35, 50-59, and 95-116 respectively) [185-187], and their neighbouring sequences up and downstream were observed for predicted increase or decrease in disorder. All software provided a threshold of 0.5 in their disorder probability results, meaning a score above 0.5 has a high probability of becoming disordered and vice-versa.

4.2.5 Homology Modelling of S10 and its Modifications

All S10 sequences were modelled in Protein Homology/Analogy Recognition Engine V 2.0 (Phyre2) under a normal modelling mode. Only S10-5 was further modelled using the intensive modelling mode to support multi-domain modelling. In Swiss-model Software, all the sequences were individually entered and modelled by searching the Swiss-model template library All models in SWISS-Modelling were built with ProdMod3 3.2.1. The software provides a GMQE and QMeanDisCo Global scores which rate the quality of the 3D structures generated in absence of experimental references. In cases where multiple templates were used to generate more than one model of a single modification, the model with highest QMeanDisCo Global score was chosen for further analysis.

All S10 sequences were modelled in Protein Homology/Analogy Recognition Engine V 2.0 (Phyre2) under a normal modelling mode [188]. Only S10-5 was further modelled using the intensive modelling mode to support multi-domain modelling. In Swiss-model Software, all the sequences were individually entered and modelled by searching the Swiss-model template

library [189-191]. All models in SWISS-Modelling were built with ProdMod3 3.2.1 [192]. The software provides GMQE and QMeanDisCo Global scores, which rate the quality of the 3D structures generated without experimental references [193]. In cases where multiple templates were used to generate more than one model of a single modification, the model with the highest QMeanDisCo Global score was chosen for further analysis.

4.3 RESULTS

4.3.1 Biophysical Properties of S10 and preC G4

The CD spectra of S10 visually differ significantly from S10-1 and S10-2 (Fig 4.1). S10-1 and S10-2 share the most similarities from 200-210 nm and 225-240 nm. Both CD spectra analysis outcomes indicate that the main portion of structured regions in S10 falls within the β-sheets category (Table 4.1). Interestingly, BESTSEL estimated that S10-1 and S10-2 are disordered in the entirety of their structure, whereas K2D3 analysis predicts that up to 90% of the protein sequence may be involved in α-folding.

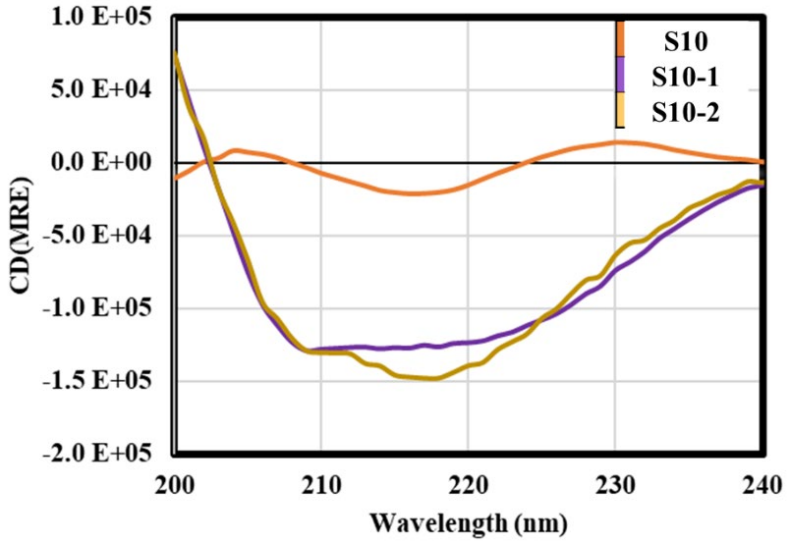

Figure 4.1. Circular dichroism spectra of S10, S10-1, and S10-2 used to determine the secondary structure content of each sdAb (n=3).

Table 4.1. Percentage of secondary structures that comprise the entirety of the sdAb structure for S10 and its modified counterparts as estimated by K2D3 and BESTSEL analysis software from circular dichroism spectra data. RMSD is the root mean square deviation which measures the similarity between the experimental CD Spectra and the spectra expected from the predicted secondary structures, where a score of 0 represents identical spectra.

	BESTSEL Analysis					**K2D3 Analysis**		
		Percentage of β-strands Parallel				**Secondary Structure Percentage**		
sdAb	**Anti - Parallel**	Left-Twist	Relaxed	Right-Twist	RMSD	α-helices	β-strands	Random Coil
S10	0	11	25	64	1	9	16	84
S10-1	100	0	0	0	21.31	90	1	99
S10-2	100	0	0	0	NA	90	1	99

The CD spectra generated from G4-folded oligos (Fig. 4.2) confirmed the presence of CD maxima and minima at the wavelengths known to be hallmark indicators of parallel G4 presence [32] in all four oligos tested. Interestingly, this observation was also seen in both FAM9a samples, taken from SEC elution fractions exhibiting a peak and a shoulder in its chromatogram.

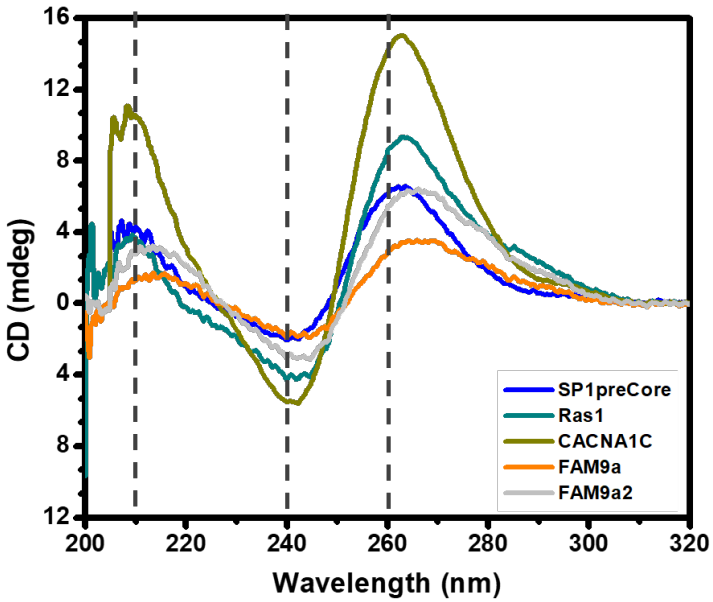

Figure 4.2. Circular dichroism spectra of the four different G4-forming DNA oligos in G4 buffer. All the oligos display the hallmark fingerprint spectra indicative of parallel G4 presence. These include 210 and 265 nm peaks and dips at 240 nm (n=3). FAM9a2 represents an elution fraction from the shoulder of the FAM9a peak in Fig 2.6.

4.3.2 Biophysical Studies of the S10 - preC G4 Interactions

As seen in Table 4.2, the difference in molecular weight of unbound S10 and Complex is 9.24 kg/mol. Considering the margin of error of the molecular weight of unbound S10 and complexed S10, the difference in molecular weight approximates the calculated sequence molecular weight of preC G4 (7.3 kg/mol). It is essential to note in Fig 4.3 that the complex and S10 did not fully separate, as observed by the elution peaks. However, as Table 4.2 describes, the polydispersity ratio is 1.000 in both peaks. Though the molecular weight of S10 and Unbound S10 (Table 4.2) differs by 2.36 kg/mol, the hydrodynamic radius calculated are almost identical. Similarly, the

complexed sample also shows an increase in the hydrodynamic radius. Lastly, all the preC G4 sample seemingly has been bound to S10, for no third peak is visible in the SEC-MALS profile.

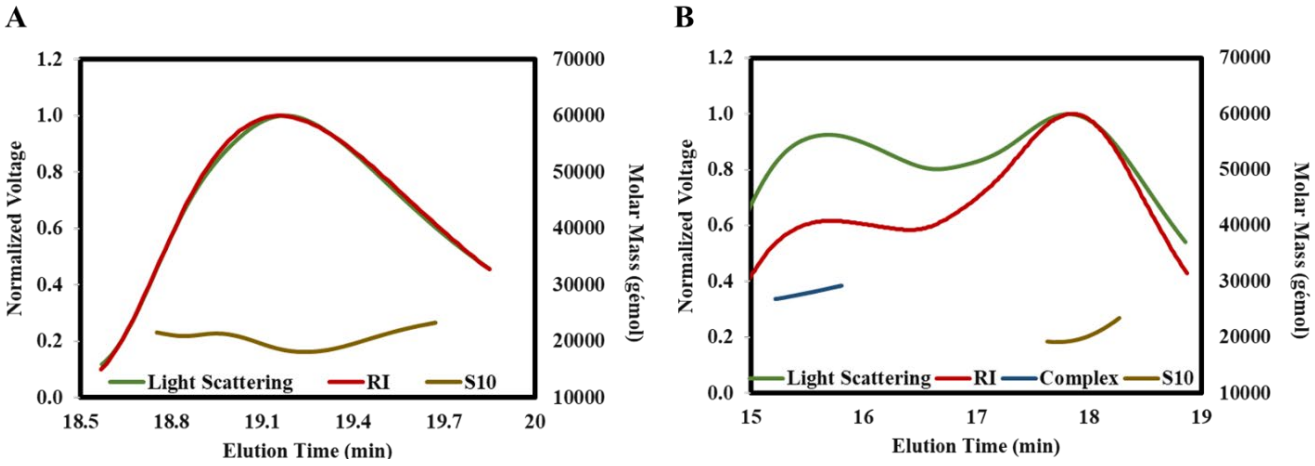

Figure 4.3. SEC-MALS analysis of (A) S10 and (B) S10 - preC G4 complex alongside unbound S10. S10 in the legend of each graph refers to S10's absolute molar mass (g/mol), and complex refers to the molar mass of the complex formed between S10 and preC G4. RI represents the refractive index.

Table 4.2 Biophysical properties estimated from SEC-MALS analysis where two separate samples were assessed: i) S10 alone, and ii) S10-preC G4 complex alongside unbound S10.

	S10	Complex	Unbound S10
Hydrodynamic Radius (Q) moments (nm)	2.74 (±2.74%)	3.64 (±3.48%)	2.78 (±3.71%)
Molecular Weight (g/mol)	1.80×10^4 (±5.51%)	2.95×10^4 (±4.10%)	2.03×10^4 (±2.79%)
Polydispersity (Mw/Mn)	1.00 (±7.87%)	1.00 (±5.77%)	1.00 (±3.95%)

Table 4.3. SAXS analysis and structure characteristics of the S10-preC G4 complex. The molecular weight for S10 and S10-preC G4 complex were estimated from SAXS analysis. "*" The molecular weight for preC G4 was calculated from the monomeric sequence. All preC G4 parameters are given as previously published [32]. The $\chi 2$ value was estimated in DAMMIN, and the NSD (normalized standard deviation) was calculated in DAMAVER.

