Received: 11 January 2024 | Accepted: 14 May 2024

DOI: 10.1002/jmv.29692

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Development of a single-domain antibody to target a G-quadruplex located on the hepatitis B virus covalently closed circular DNA genome

Gerardo B. Figueroa[1] | Simmone D'souza[1,2,3] | Higor S. Pereira[1] | Gunjan Vasudeva[1] | Sara B. Figueroa[1] | Zachary E. Robinson[1] | Maulik D. Badmalia[1,4] | Vanessa Meier-Stephenson[1,2,4] | Jennifer A. Corcoran[2] | Guido van Marle[2] | Yi Ni[5,6] | Stephan Urban[5,6] | Carla S. Coffin[2,3] | Trushar R. Patel[1,2,4]

[1]Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, Lethbridge, Alberta, Canada

[2]Department of Microbiology, Immunology and Infectious Diseases, Cumming School of Medicine, University of Calgary, Alberta, Canada

[3]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[4]Li Ka Shing Institute of Virology, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada

[5]Department of Infectious Diseases, Molecular Virology, University Hospital Heidelberg, Heidelberg, Germany

[6]German Center for Infection Research, Heidelberg University, Heidelberg, Germany

Correspondence

Trushar R. Patel, Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada.
Email: trushar.patel@uleth.ca

Funding information

Canada Research Chairs; Alberta Innovates; Natural Sciences and Engineering Research Council of Canada; Canadian Institutes of Health Research; Diamond Light Source; Canadian Network on Hepatitis C; Cumming School of Medicine, University of Calgary

Abstract

To achieve a virological cure for hepatitis B virus (HBV), innovative strategies are required to target the covalently closed circular DNA (cccDNA) genome. Guanine-quadruplexes (G4s) are a secondary structure that can be adopted by DNA and play a significant role in regulating viral replication, transcription, and translation. Antibody-based probes and small molecules have been developed to study the role of G4s in the context of the human genome, but none have been specifically made to target G4s in viral infection. Herein, we describe the development of a humanized single-domain antibody (S10) that can target a G4 located in the PreCore (PreC) promoter of the HBV cccDNA genome. MicroScale Thermophoresis demonstrated that S10 has a strong nanomolar affinity to the PreC G4 in its quadruplex form and a structural electron density envelope of the complex was determined using Small-Angle X-ray Scattering. Lentiviral transduction of S10 into HepG2-NTCP cells shows nuclear localization, and chromatin immunoprecipitation coupled with next-generation sequencing demonstrated that S10 can bind to the HBV PreC G4 present on the cccDNA. This research validates the existence of a G4 in HBV cccDNA and demonstrates that this DNA

Gerardo Balderas Figueroa and Simmone D'souza have made equal contributions to the work, as such we request for them to be considered co-first authors on this manuscript.

secondary structure can be targeted with high structural and sequence specificity using S10.

KEYWORDS
cccDNA chromatin immunoprecipitation, covalently closed circular DNA, G-quadruplex or guanine-quadruplex, hepatitis B virus, single-domain antibody

1 | INTRODUCTION

Hepatitis B virus (HBV) is a partially double-stranded DNA virus that is a member of the *Hepadnaviridae* family, genus *Orthohepadnaviridae*. Despite the availability of an effective vaccine against HBV, it is estimated that over 296 million people globally are living with chronic HBV infection.[1] HBV is the leading cause of liver cancer and end-stage liver disease, claiming approximately 820 000 lives/year.[1] The HBV life cycle is complex, and many reviews highlight the currently approved and potential antiviral targets in the pipeline.[2–5] Currently approved therapies for the treatment of chronic HBV infection include life-long therapy with nucleos(t)ide analogs (NAs) and/or a 48-week pegylated (PEG) interferon-α (PEG-IFN- α) treatment course.[6] Overall, NAs can control viral replication, but do not lead to viral clearance due to the persistence of a stable covalently closed circular DNA (cccDNA) episome in the nucleus of infected hepatocytes.[3,7–9]

Transcription factors regulate gene expression activity on promoter regions of DNA. Previous work by Li and Ou,[10] has described that the human transcription factor Sp1 aids in HBV replication by interacting with the PreCore/Core (PreC) promoter. Sp1 is highly enriched in the liver and is a transcription factor that binds to GC-rich elements.[11] At the atomic level, guanine-rich DNA or RNA sequences can form a secondary structure called a G-quadruplex (G4) which are stabilized through Hoogsteen hydrogen bonds and by monovalent cations.[12] These G4 secondary structures can be bound by G4 binding proteins, and as such they are dynamic structures that can be stabilized and resolved.[13]