Data collection parameters	preC G4	S10	S10 - preC G4
I(0) from P(r)	$0.0034 \pm 0.7 \times 10^{-5}$	$0.01245 \pm 1.1 \times 10^{-5}$	$0.03804 \pm 5.3 \times 10^{-5}$
I(0) from Guinier plot	$0.0035 \pm 1.1 \times 10^{-5}$	0.01245	0.03803
Rg (Å) from P(r)	17.18 ± 0.03	24.44 ± 0.04	32.56 ± 0.07
Rg (Å) from Guinier plot	18.37 ± 0.11	24.39	32.32
Dmax (Å)	46.85	79.22	112.27
qRg Limits	0.24–1.30	0.28 - 1.30	0.50 - 1.23
Molecular Weight (kDa)	7.3*	21.927	28.85
χ^2	~1.80	~1.09	~1.13
NSD	0.55 ± 0.01	0.58 ± 0.02	0.59 ± 0.01

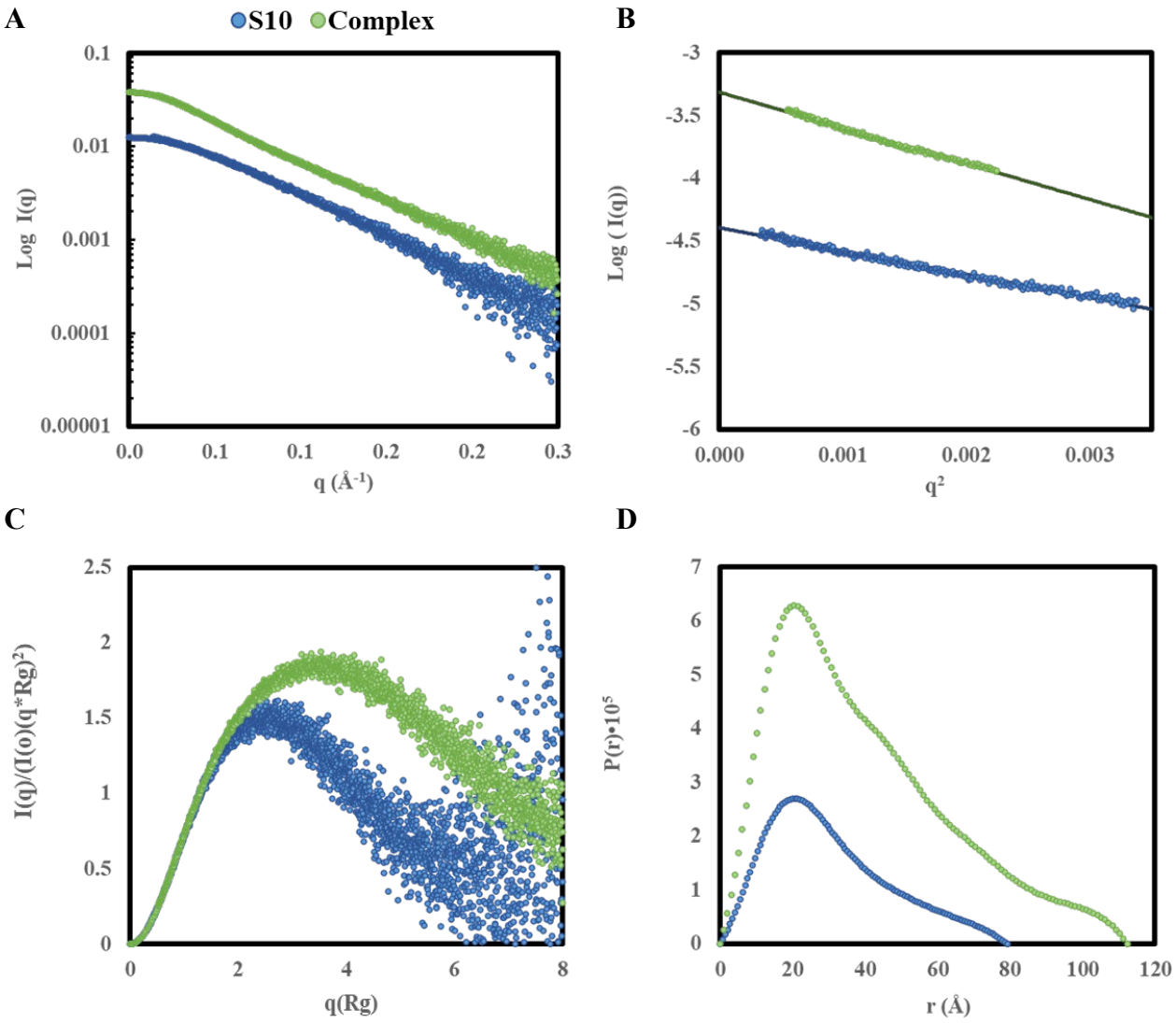

Figure 4.4. Low-resolution biophysical studies of S10 (blue) and S10-preC G4 complex (green). A) Raw scattering plot in log Y scale, where momentum transfer (q) and scattered light intensity are shown. B) Guinier plot following within limits shown in the SAXS software (S10: points 68-372; Complex: points 86-269). C) Kratky dimensionless plots calculated from the Guinier I(0) and Rg. D) Electron pair-distance distribution function plot. P(r) graph.

The scattering from both sample sets, S10 and Complex (S10-preC G4), had enough intensity of scattered light sets to proceed with SAXS analysis. The raw scattering (Fig 4.4A) also shows a pattern consistent with a rod in the complex and slightly more globular in S10. The Guinier region (Fig 4.4B) shows that all the data of both subjects fall over the regression of the points within this region. This linear relationship is indicative of a homogenous and aggregate-free sample. The Kratky plot (Fig 4.4C) indicates well-folded and globular protein for S10, whereas the plot for the complex has a similar trend with increased flexibility and less foldedness or globularity. The electron distance pair-distribution function plot (Fig 4.4D) for both analytes shows a relative globular concentration of electrons as indicated by the initial normal distribution from 0 to 40 Å with a tail sticking out at one end, as seen by the trailing shoulder to the right. The complex also displays a potential secondary concentration of electrons at one end of the complex. The difference between the D_{max} acquired from the P(r) graph is 33.05 nm, where the expected D_{max} of preC G4 is 46.85, as seen in Table 4.3. However, the difference in molecular weight of the complex and S10 calculated from these SAXS experiments is 6.92 kg/mol, which closely approximates the calculated molecular weight from the sequence of preC G4. The overall shape of the modelled preC G4 is relatively dumbbell in shape; thereby, it is difficult to estimate the exact orientation of this molecule accurately. However, the orientation of the S10 can be estimated more efficiently with the model generated of the complex, as observed in Fig 4.5.

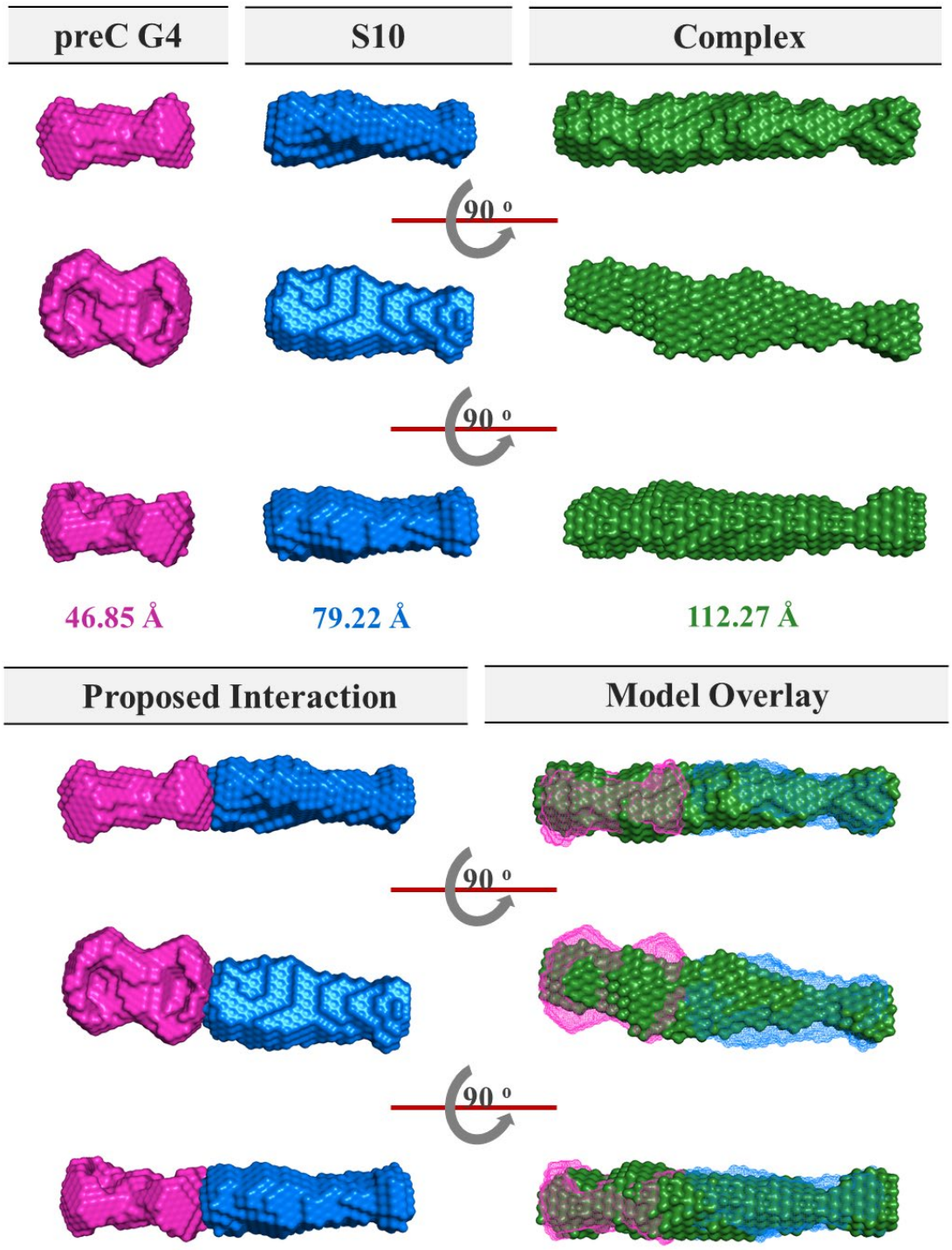

Figure 4.5. Low-resolution average-filtered models of S10 and S10-preC G4 complex and derived by DAMMIN and DAMAVER calculations. Two images from each structure at 90°

rotation about the x-axis from the middle of the model. The D_{max} of the models is shown below the first three models. The bottom row shows the proposed interaction between the individual analytes fitted in PyMOL and based by the complexed model.

4.3.3 Prediction of Structural Changes in Peptide Fused – S10

As observed in Fig 4.7B, PONDR software predicted S10-1, S10-2, and S10-5 to have a higher likelihood of intrinsic disorder in the third CDR region of the sdAb. S10-1 also showed increased disorder likelihood in the N-terminal of the sdAb. IUPred3 (Fig 4.7C) shows a similar pattern as PONDR software for S10-1 and S10-2, but the increase in disorder likelihood is relatively more discrete than previously observed in PONDR. The predictions observed in RIDAO (Fig 4.7D-I) show the N-terminal mean disorder prediction (MDP) range for S10-1 and S10-3, and S10-5 deviating from S10's. Similarly, the C-terminal MDP range for S10-1, S10-2, and S10-4 deviate from S10's. Close to the N-terminal of the CDR3 region, it can be observed that S10-1, S10-2, and S10-5 have an overall increase in MDP's intrinsic disorder likelihood. Overall, the modified candidates that were observed to deviate most times from S10's intrinsic disorder under the parameters explained above are S10-1 and S10-2.

Interestingly, Phyre2 homology modelling of S10-1 and S10-2 resulted in a high amount of predicted disorder (table 4.4), which is consistent with the intrinsic disorder predictors. Additionally, these two models also displayed the lowest coverage in Phyre2. The SWISS Modelling for S10-1 and S10-2 shows long trailing disorder peptide chains at the C-terminals. Interestingly, these two models also demonstrated the lowest QMEAN DisCo Global scores of all the models generated, which suggests low confidence in the accuracy of the models. Furthermore, these two models had the lowest sequence similarity and identity to their templates.

SWISS modelling of S10, S10-3, and S10-4 scored the same QMEAN DisCo score and automatically used the same template. S10-5 was a model from Phyre2 that was limited in its data provided since it required assembly using a multi-domain approach to modelling. Similarly, S10-5's low scoring and poor homology shown in Fig 5 are also influenced by the presence of entire domains (GFP) in the S10 sequence.

Table 4.4. Statistical outcomes of S10 homology modelling in Phyre2 and SWISS Model Software. "*" in Phyre2 data shows the sdAbs scoring above 50% disorder which lowers the likelihood that the model is accurate according to developers. "*" in SWISS Model data shows those QMEAN DisCo Global below 0.80.

Phyre2							
sdAb	Coverage	Confidence	Disordered	Alpha helix	Beta strand	Beta r in VHH	Template
S10	89% (113r)	99.90%	41	0	55	69	d1vhpa_
S10-1*	64% (117r)	100%	61	0	47	69	d1vhpa_
S10-2*	64% (117r)	100%	54	0	53	69	d1vhpa_
S10-3	73% (115r)	100%	39	6	46	71	d1vhpa_
S10-4	76% (115r)	99.90%	48	0	53	68	d1xfpa_
S10-5	88% (366r)	>90%	-	-	-	-	c6wznA_
SWISS Model							
sdAb	Coverage	GMQE	QMEAN DisCo Global	Sequence Identity	Sequence Similarity	Residue Range	Template
S10	0.94	0.78	0.80±0.08	83.33	0.54	3-121	7te8.1.A
S10-1	0.74	0.59	0.71±0.07*	57.66%	0.45	18-158	5xcv.1.A
S10-2	0.87	0.63	0.66±0.07*	56.88	0.44	3-162	5HI5.1.C
S10-3	0.76	0.67	0.80±0.08	83.33	0.54	31-149	7te8.1.A
S10-4	0.79	0.69	0.80±0.08	83.33	0.54	3-121	7te8.1.A
S10-5	0.93	0.15	0.44±0.07*	61.08	0.46	6-136	7xm9.1.B

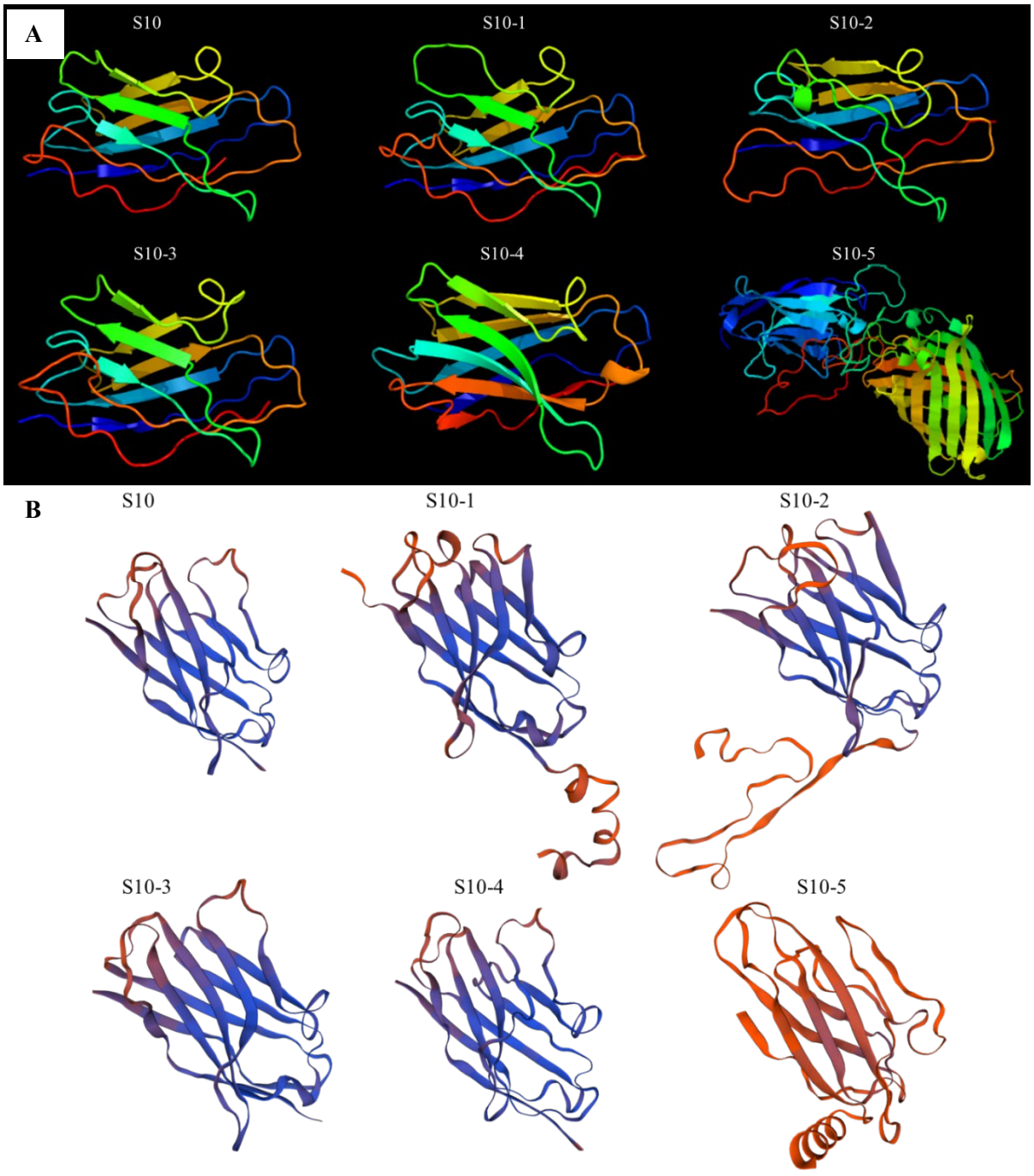

Figure 4.6. Homology models of S10 and peptide tag fused S10s. (A) Models generated by Phyre2. (B) Models generated by the SWISS model. The blue regions represent highly conserved regions, and the red regions represent poorly conserved regions. The N-terminal of all the sdAb models in (B) were positioned to have the N-terminal facing the top left of the image.

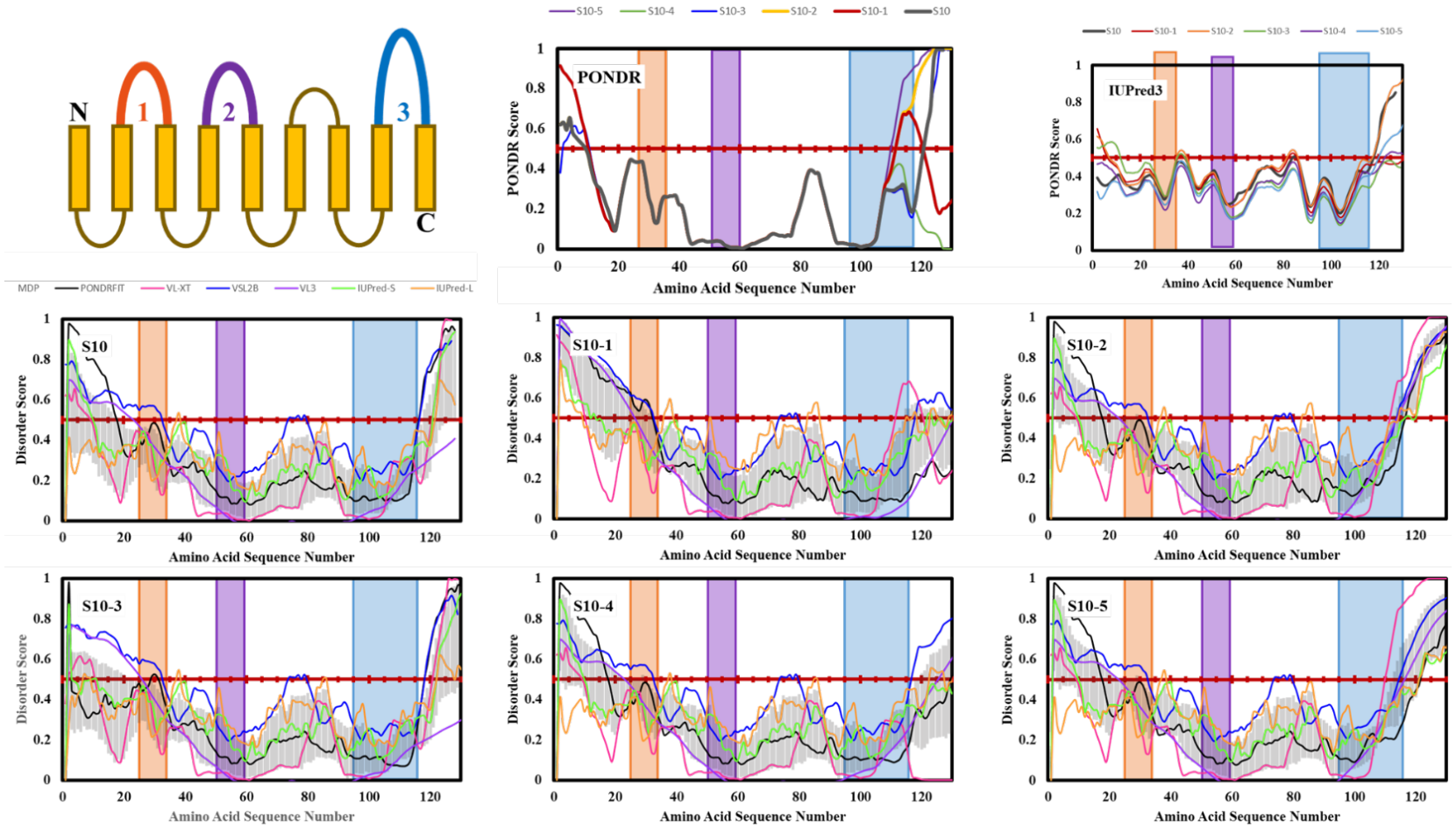

Figure 4.7. Sequence-disorder analysis of the S10 modifications as calculated by predicting software. A) shows the overall secondary structure of S10, where the yellow boxes represent beta sheets, 1 (orange) represents CDR1, 2 (purple) represents CDR2, and 3 (blue) represents CDR3. B) The initial PONDR ® software analysis depicting disorder changes in the N and C terminal regions. C) IUPred3 software results of all S10 modifications. D-I) Show a compilation of disorder region analysis by various software where MDP (in gray) represents the average disorder predicted from all the software used along its error margin. The software calculates the disorder probability and indicates 0.5 (in red) as the threshold to determine whether something is

4.4 DISCUSSION

The proper conditions must exist for G4 structures to fold and exhibit the G4-specific interactions being studied [32, 132]. It has already been shown that disruption of the G4 structure inhibits core protein expression [32]. Therefore, it is vital for the sdAb to specifically target the G4 structure during therapeutic development to ensure inhibition of the intended target is successful. Here, the results in CD spec confirm the presence of this G4 structure in preC G4 based on G4 footprint spectra inflections observed at 210, 240 and 260 nm [194, 195]. This confirms that the structure-specific interactions observed in MST (Table 3.1) are a product of the presence of G4 structures. Similarly, it can also be confirmed that the off-target G4s tested were also correctly folded, and thereby, the weak interactions observed in MST suggest that upon exposure to other parallel G4 structures, S10 is unlikely to interact.