In recent years, there have been several studies focused on identifying G4s in the HBV genome and to elucidate their role in the viral life cycle.[14,15] Through a comprehensive biophysical and cell biology approach, our previous research has shown the existence of a highly ordered parallel G-quadruplex (G4) structure in the EnhII/PreC promoter region of the HBV genome and its role in regulating HBV core protein translation.[16] Moreover, researchers have also identified 12 potential quadruplex-forming sequences (PQS), some of which have been attributed to have roles in cccDNA phase separation, as well as regulation of HBsAg and HBcAg production.[16–19] Those findings are consistent with previous reports of G4 involvement in viral replication, transcription, and translation, all of which have highlighted the relevance of G4s as possible therapeutic targets.[20]

Previously, BG4, a single-chain variable fragment (scFV) antibody was routinely used in biochemical assays to target G4s but was unable to be expressed intracellularly.[21] Further development has shown that single-domain antibodies (sdAb) can be engineered against G4s and can be expressed intracellularly to target these structures.[22] Through the development of phage display technologies, libraries containing billions of sdAb clones have allowed researchers to find an antibody for any antigen.[23] Owing to their small size (~14 kDa) and presence of only a fragment antigen binding domain, sdAbs can be cheaply produced using bacterial systems, whilst maintaining similar antigen specificity and affinity as traditional antibodies.[24]

Our goal for this study was to develop a single-domain antibody capable of targeting a G4 structure with high specificity located in the HBV PreC promoter (PreC G4). Utilizing a humanized-sdAb library, we identified a sdAb (S10) that can target the PreC G4. The interaction between S10 and PreC G4 was characterized using comprehensive biophysical and biochemical techniques which show that S10 binds to PreC G4 at a nanomolar affinity, thus forming a stable complex. We also demonstrate that S10 can localize into the cell nucleus, and upon HBV infection can bind to the cccDNA PreC G4 using cross-linked chromatin immunoprecipitation (X-ChIP) paralleled with qPCR and next-generation sequencing. Ultimately, the work shown here demonstrates the ability of sdAbs to be considered as a useful G4-targeting molecule that can be used in future studies to better understand G4s in the context of viruses, including infection with the hepatitis B virus.

2 | RESULTS

2.1 | Screening and selection of sdAbs

To develop a single-domain antibody specifically against the HBV PreC G4, we used a 23nt wild-type HBV PreC G4 DNA oligo as our target of interest and a double mutant sequence (containing two G–A mutations) with disrupted G4 quartets in negative selection (see Materials and Methods in Supporting Information S1), as previously described.[16] Biotinylated wild-type and mutant PreC oligos were folded and used as bait in the SAS hsd2Ab library screen of humanized sdAbs by phage display (Hybrigenics; Evry-Courouronnes). Before each round of positive selection, a negative selection was performed using the double mutant G4. Following three rounds of positive and negative selection, 90 clones were identified. Further screening using non-adsorbed phage enzyme-linked immunosorbent assay (ELISA) and sequencing identified 11 different sdAb sequences that were specific to the WT PreC G4 (Figure 1). The top two strongest binding sdAbs (designated S10 and S11) also had

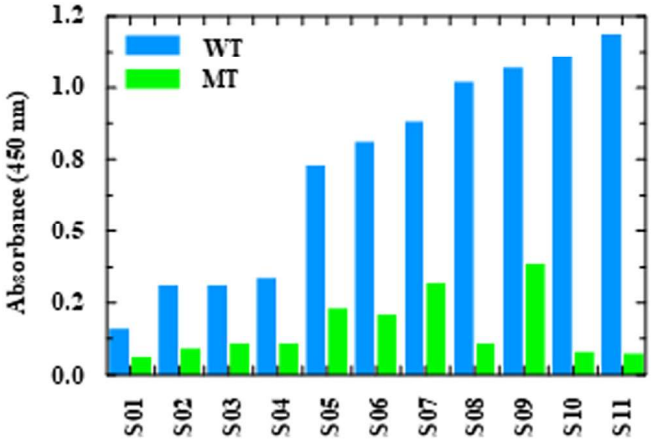


FIGURE 1 Phage-enzyme-linked immunosorbent assay (ELISA) demonstrating binding interaction of the 11 phage display-identified sdAbs against wild-type (WT) preC G4 and mutant (MUT) preC G4. Performed in biological duplicate.

the highest signal-to-noise ratio (WT PreC G4: mutant PreC G4) and thus were selected for further characterization. Moreover, the weakest binding sdAb (S1) was also selected as a weak binding control for further studies (Figure 1).