It is known that a large portion of single-domain antibodies has a high degree of sequence conservation, apart from their CDR regions [185]. A large percentage of the sdAb sequence is known to participate in the formation of β-sandwich motifs [196, 197], allowing the CDRs to be positioned towards one end of the molecule. S10 shows the ability to engage in β-sheet formation according to BESTSEL and K2D3 analysis, where this secondary structure is the most represented. In contrast, S10-1 and S10-2 have conflicting results where K2D3 estimates that 90% of the sequence is involved in α-folding versus 100% involvement in antiparallel β-sheet formation. Regardless, both results differ from the expected number of residues involved in β-sheet formation. We suggest that the failure of S10-1 and S10-2 to interact with preC G4 in previous MST studies may result from changes in the structure of S10 associated with the fusion of these specific peptide tag configurations (Table 3.3).

The sequence-calculated molecular weight of preC G4 is 7.3 kg/mol. Based on the difference observed in SAXS of molecular weight between S10 and S10 – preC G4, the G4 was estimated to be ~6.9 kg/mol and SEC-MALS ~9.24 kg/mol. Both results closely approximate the expected size, where SEC-MALS results differ the most by 1.9 kg/mol. Furthermore, the presence of a sdAb dimer is unlikely since it is expected that the size of an S10-dimer would be that of 30-34 kg/mol, based on a monomer size of 15-18 kg/mol as shown in SDS-PAGEs (Fig 2.3). The RH and Rg changes observed also have a ~150% increase when complex bound, which reflects the change in the dimensions of the molecule consistent with S10 and preC G4 dimensions. Furthermore, the Guinier region of the SAXS scattering (Fig 4.4B) follows the straight regression line, which indicates aggregate-free samples [198]. The polydispersity ratio indicates the non-uniformity of the size distribution of particles, where a value of 1 indicates a monodispersed sample [199, 200]. SEC-MALS indicates that the ranges studied within the SEC peaks (Fig 4.3) contain mainly complex or mainly unbound S10, and the likelihood of a significant number of other compounds is low. Based on these results, it has been determined that the stoichiometry of the S10 – preC G4 interaction is 1:1.

The patterns observed in the Kratky plot (qRg decrease post-peak maxima is slow) and the electron distance pair-distribution function plot (not normal distribution and longer lagging right shoulder) suggest that the complex is more elongated than unbound S10 [201]. Additionally, the difference of Dmax - observed in the complex (112.27 Å) is close to the sum of S10's (79.22 Å) and preC G4's (46.85 Å) Dmax-, 126.07 Å. As expected from the positioning of the CDR regions in sdAb [202], the nature of the interaction is to be head-on at one end of the sdAb. This is consistent with the models generated with DAMAVER and DAMMIN, where a head-to-head

interaction is expected to occur, as suggested by the overlay of models generated by unbound and in-complex S10 and preC G4. The symmetry of preC G4 makes it challenging to discern the exact orientation of the G4 in the interaction. However, it can be observed that the surface area of preC G4 exposed may allow for a different secondary ligand to be used in the future, as has been done before [53] to improve the neutralization outcomes of sdAb therapy.

The topology of the sdAb binding site is sculpted in part by noncanonical disulfide bridges where the CDR3 loop is heavily involved in its binding activity and affinity [51]. As observed from CD Spec and MST data (Fig 4.1, Table 3.1), it was shown that adding peptide tags (S10-1 and S10-2) could impact function and folding. However, modifications in S10-3 showed less than a 10-fold decrease in binding affinity performance (Fig 3.2, Table 3.1). The positioning and type of peptide tag used can impact the function of either the protein of interest or the peptide [163]. The IDPs were used to show a correlation between the observed results in CD spec, MST and disorder prediction, where an increase in the probability of disorder specifically in the CDR3 loop region or C-terminal was observed specifically for S10-1 and S10-2. Homology modellings by the SWISS model and Phyre2 showed consistent results where S10-1 and S10-2 were deemed to contain above 50% intrinsic disorder, and the models generated also scored in quality (QMEAN DisCo Global), coverage, and sequence similarity. S10-3 does not show the degree of change in the disorder probability as S10-1 and S10-2 do. Furthermore, the homology models suggest that the predicted model of S10-3 score similarly to S10's models based on the QMEANDisCo score, sequence similarity and identity.

Based on these results, previously used functioning modifications in G4-binding sdAbs[62] were adapted to S10 to assess whether IDPs could predict the ability of these modifications to avoid sdAb disruption. Interestingly, the S10-4 CDR3 loop region and C-terminal did not increase the likelihood of intrinsic disorder. S10-4 also performed similarly to S10 and S10-3 in quality and sequence similarity. Additionally, the program chose the same template for S10, S10-3, and S10-4, providing further evidence of their shared biochemical and biophysical properties. However, S10-5 shows a slight increase in probability in these regions, and its modelling did not provide insight into its ability to function. Adding GFP to S10 may function as a folding chaperone in vitro since the modification contains a fully formed secondary domain rather than disordered peptide tags. Phyre2 and SWISS-model mainly function as a single domain predicting tool; the ability to gain insight into more complex fusion tags through this approach may be limited. Overall, IDPs and homology modelling software can be a tool to check for potential structural changes related to peptide tags that can affect sdAb binding function. This prediction approach serves as a quick check for potential changes in peptide tagged-sdAb structure that may result in loss of function, thereby saving time and money invested in non-functional candidates.

4.5 CONCLUSION

In conclusion, the structural data described in this study provides insight into the development of therapeutic candidates such as S10. CD Spec is a useful tool to examine the secondary structure of proteins and DNA. Here, we used it to examine the presence of G4s in the DNA sequences tested by MST. CD Spec also provided evidence that S10-1's and S10-2's loss of function may be a product of structural changes in the S10 sequence as reported by K2D3 and BESTSEL. SEC-MALS and SAXS showed that the S10-preC G4 complex was formed with a 1:1

stoichiometry based on the change of D_{MAX}, and molecular weight. A low-resolution model was generated using SAXS, which revealed the orientation of the analytes within the interaction and the potential for developing a secondary preC G4 based on G4 surface area exposure. IDP and homology modelling software was used to discern and predict functional versus non-functional peptide tag modifications. The predictions of S10 and all the peptide modified S10s closely resembled the interactions observed in MST, the trends observed in CD Spec and other previously published data by other groups. The use of this software represents an opportunity to decrease the cost and time invested in investigating whether a modification to the sdAb sequence may have any detrimental effects on the structure of sdAbs.

CHAPTER 5: CONCLUDING STATEMENTS

The number of chronic infections of HBV has grown from 250 million in 2017 [10] to 296 million estimated cases in 2019 [203] despite the existence of vaccines. There exist three main approaches to the treatment of chronic HBV infection at this time: IFN therapy, NAs or a combination of NA and PEG-IFN therapy [204]. Various conditions affect the efficacy of IFN therapy in patients, including their gender, age, viral load, presence of anti-IFN antibodies, and viral genotype or mutant [205]. These variables often determine whether a patient is an appropriate candidate for this therapeutic approach. NAs have been proven effective at controlling the infection, but some side effects exist, and their use does not prevent the establishment of cccDNA within the hepatocyte's nucleus[204]. Apart from current approaches' side effects and limitations, none have successfully provided a functional cure. Infection relapse is still challenging due to cccDNA reservoir activity once therapy stops [206]. The presence of cccDNA reservoirs and their contribution to relapse makes it an attractive target for therapeutic development.

Other groups are currently developing therapeutic candidates with cccDNA disruptive properties [26, 204, 207]. However, to our knowledge, these candidates either target human molecules that inhibit the HBV viral cycle, or they fall under the gene editing category [208-210]. Like currently available IFN therapies [205], targeting molecules involved in other host pathways increases the risk of developing secondary effects detrimental to the patient's health. As is the example of the human SP1 protein, where through its interaction with the HBV genome has been described [38], SP1 is also involved in disease [211] and predicted to interact with undescribed

cellular function [212]. Gene editing therapies have an inherently increased risk of malignant tumorigenesis, off-target mutagenesis, and insertional mutagenesis, and in the case of CRISPR/Cas9, appropriate delivery systems and avoiding anti-Cas9 responses are difficult [213, 214]. Therefore, there exists a need for therapeutic approaches that are highly specific and do not exhibit therapeutic gene risks.

Antibody therapy has shown great success due to its high affinity and specificity to their target in cancer and infectious diseases [215, 216]. Traditionally, this approach was considered costly due to its production, storage and administration [215]. However, single-domain antibodies (sdAbs) present an alternative to conventional antibodies that preserve the specificity and affinity qualities while providing advantages such as their small size, high stability and ease of recombinant expression [217]. Additionally, time-consuming production approaches such as those requiring animal immunizations[218] are unnecessary due to synthetic libraries [219] that allow for antibody screening to be sped up. Other proteins can target G4s with a high affinity, but they do so non-specifically [220].

Current sdAbs being developed for therapeutics or theranostics target proteins [217, 221] such as CD20 due to their presence in B-cell lymphocytes in cancer [222] or such as the receptor binding domain (RBD) of the SARS-Cov-2 spike protein [53]. Recently, the emergence of SG4 has proved the ability of sdAbs to target G4 structures [62]. Previously reported sdAbs being developed as therapeutics have an affinity to their protein epitope ranging from 1-60 nM [53, 111, 222]; testing conditions (buffers) and instruments used to differ. Further work is necessary

to assess affinity effects on the thermodynamics of the interactions as well as how they compare *in vivo*. IFN therapy has been reported to lower the levels of intrahepatocellular cccDNA [22, 223]. This particular property can be exploited in combination therapy approaches with S10 to increase the efficacy of cccDNA targeting by decreasing the amount of cccDNA present in the host.

In conclusion, in this work, we present the ability of S10 to be expressed in bacterial systems and purified through various chromatography steps. These steps can be quickly adopted by other sdAbs with minimal adjustments. The dissociation constant of S10 for preC G4 is ~90 nM in MST studies, comparable to previously studied sdAb candidates. The MST interaction studies confirm that this is the first time a sdAb has been developed to target DNA based on the parallel G4 structure and the oligo's overall sequence. SEC-MALS and SAXS determined the stoichiometry of the complex formed to be 1:1. A low-resolution model describing the nature of the interactions was generated and fitted with SAXS data. Based on the surface area of preC G4 exposed, a secondary ligand could be generated to improve the ability of S10 to neutralize G4 function to improve inhibition of G4 function in future cellular experiments. Lastly, modifications looking to improve tissue-specificity can impact S10 function due to structural changes, as observed in CD spec and predicted by IDPs and homology modelling software. However, modifications such as S10-3, S10-4, and S10-5 demonstrate the feasibility of using peptide tags to promote tissue specificity and nuclear localization in the future.

REFERENCES

1. Organization, W.H. *Hepatitis B*. 2019 2022-08-24]; Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b.
2. Gentile, I. and G. Borgia, *Vertical transmission of hepatitis B virus: challenges and solutions.* International Journal of Women's Health, 2014. **6**: p. 605-611.
3. Nanwa, N., et al., *The mean attributable health care costs associated with hepatitis B virus in Ontario, Canada: A matched cohort study.* Canadian Liver Journal, 2022. **5**(3): p. 339-361.
4. Canada, P.H.A.o., *Report on Hepatitis B and C Surveillance in Canada: 2019*. 2022-01-17.
5. Reuten, R., et al., *Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. Nat Commun 7*. 2016.
6. Canada, G.o. *Hepatitis B vaccine: Canadian Immunization Guide*. May 2022; Available from: https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-4-active-vaccines/page-7-hepatitis-b-vaccine.html.
7. Columbia, G.o.B. *Hepatitis B*. 2022 [cited 2023; Available from: https://immunizebc.ca/hepatitis-b.
8. Services, A.H. *Hepatitis B*. 2018; Available from: https://immunizealberta.ca/i-need-know-more/diseases-covered/hepatitis-b.
9. Ministry of Health, M.o.L.-t.C., *Immunization: Hepatitis B Vaccine for babies.* 2022.
10. WHO, *Global Hepatitis Report, 2017 [Internet]*. 2017.
11. Değertekin, H. and G. Güneş, *Horizontal transmission of hepatitis B virus in Turkey.* Public Health, 2008. **122**(12): p. 1315-1317.
12. Doganci, T., et al., *Horizontal transmission of hepatitis B virus in children with chronic hepatitis B.* World J Gastroenterol, 2005. **11**(3): p. 418-20.
13. Veronese, P., et al., *Prevention of vertical transmission of hepatitis B virus infection.* World J Gastroenterol, 2021. **27**(26): p. 4182-4193.
14. Wu, J.-F. and M.-H. Chang, *Natural history of chronic hepatitis B virus infection from infancy to adult life -the mechanism of inflammation triggering and long-term impacts.* Journal of Biomedical Science, 2015. **22**(1): p. 92.
15. Rajoriya, N., et al., *How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach?* Journal of Hepatology, 2017. **67**(6): p. 1281-1297.
16. Liu, Z., et al., *Distribution of hepatitis B virus genotypes and subgenotypes: A meta-analysis.* Medicine (Baltimore), 2021. **100**(50): p. e27941.
17. Tsukuda, S. and K. Watashi, *Hepatitis B virus biology and life cycle.* Antiviral Research, 2020. **182**: p. 104925.
18. Wagner, J., et al., *Analysis of Hepatitis B Virus Haplotype Diversity Detects Striking Sequence Conservation Across Genotypes and Chronic Disease Phase.* Hepatology, 2021. **73**(5): p. 1652-1670.
19. D'Souza, S., et al., *Molecular mechanisms of viral hepatitis induced hepatocellular carcinoma.* World J Gastroenterol, 2020. **26**(38): p. 5759-5783.

20. Chuang, Y.-C., K.-N. Tsai, and J.-H.J. Ou, *Pathogenicity and virulence of Hepatitis B virus.* Virulence, 2022. **13**(1): p. 258-296.
21. Nassal, M., *HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B.* Gut, 2015. **64**(12): p. 1972-1984.
22. Bockmann, J.H., et al., *Comparative Analysis of the Antiviral Effects Mediated by Type I and III Interferons in Hepatitis B Virus-Infected Hepatocytes.* J Infect Dis, 2019. **220**(4): p. 567-577.
23. Newbold, J.E., et al., *The covalently closed duplex form of the hepadnavirus genome exists in situ as a heterogeneous population of viral minichromosomes.* J Virol, 1995. **69**(6): p. 3350-7.
24. Guo, J.-T. and H. Guo, *Metabolism and function of hepatitis B virus cccDNA: Implications for the development of cccDNA-targeting antiviral therapeutics.* Antiviral Research, 2015. **122**: p. 91-100.
25. Allweiss, L. and M. Dandri, *The Role of cccDNA in HBV Maintenance.* Viruses, 2017. **9**(6): p. 156.
26. Tsounis, E.P., et al., *Toward a new era of hepatitis B virus therapeutics: The pursuit of a functional cure.* World Journal of Gastroenterology, 2021. **27**(21): p. 2727-2757.
27. Bhattacharya, D. and Chloe, *Review of Hepatitis B Therapeutics.* Clinical Infectious Diseases, 2010. **51**(10): p. 1201-1208.
28. Tawada, A., T. Kanda, and O. Yokosuka, *Current and future directions for treating hepatitis B virus infection.* World J Hepatol, 2015. **7**(11): p. 1541-52.
29. Zoulim, F., F. Lebossé, and M. Levrero, *Current treatments for chronic hepatitis B virus infections.* Current Opinion in Virology, 2016. **18**: p. 109-116.
30. Dandri, M., *Epigenetic modulation in chronic hepatitis B virus infection.* Seminars in Immunopathology, 2020. **42**(2): p. 173-185.
31. Wong, G.L.H., E. Gane, and A.S.F. Lok, *How to achieve functional cure of HBV: Stopping NUCs, adding interferon or new drug development?* Journal of Hepatology, 2022. **76**(6): p. 1249-1262.
32. Meier-Stephenson, V., et al., *Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA.* Journal of Biological Chemistry, 2021. **296**.
33. An, J.-Y., et al., *Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder.* Science, 2018. **362**(6420): p. eaat6576.
34. Perera, D., et al., *Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes.* Nature, 2016. **532**(7598): p. 259-263.
35. Heidenreich, B., et al., *TERT promoter mutations in cancer development.* Current Opinion in Genetics & Development, 2014. **24**: p. 30-37.
36. Maraganore, D.M., et al., *Collaborative Analysis of α-Synuclein Gene Promoter Variability and Parkinson Disease.* JAMA, 2006. **296**(6): p. 661-670.
37. Li, J. and J.-h. Ou, *Differential Regulation of Hepatitis B Virus Gene Expression by the Sp1 Transcription Factor.* Journal of Virology, 2001. **75**(18): p. 8400-8406.
38. Meier-Stephenson, V., et al., *Comprehensive Analysis of Hepatitis B Virus Promoter Region Mutations.* Viruses, 2018. **10**(11): p. 603.
39. Eddy, J., et al., *G4 motifs correlate with promoter-proximal transcriptional pausing in human genes.* Nucleic Acids Research, 2011. **39**(12): p. 4975-4983.

40. Moon, J., et al., *Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin.* Biochemistry and Biophysics Reports, 2015. **2**: p. 29-35.
41. Lavezzo, E., et al., *G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide.* PLOS Computational Biology, 2018. **14**(12): p. e1006675.
42. Métifiot, M., et al., *G-quadruplexes in viruses: function and potential therapeutic applications.* Nucleic Acids Research, 2014. **42**(20): p. 12352-12366.
43. Abiri, A., et al., *Unlocking G-Quadruplexes as Antiviral Targets.* Pharmacological Reviews, 2021. **73**(3): p. 897-923.
44. Leavy, O., *Therapeutic antibodies: past, present and future.* Nature Reviews Immunology, 2010. **10**(5): p. 297-297.
45. Jovčevska, I. and S. Muyldermans, *The Therapeutic Potential of Nanobodies.* BioDrugs, 2020. **34**(1): p. 11-26.
46. Huang, C.-J., H. Lin, and X. Yang, *Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements.* Journal of Industrial Microbiology and Biotechnology, 2012. **39**(3): p. 383-399.
47. Samaranayake, H., et al., *Challenges in monoclonal antibody-based therapies.* Annals of Medicine, 2009. **41**(5): p. 322-331.
48. Cruz, E. and V. Kayser, *Monoclonal antibody therapy of solid tumors: clinical limitations and novel strategies to enhance treatment efficacy.* Biologics: Targets and Therapy, 2019. **13**: p. 33-51.
49. Harmsen, M.M. and H.J. De Haard, *Properties, production, and applications of camelid single-domain antibody fragments.* Applied Microbiology and Biotechnology, 2007. **77**(1): p. 13-22.
50. Krah, S., et al., *Single-domain antibodies for biomedical applications.* Immunopharmacology and Immunotoxicology, 2016. **38**(1): p. 21-28.
51. Rossotti, M.A., et al., *Immunogenicity and humanization of single-domain antibodies.* The FEBS Journal, 2022. **289**(14): p. 4304-4327.
52. Van Heeke, G., et al., *Nanobodies®††Nanobody is a registered trademark of Ablynx NV. as inhaled biotherapeutics for lung diseases.* Pharmacology & Therapeutics, 2017. **169**: p. 47-56.
53. Custódio, T.F., et al., *Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2.* Nature communications, 2020. **11**(1): p. 5588-5588.
54. Zimmermann, I., et al., *Generation of synthetic nanobodies against delicate proteins.* Nature Protocols, 2020. **15**(5): p. 1707-1741.
55. Bannas, P., J. Hambach, and F. Koch-Nolte, *Nanobodies and Nanobody-Based Human Heavy Chain Antibodies As Antitumor Therapeutics.* Front Immunol, 2017. **8**: p. 1603.
56. Pardridge, W.M., *Blood–brain barrier delivery.* Drug Discovery Today, 2007. **12**(1): p. 54-61.
57. Begley, D.J., *Delivery of therapeutic agents to the central nervous system: the problems and the possibilities.* Pharmacology & Therapeutics, 2004. **104**(1): p. 29-45.
58. Mashel, T.V., et al., *Overcoming the delivery problem for therapeutic genome editing: Current status and perspective of non-viral methods.* Biomaterials, 2020. **258**: p. 120282.

59. Esposito, D. and D.K. Chatterjee, *Enhancement of soluble protein expression through the use of fusion tags.* Current Opinion in Biotechnology, 2006. **17**(4): p. 353-358.
60. Bolhassani, A., B.S. Jafarzade, and G. Mardani, *In vitro and in vivo delivery of therapeutic proteins using cell penetrating peptides.* Peptides, 2017. **87**: p. 50-63.
61. Kim, B.-K., et al., *Homodimeric SV40 NLS peptide formed by disulfide bond as enhancer for gene delivery.* Bioorganic & Medicinal Chemistry Letters, 2012. **22**(17): p. 5415-5418.
62. Galli, S., et al., *DNA G-Quadruplex Recognition In Vitro and in Live Cells by a Structure-Specific Nanobody.* Journal of the American Chemical Society, 2022. **144**(50): p. 23096-23103.
63. Casteleijn, M.G., A. Urtti, and S. Sarkhel, *Expression without boundaries: Cell-free protein synthesis in pharmaceutical research.* International Journal of Pharmaceutics, 2013. **440**(1): p. 39-47.
64. Giegé, R., et al., *The role of purification in the crystallization of proteins and nucleic acids.* Journal of Crystal Growth, 1986. **76**(3): p. 554-561.
65. Chirino, A.J. and A. Mire-Sluis, *Characterizing biological products and assessing comparability following manufacturing changes.* Nature Biotechnology, 2004. **22**(11): p. 1383-1391.
66. Raynal, B., et al., *Quality assessment and optimization of purified protein samples: why and how?* Microb Cell Fact, 2014. **13**: p. 180.
67. Wakelin, S.J., et al., *"Dirty little secrets"—Endotoxin contamination of recombinant proteins.* Immunology Letters, 2006. **106**(1): p. 1-7.
68. Schmidt, F.R., *Recombinant expression systems in the pharmaceutical industry.* Applied Microbiology and Biotechnology, 2004. **65**(4): p. 363-372.
69. Swiech, K., V. Picanço-Castro, and D.T. Covas, *Human cells: New platform for recombinant therapeutic protein production.* Protein Expression and Purification, 2012. **84**(1): p. 147-153.
70. Khambhati, K., et al., *Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems.* Frontiers in Bioengineering and Biotechnology, 2019. **7**.
71. Conibear, A.C., *Deciphering protein post-translational modifications using chemical biology tools.* Nature Reviews Chemistry, 2020. **4**(12): p. 674-695.
72. Walsh, G., *Post-translational modifications of protein biopharmaceuticals.* Drug Discovery Today, 2010. **15**(17): p. 773-780.
73. Jenkins, N., L. Murphy, and R. Tyther, *Post-translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals.* Molecular Biotechnology, 2008. **39**(2): p. 113-118.
74. Shrivastava, A., et al., *N-Glycosylation of monoclonal antibody therapeutics: A comprehensive review on significance and characterization.* Analytica Chimica Acta, 2022. **1209**: p. 339828.
75. Liu, Y. and H. Huang, *Expression of single-domain antibody in different systems.* Applied Microbiology and Biotechnology, 2018. **102**(2): p. 539-551.
76. Porath, J., et al., *Metal chelate affinity chromatography, a new approach to protein fractionation.* Nature, 1975. **258**(5536): p. 598-599.
77. Lichty, J.J., et al., *Comparison of affinity tags for protein purification.* Protein Expression and Purification, 2005. **41**(1): p. 98-105.

78. Young, C.L., Z.T. Britton, and A.S. Robinson, *Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications.* Biotechnology Journal, 2012. **7**(5): p. 620-634.
79. Kyo, S., et al., *Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT).* Nucleic Acids Res, 2000. **28**(3): p. 669-77.
80. Kikin, O., L. D'Antonio, and P.S. Bagga, *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.* Nucleic Acids Research, 2006. **34**(suppl_2): p. W676-W682.
81. Maggi, M. and C. Scotti, *Enhanced expression and purification of camelid single domain VHH antibodies from classical inclusion bodies.* Protein Expression and Purification, 2017. **136**: p. 39-44.
82. Rahbarizadeh, F., et al., *High expression and purification of the recombinant camelid anti-MUC1 single domain antibodies in Escherichia coli.* Protein Expression and Purification, 2005. **44**(1): p. 32-38.
83. Vaneycken, I., et al., *Preclinical screening of anti-HER2 nanobodies for molecular imaging of breast cancer.* The FASEB Journal, 2011. **25**(7): p. 2433-2446.
84. Chi, X., et al., *Humanized single domain antibodies neutralize SARS-CoV-2 by targeting the spike receptor binding domain.* Nature Communications, 2020. **11**(1).
85. Crowell, L.E., et al., *Development of a platform process for the production and purification of single-domain antibodies.* Biotechnology and Bioengineering, 2021. **118**(9): p. 3348-3358.
86. Bao, X., et al., *Optimization of dilution refolding conditions for a camelid single domain antibody against human beta-2-microglobulin.* Protein Expression and Purification, 2016. **117**: p. 59-66.
87. Jalali-Javaran, M., et al., *Production and characterization of anti-(mucin MUC1) single-domain antibody in tobacco (Nicotiana tabacum cultivar Xanthi).* Biotechnology and Applied Biochemistry, 2007. **47**(1): p. 11.
88. Snøve, O. and T. Holen, *Many commonly used siRNAs risk off-target activity.* Biochemical and Biophysical Research Communications, 2004. **319**(1): p. 256-263.
89. Young, M.J., *Off-Target Effects of Drugs that Disrupt Human Mitochondrial DNA Maintenance.* Frontiers in Molecular Biosciences, 2017. **4**.
90. MacDonald, M.L., et al., *Identifying off-target effects and hidden phenotypes of drugs in human cells.* Nature Chemical Biology, 2006. **2**(6): p. 329-337.
91. Saerens, D., et al., *Identification of a Universal VHH Framework to Graft Non-canonical Antigen-binding Loops of Camel Single-domain Antibodies.* Journal of Molecular Biology, 2005. **352**(3): p. 597-607.
92. Muyldermans, S., *Applications of Nanobodies.* Annual Review of Animal Biosciences, 2021. **9**(1): p. 401-421.
93. Ni, Y., et al., *Hepatitis B and D Viruses Exploit Sodium Taurocholate Co-transporting Polypeptide for Species-Specific Entry into Hepatocytes.* Gastroenterology, 2014. **146**(4): p. 1070-1083.e6.
94. Lauer, S.A. and J.P. Nolan, *Development and characterization of Ni-NTA-bearing microspheres.* Cytometry, 2002. **48**(3): p. 136-145.

95. Krielgaard, L., et al., *Effect of tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant human factor XIII.* Journal of Pharmaceutical Sciences, 1998. **87**(12): p. 1593-1603.
96. George, J., et al., *Structural and mutational analysis of a monomeric and dimeric form of a single domain antibody with implications for protein misfolding.* Proteins: Structure, Function, and Bioinformatics, 2014. **82**(11): p. 3101-3116.
97. Kypr, J., et al., *Circular dichroism and conformational polymorphism of DNA.* Nucleic Acids Research, 2009. **37**(6): p. 1713-1725.
98. Del Villar-Guerra, R., J.O. Trent, and J.B. Chaires, *G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy.* Angewandte Chemie International Edition, 2018. **57**(24): p. 7171-7175.
99. Carvalho, J., J.A. Queiroz, and C. Cruz, *Circular Dichroism of G-Quadruplex: a Laboratory Experiment for the Study of Topology and Ligand Binding.* Journal of Chemical Education, 2017. **94**(10): p. 1547-1551.
100. Małgowska, M., et al., *How to study G-quadruplex structures.* BioTechnologia, 2012. **4**: p. 381-390.
101. Dai, W.-J., et al., *CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles.* Molecular Therapy - Nucleic Acids, 2016. **5**: p. e349.
102. Deng, Y., et al., *Therapeutic potentials of gene silencing by RNA interference: Principles, challenges, and new strategies.* Gene, 2014. **538**(2): p. 217-227.
103. Rázga, F. and V. Némethová, *Selective Therapeutic Intervention: A Challenge against Off-Target Effects.* Trends in Molecular Medicine, 2017. **23**(8): p. 671-674.
104. Rao, D.D., et al., *Comparative assessment of siRNA and shRNA off target effects: what is slowing clinical development.* Cancer Gene Therapy, 2009. **16**(11): p. 807-809.
105. Zhang, X.-H., et al., *Off-target Effects in CRISPR/Cas9-mediated Genome Engineering.* Molecular Therapy - Nucleic Acids, 2015. **4**: p. e264.
106. Ma, W., L. Yang, and L. He, *Overview of the detection methods for equilibrium dissociation constant KD of drug-receptor interaction.* Journal of Pharmaceutical Analysis, 2018. **8**(3): p. 147-152.
107. Hoare, S.R.J., *The Problems of Applying Classical Pharmacology Analysis to Modern In Vitro Drug Discovery Assays: Slow Binding Kinetics and High Target Concentration.* SLAS DISCOVERY: Advancing the Science of Drug Discovery, 2021. **26**(7): p. 835-850.
108. Shakib, S., et al., *Microscale Thermophoresis in Liquids Induced by Plasmonic Heating and Characterized by Phase and Fluorescence Microscopies.* The Journal of Physical Chemistry C, 2021. **125**(39): p. 21533-21542.
109. Mrozowich, T., V. Meierstephenson, and T.R. Patel, *Microscale thermophoresis: warming up to a new biomolecular interaction technique.* The Biochemist, 2019. **41**(2): p. 8-12.
110. Valenzuela Nieto, G., et al., *Potent neutralization of clinical isolates of SARS-CoV-2 D614 and G614 variants by a monomeric, sub-nanomolar affinity nanobody.* Scientific Reports, 2021. **11**(1): p. 3318.
111. Broos, K., et al., *Evaluating a Single Domain Antibody Targeting Human PD-L1 as a Nuclear Imaging and Therapeutic Agent.* Cancers, 2019. **11**(6): p. 872.
112. Zyner, K.G., et al., *G-quadruplex DNA structures in human stem cells and differentiation.* Nature Communications, 2022. **13**(1): p. 142.

113. Biffi, G., et al., *Quantitative visualization of DNA G-quadruplex structures in human cells.* Nature Chemistry, 2013. **5**(3): p. 182-186.
114. Ligat, G., et al., *Targeting Viral cccDNA for Cure of Chronic Hepatitis B.* Current Hepatology Reports, 2020. **19**(3): p. 235-244.
115. Lucifora, J. and U. Protzer, *Attacking hepatitis B virus cccDNA – The holy grail to hepatitis B cure.* Journal of Hepatology, 2016. **64**(1, Supplement): p. S41-S48.
116. Kim, S.W., et al., *Toward a complete cure for chronic hepatitis B: Novel therapeutic targets for hepatitis B virus.* Clin Mol Hepatol, 2022. **28**(1): p. 17-30.
117. Shahjin, F., S. Chand, and S.V. Yelamanchili, *Extracellular Vesicles as Drug Delivery Vehicles to the Central Nervous System.* Journal of Neuroimmune Pharmacology, 2020. **15**(3): p. 443-458.
118. Maus, A., L. Strait, and D. Zhu, *Nanoparticles as delivery vehicles for antiviral therapeutic drugs.* Engineered Regeneration, 2021. **2**: p. 31-46.
119. Vargason, A.M., A.C. Anselmo, and S. Mitragotri, *The evolution of commercial drug delivery technologies.* Nature Biomedical Engineering, 2021. **5**(9): p. 951-967.
120. Deshayes, S., et al., *Cell-penetrating peptides: tools for intracellular delivery of therapeutics.* Cellular and Molecular Life Sciences CMLS, 2005. **62**(16): p. 1839-1849.
121. Li, D., et al., *Improving pharmacological activities of thrombopoietin mimetic peptide by genetic fusion to albumin-binding domain.* Biotechnology Letters, 2023. **45**(4): p. 439-448.
122. Biffi, G., et al., *Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells.* Nature chemistry, 2014. **6**(1): p. 75-80.
123. Jaroch, K., A. Jaroch, and B. Bojko, *Cell cultures in drug discovery and development: The need of reliable in vitro-in vivo extrapolation for pharmacodynamics and pharmacokinetics assessment.* Journal of Pharmaceutical and Biomedical Analysis, 2018. **147**: p. 297-312.
124. Wang, W., et al., *Human MYC G-quadruplex: From discovery to a cancer therapeutic target.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2020. **1874**(2): p. 188410.
125. Tian, T., et al., *G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting.* Chem, 2018. **4**(6): p. 1314-1344.
126. Marzano, M., et al., *Evaluation of an Analogue of the Marine ε-PLL Peptide as a Ligand of G-quadruplex DNA Structures.* Marine Drugs, 2020. **18**(1): p. 49.
127. Sengupta, P., et al., *Site-specific amino acid substitution in dodecameric peptides determines the stability and unfolding ofc-MYCquadruplex promoting apoptosis in cancer cells.* Nucleic Acids Research, 2018. **46**(19): p. 9932-9950.
128. Yang, M., et al., *Targeting a noncanonical, hairpin-containing G-quadruplex structure from the MYCN gene.* Nucleic Acids Research, 2021. **49**(14): p. 7856-7869.
129. Fan, X., et al., *A homogeneous time-resolved fluorescence-based high-throughput screening for discovery of inhibitors of Nef-sdAb19 interaction.* Int J Oncol, 2015. **47**(4): p. 1485-1493.
130. Zyner, K.G., et al., *Genetic interactions of G-quadruplexes in humans.* eLife, 2019. **8**.
131. Li, Y., et al., *Ultrafast Microfluidic Mixer for Tracking the Early Folding Kinetics of Human Telomere G-Quadruplex.* Analytical Chemistry, 2014. **86**(9): p. 4333-4339.

132. Dahal, S., et al., *Characterization of G4 DNA formation in mitochondrial DNA and their potential role in mitochondrial genome instability.* The FEBS Journal, 2022. **289**(1): p. 163-182.
133. Rehm, C., et al., *A bacterial DNA quadruplex with exceptional K+ selectivity and unique structural polymorphism.* Chemical Science, 2014. **5**(7): p. 2809-2818.
134. Buglione, E., et al., *Nanomechanics of G-quadruplexes within the promoter of the KIT oncogene.* Nucleic Acids Research, 2021. **49**(8): p. 4564-4573.
135. Bardin, C. and J.L. Leroy, *The formation pathway of tetramolecular G-quadruplexes.* Nucleic Acids Research, 2007. **36**(2): p. 477-488.
136. Robinson, P.C., et al., *COVID-19 therapeutics: Challenges and directions for the future.* Proceedings of the National Academy of Sciences, 2022. **119**(15): p. e2119893119.
137. Uddin, M.N. and M.A. Roni, *Challenges of Storage and Stability of mRNA-Based COVID-19 Vaccines.* Vaccines, 2021. **9**(9): p. 1033.
138. Leader, B., Q.J. Baca, and D.E. Golan, *Protein therapeutics: a summary and pharmacological classification.* Nature Reviews Drug Discovery, 2008. **7**(1): p. 21-39.
139. Weiss, W.F., T.M. Young, and C.J. Roberts, *Principles, approaches, and challenges for predicting protein aggregation rates and shelf life.* Journal of Pharmaceutical Sciences, 2009. **98**(4): p. 1246-1277.
140. Szlachcic, A., M. Zakrzewska, and J. Otlewski, *Longer action means better drug: Tuning up protein therapeutics.* Biotechnology Advances, 2011. **29**(4): p. 436-441.
141. Crommelin, D.J.A., et al., *Addressing the Cold Reality of mRNA Vaccine Stability.* Journal of Pharmaceutical Sciences, 2021. **110**(3): p. 997-1001.
142. Daugherty, A.L. and R.J. Mrsny, *Formulation and delivery issues for monoclonal antibody therapeutics.* Advanced Drug Delivery Reviews, 2006. **58**(5): p. 686-706.
143. Ma, H., C. Ó'Fágáin, and R. O'Kennedy, *Antibody stability: A key to performance - Analysis, influences and improvement.* Biochimie, 2020. **177**: p. 213-225.
144. Miller, M.A., et al., *Frozen-State Storage Stability of a Monoclonal Antibody: Aggregation is Impacted by Freezing Rate and Solute Distribution.* Journal of Pharmaceutical Sciences, 2013. **102**(4): p. 1194-1208.
145. Fishman, J.B. and E.A. Berg, *Antibody purification and storage.* Cold Spring Harbor Protocols, 2019. **2019**(5): p. pdb. top099101.
146. Sun, M.-F., et al., *Identification and characterization of chemical and physical stability of insulin formulations utilizing degraded glycerol after repeated use and storage.* European Journal of Pharmaceutics and Biopharmaceutics, 2022. **177**: p. 147-156.
147. Wilamowski, M., et al., *Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication.* Biophysical Journal, 2021. **120**(15): p. 3152-3165.
148. Dejaegere, A., et al., *Variations in Antigen−Antibody Association Kinetics as a Function of pH and Salt Concentration: A QSAR and Molecular Modeling Study.* Biochemistry, 2005. **44**(44): p. 14409-14418.
149. Renaud, J.-P., et al., *Biophysics in drug discovery: impact, challenges and opportunities.* Nature Reviews Drug Discovery, 2016. **15**(10): p. 679-698.
150. Blundell, T.L. and S. Patel, *High-throughput X-ray crystallography for drug discovery.* Current Opinion in Pharmacology, 2004. **4**(5): p. 490-496.
151. Ladbury, J.E., G. Klebe, and E. Freire, *Adding calorimetric data to decision making in lead discovery: a hot tip.* Nature Reviews Drug Discovery, 2010. **9**(1): p. 23-27.

152. Claveria-Gimeno, R., et al., *A look at ligand binding thermodynamics in drug discovery.* Expert Opinion on Drug Discovery, 2017. **12**(4): p. 363-377.
153. Kairys, V., et al., *Binding affinity in drug design: experimental and computational techniques.* Expert Opinion on Drug Discovery, 2019. **14**(8): p. 755-768.
154. Zhang, A., H. Sun, and X. Wang, *Mass spectrometry-driven drug discovery for development of herbal medicine.* Mass Spectrometry Reviews, 2018. **37**(3): p. 307-320.
155. Cheung, E., et al., *Tools shaping drug discovery and development.* Biophysics Reviews, 2022. **3**(3).
156. Lua, L.H.L., et al., *Bioengineering virus-like particles as vaccines.* Biotechnology and Bioengineering, 2014. **111**(3): p. 425-440.
157. Leopold, P.E., M. Montal, and J.N. Onuchic, *Protein folding funnels: a kinetic approach to the sequence-structure relationship.* Proceedings of the National Academy of Sciences, 1992. **89**(18): p. 8721-8725.
158. Starr, C.G. and P.M. Tessier, *Selecting and engineering monoclonal antibodies with drug-like specificity.* Current Opinion in Biotechnology, 2019. **60**: p. 119-127.
159. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy.* Proceedings of the National Academy of Sciences, 2015. **112**(24): p. E3095-E3103.
160. Chiu, M.L., et al., *Antibody Structure and Function: The Basis for Engineering Therapeutics.* Antibodies, 2019. **8**(4): p. 55.
161. Barnes, C.O., et al., *Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies.* Cell, 2020. **182**(4): p. 828-842.e16.
162. Vauquelin, G. and S.J. Charlton, *Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands.* British Journal of Pharmacology, 2013. **168**(8): p. 1771-1785.
163. Ki, M.-R. and S.P. Pack, *Fusion tags to enhance heterologous protein expression.* Applied Microbiology and Biotechnology, 2020. **104**(6): p. 2411-2425.
164. D'Imprima, E. and W. Kühlbrandt, *Current limitations to high-resolution structure determination by single-particle cryoEM.* Quarterly Reviews of Biophysics, 2021. **54**: p. e4.
165. Ruff, K.M. and R.V. Pappu, *AlphaFold and Implications for Intrinsically Disordered Proteins.* Journal of Molecular Biology, 2021. **433**(20): p. 167208.
166. Muhammed, M.T. and E. Aki-Yalcin, *Homology modeling in drug discovery: Overview, current applications, and future perspectives.* Chemical Biology & Drug Design, 2019. **93**(1): p. 12-20.
167. He, B., et al., *Predicting intrinsic disorder in proteins: an overview.* Cell Research, 2009. **19**(8): p. 929-949.
168. Dayhoff II, G.W. and V.N. Uversky, *Rapid prediction and analysis of protein intrinsic disorder.* Protein Science, 2022. **31**(12): p. e4496.
169. Louis-Jeune, C., M.A. Andrade-Navarro, and C. Perez-Iratxeta, *Prediction of protein secondary structure from circular dichroism using theoretically derived spectra.* Proteins: Structure, Function, and Bioinformatics, 2012. **80**(2): p. 374-381.
170. Micsonai, A., et al., *BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra.* Nucleic Acids Research, 2018. **46**(W1): p. W315-W322.

171. Micsonai, A., É. Bulyáki, and J. Kardos, *BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy*, in *Structural Genomics: General Applications*, Y.W. Chen and C.-P.B. Yiu, Editors. 2021, Springer US: New York, NY. p. 175-189.
172. Micsonai, A., et al., *BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy.* Nucleic Acids Research, 2022. **50**(W1): p. W90-W98.
173. Micsonai, A., et al., *Disordered–Ordered Protein Binary Classification by Circular Dichroism Spectroscopy.* Frontiers in Molecular Biosciences, 2022. **9**.
174. Pusey, P., *Photon correlation and light beating spectroscopy.* Plenum Press, New York, 1974.
175. Meier, M., et al., *Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge.* Nucleic acids research, 2018. **46**(10): p. 5319-5331.
176. Panjkovich, A. and D.I. Svergun, *CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data.* Bioinformatics, 2018. **34**(11): p. 1944-1946.
177. Guinier, A., G. Fournet, and K.L. Yudowitch, *Small-angle scattering of X-rays.* 1955.
178. Durand, D., et al., *NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers.* Journal of structural biology, 2010. **169**(1): p. 45-53.
179. Svergun, D., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of applied crystallography, 1992. **25**(4): p. 495-503.
180. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophysical journal, 1999. **76**(6): p. 2879-2886.
181. Reuten, R., et al., *Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes.* Nature communications, 2016. **7**(1): p. 13515.
182. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of applied crystallography, 2003. **36**(3): p. 860-864.
183. Dass, R., F.A.A. Mulder, and J.T. Nielsen, *ODiNPred: comprehensive prediction of protein order and disorder.* Scientific Reports, 2020. **10**(1): p. 14780.
184. Nielsen, J.T. and F.A.A. Mulder, *Quality and bias of protein disorder predictors.* Scientific Reports, 2019. **9**(1): p. 5137.
185. Mitchell, L.S. and L.J. Colwell, *Comparative analysis of nanobody sequence and structure data.* Proteins: Structure, Function, and Bioinformatics, 2018. **86**(7): p. 697-706.
186. Al-Lazikani, B., A.M. Lesk, and C. Chothia, *Standard conformations for the canonical structures of immunoglobulins11Edited by I. A. Wilson.* Journal of Molecular Biology, 1997. **273**(4): p. 927-948.
187. Sircar, A., et al., *Analysis and Modeling of the Variable Region of Camelid Single-Domain Antibodies.* The Journal of Immunology, 2011. **186**(11): p. 6357-6367.
188. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature Protocols, 2015. **10**(6): p. 845-858.
189. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

190. Bienert, S., et al., *The SWISS-MODEL Repository-new features and functionality.* Nucleic Acids Res, 2017. **45**(D1): p. D313-d319.
191. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-w303.
192. Studer, G., et al., *ProMod3—A versatile homology modelling toolbox.* PLOS Computational Biology, 2021. **17**(1): p. e1008667.
193. Studer, G., et al., *QMEANDisCo—distance constraints applied on model quality estimation.* Bioinformatics, 2019. **36**(6): p. 1765-1771.
194. Lam, E.Y., et al., *G-quadruplex structures are stable and detectable in human genomic DNA.* Nat Commun, 2013. **4**: p. 1796.
195. Bagheri, Z., et al., *Spectral properties and thermal stability of AS1411 G-quadruplex.* International Journal of Biological Macromolecules, 2015. **72**: p. 806-811.
196. Zabetakis, D., et al., *Evaluation of Disulfide Bond Position to Enhance the Thermal Stability of a Highly Stable Single Domain Antibody.* PLOS ONE, 2014. **9**(12): p. e115405.
197. Bekker, G.-J., B. Ma, and N. Kamiya, *Thermal stability of single-domain antibodies estimated by molecular dynamics simulations.* Protein Science, 2019. **28**(2): p. 429-438.
198. Skou, S., R.E. Gillilan, and N. Ando, *Synchrotron-based small-angle X-ray scattering of proteins in solution.* Nature Protocols, 2014. **9**(7): p. 1727-1739.
199. Rogošić, M., H.J. Mencer, and Z. Gomzi, *Polydispersity index and molecular weight distributions of polymers.* European Polymer Journal, 1996. **32**(11): p. 1337-1344.
200. Danaei, M., et al., *Impact of Particle Size and Polydispersity Index on the Clinical Applications of Lipidic Nanocarrier Systems.* Pharmaceutics, 2018. **10**(2): p. 57.
201. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Letters, 2015. **589**(19, Part A): p. 2570-2577.
202. White, B., I. Huh, and C.L. Brooks, *Structure of a VHH isolated from a naïve phage display library.* BMC Research Notes, 2019. **12**(1): p. 154.
203. WHO, *Global progress report on HIV, viral hepatitis and sexually transmitted infections*. 2021, World Health Organization Geneva, Switzerland.
204. Leowattana, W. and T. Leowattana, *Chronic hepatitis B: New potential therapeutic drugs target.* World Journal of Virology, 2022. **11**(1): p. 57-72.
205. Ye, J. and J. Chen, *Interferon and Hepatitis B: Current and Future Perspectives.* Frontiers in Immunology, 2021. **12**.
206. Zhu, A., et al., *HBV cccDNA and Its Potential as a Therapeutic Target.* J Clin Transl Hepatol, 2019. **7**(3): p. 258-262.
207. Naggie, S. and A.S. Lok, *New Therapeutics for Hepatitis B: The Road to Cure.* Annual Review of Medicine, 2021. **72**(1): p. 93-105.
208. Cradick, T.J., et al., *Zinc-finger Nucleases as a Novel Therapeutic Strategy for Targeting Hepatitis B Virus DNAs.* Molecular Therapy, 2010. **18**(5): p. 947-954.
209. Ramanan, V., et al., *CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus.* Scientific Reports, 2015. **5**(1): p. 10833.
210. Chen, J., et al., *An Efficient Antiviral Strategy for Targeting Hepatitis B Virus Genome Using Transcription Activator-Like Effector Nucleases.* Molecular Therapy, 2014. **22**(2): p. 303-311.

211. Li, L. and J.R. Davie, *The role of Sp1 and Sp3 in normal and cancer cell biology.* Annals of Anatomy - Anatomischer Anzeiger, 2010. **192**(5): p. 275-283.
212. Mahmoud, M.M., N.A. Belal, and A. Youssif, *Prediction of Transcription Factor Binding Sites of SP1 on Human Chromosome1.* Applied Sciences, 2021. **11**(11): p. 5123.
213. Gao, Q., et al., *Therapeutic potential of CRISPR/Cas9 gene editing in engineered T-cell therapy.* Cancer Medicine, 2019. **8**(9): p. 4254-4264.
214. LaFountaine, J.S., K. Fathe, and H.D.C. Smyth, *Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9.* International Journal of Pharmaceutics, 2015. **494**(1): p. 180-194.
215. Casadevall, A., E. Dadachova, and L.-A. Pirofski, *Passive antibody therapy for infectious diseases.* Nature Reviews Microbiology, 2004. **2**(9): p. 695-703.
216. Scott, A.M., J.D. Wolchok, and L.J. Old, *Antibody therapy of cancer.* Nature Reviews. Cancer, 2012. **12**(4): p. 278-87.
217. Kang, X., et al., *[Single-domain antibody - advances in research and application].* Sheng wu gong cheng xue bao = Chinese journal of biotechnology, 2018. **34**(12): p. 1974-1984.
218. Pardon, E., et al., *A general protocol for the generation of Nanobodies for structural biology.* Nature protocols, 2014. **9**(3): p. 674-693.
219. Moutel, S., et al., *NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies.* eLife, 2016. **5**.
220. Zheng, K.-W., et al., *Detection of genomic G-quadruplexes in living cells using a small artificial protein.* Nucleic Acids Research, 2020. **48**(20): p. 11706-11720.
221. Czajka, T.F., D.J. Vance, and N.J. Mantis, *Slaying SARS-CoV-2 One (Single-domain) Antibody at a Time.* Trends in Microbiology, 2021. **29**(3): p. 195-203.
222. Krasniqi, A., et al., *Theranostic Radiolabeled Anti-CD20 sdAb for Targeted Radionuclide Therapy of Non-Hodgkin Lymphoma.* Molecular Cancer Therapeutics, 2017. **16**(12): p. 2828-2839.
223. Xia, Y., et al., *Interferon-γ and Tumor Necrosis Factor-α Produced by T Cells Reduce the HBV Persistence Form, cccDNA, Without Cytolysis.* Gastroenterology, 2016. **150**(1): p. 194-205.

APPENDIX A: Supplementary Information - SDS-PAGES OF PURIFIED SDABS

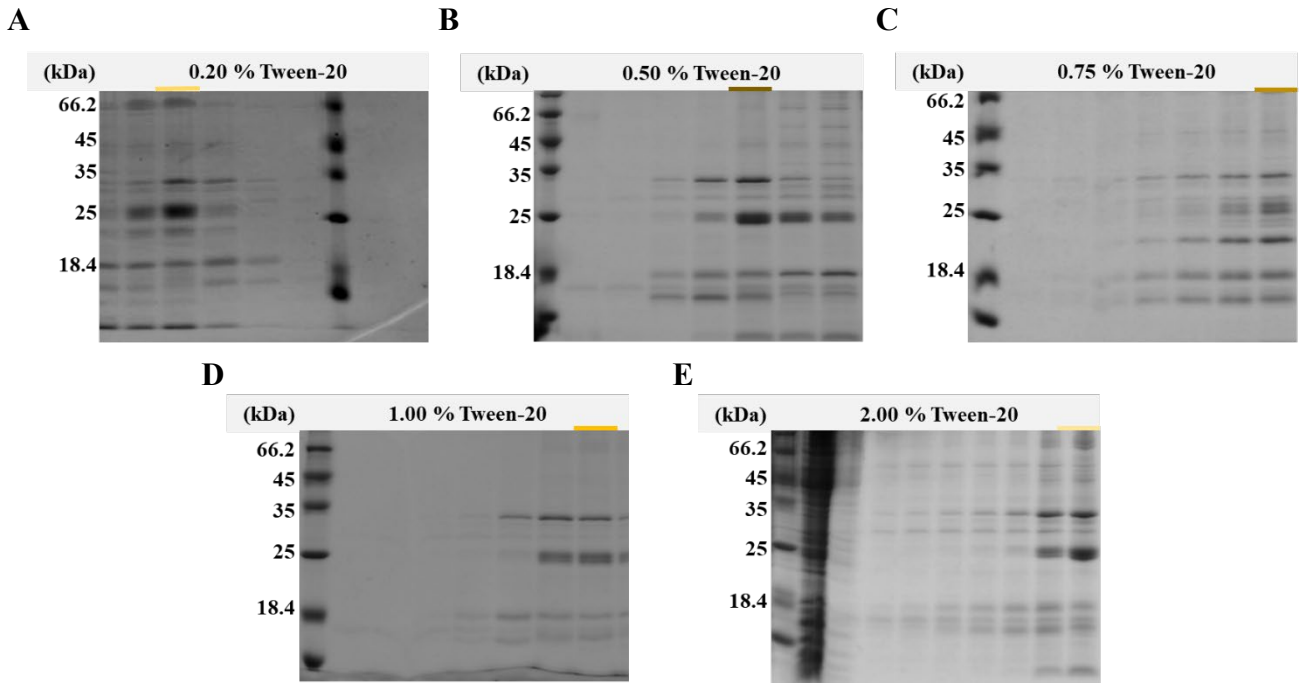

Figure S1. SDS-PAGE of post-Ni-NTA sdAbs showing differences in elution profile when subjected to different Tween-20 concentrations from the elution buffer. Each sample studied is highlighted and color-coded similarly to Fig 2.3F.

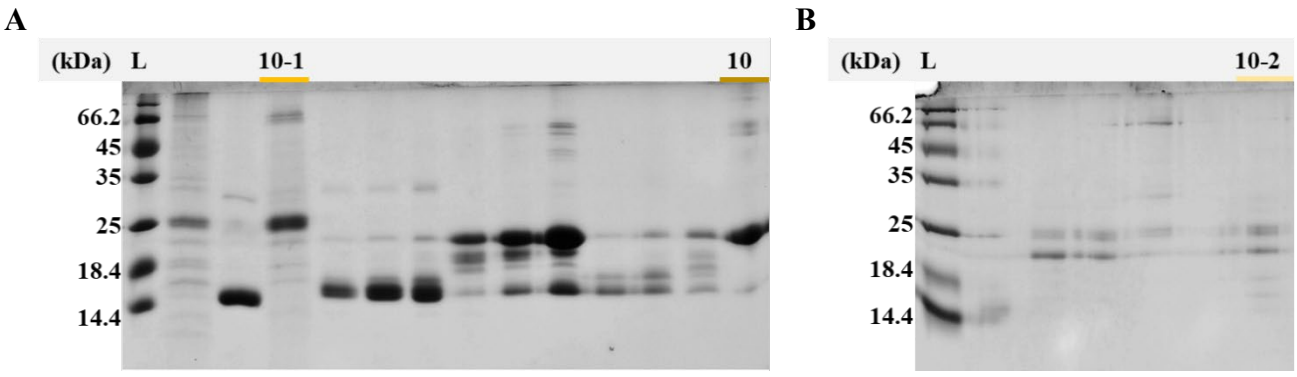

Figure S2. SDS-PAGE of post-SEC sdAbs following a week-long storage of (A) S10, (A) S10-1, and (B) S10-2 at -20 ºC. Each sample studied is highlighted and color-coded similarly to Fig 2.3G