2.2 | Production, purification, and quality check of sdAbs

We recombinantly produced S1, S10, S11, and BG4 (known as a G-quadruplex binding scFv) to determine the binding affinity to HBV PreC G4.[25] *Escherichia coli* Lemo21(DE3) cells were transformed with the respective plasmids, and these proteins were overexpressed using 1 mM IPTG and purified through Ni-NTA affinity and size exclusion (SEC) chromatography (Supporting Information S1: Figure S1A). Following SEC, proteins larger than approximately 30 kDa were not observed in sodium dodecyl sulfate-polyacrylamide gel electrophoresis indicating a clean protein preparation (Supporting Information S1: Figure S1A). Additionally, circular dichroism (CD) spectroscopy was used as a secondary structure quality control for S10 folding (Supporting Information S1: Figure S1B) that was estimated to contain a high percentage of β-sheets by K2D3 (16%) and BESTSEL (11% left-twist, 25% relaxed, 64% right-twist) content. This is expected as previous analysis of 90 co-crystallized structures revealed that the framework regions of sdAbs (69% of the sequence residues), where 9 beta-sheets are present, are highly conserved across structures.[26]

2.3 | Interactions of sdAb with G4-forming sequences

We used Microscale Thermophoresis (MST) to assess the binding strength of the sdAbs and BG4 scFv to PreC G4. A constant concentration of 3′FITC labeled HBV PreC G4 and a twofold serial dilution of the antibodies over the course of 16 capillaries were used to obtain a titration curve. All the assays showed no evidence of photobleaching and had signal-to-noise ratios above the threshold of 5, which are parameters used to score data trends as meaningful in their representation of a molecular interaction by Nanotemper.

Under the detection limits of MST, S1 did not interact with PreC G4, and the scFv BG4 interacted with PreC G4 with a K_D 6.640 ± 0.189 μM (Figure 2). S10 (K_D 0.091 ± 0.019 μM) outperformed S11 (K_D 2.126 ± 0.143 μM) in binding affinity to PreC G4 by approximately 200 fold. From this experiment, we decided to use S10 as our primary antibody candidate for further investigation and S1 as our weak sequence specific binding control.

Next, we investigated the ability of S10 to differentiate between folded and unfolded forms of the HBV PreC G4 sequence. To do this, we folded and purified the HBV PreC G4 in a LiCl-based buffer (denoted as PreC G4*) to disrupt the formation of the G4 structure. S10's affinity for the PreC G4* (K_D 8.282 ± 1.563 μM) which is approximately a ~800-fold weaker interaction compared to the KCl folded PreC G4. Additionally, the double mutant PreC G4 sequence used for negative selection in phage display was tested in KCl-based buffer to test the interaction between S10 and a disrupted G4 structure (Supporting Information S1: Figure S2). Here, no evidence of an interaction was observed under the parameters of MST.

To determine the ability of S10 to differentiate between sequences, we employed a BlastN search to find comparable 23nt sequences within the human genome.[27] Ras1 (47.8% sequence identity) was chosen because the sequence is known to bind to the human Sp1 protein.[28] CACNA1C (82.6% sequence identity) and FAM9A (65.2% sequence identity), were chosen due to their higher sequence identity to the HBV PreC G4. These sequences were folded as per the same protocol described for the HBV PreC G4 and were subjected to SEC to ensure the presence of monomers (Supporting Information S1: Figure S1C). Once folded, all four oligos (including the PreC G4) were subject to CD spec. As described previously,[16,29,30] all the oligo's spectra were consistent with the presence of parallel G4s as shown by their inflection points at 210, 240, and 260 nm (Supporting Information S1: Figure S1D).

MST results show that S10 has at least a 10-fold stronger interaction with the PreC G4 than to any of the other parallel G4s with different sequences found in the human genome (Figure 2). The binding strength of S10 to different sequences of G4s can be summarized as follows: PreC G4 > CACNA1C > Ras1 > Fam9a (strongest interaction to weakest). The FAM9a and Ras1 data sets did not reach saturation in their binding affinity curve, thereby the estimated affinity strength is an approximation, and the true number may be much weaker. Additionally, apart from CACNA1C's assay, all fall above the S:N ratio (12) considered to be indicative of an excellent assay indicative of interactions.

To evaluate intracellular nuclear targeting of S1 and S10, we modified the residue sequence to include a C-myc derived nuclear localization (NLS) signal sequence (PAAKRVKLD) and HA-tag

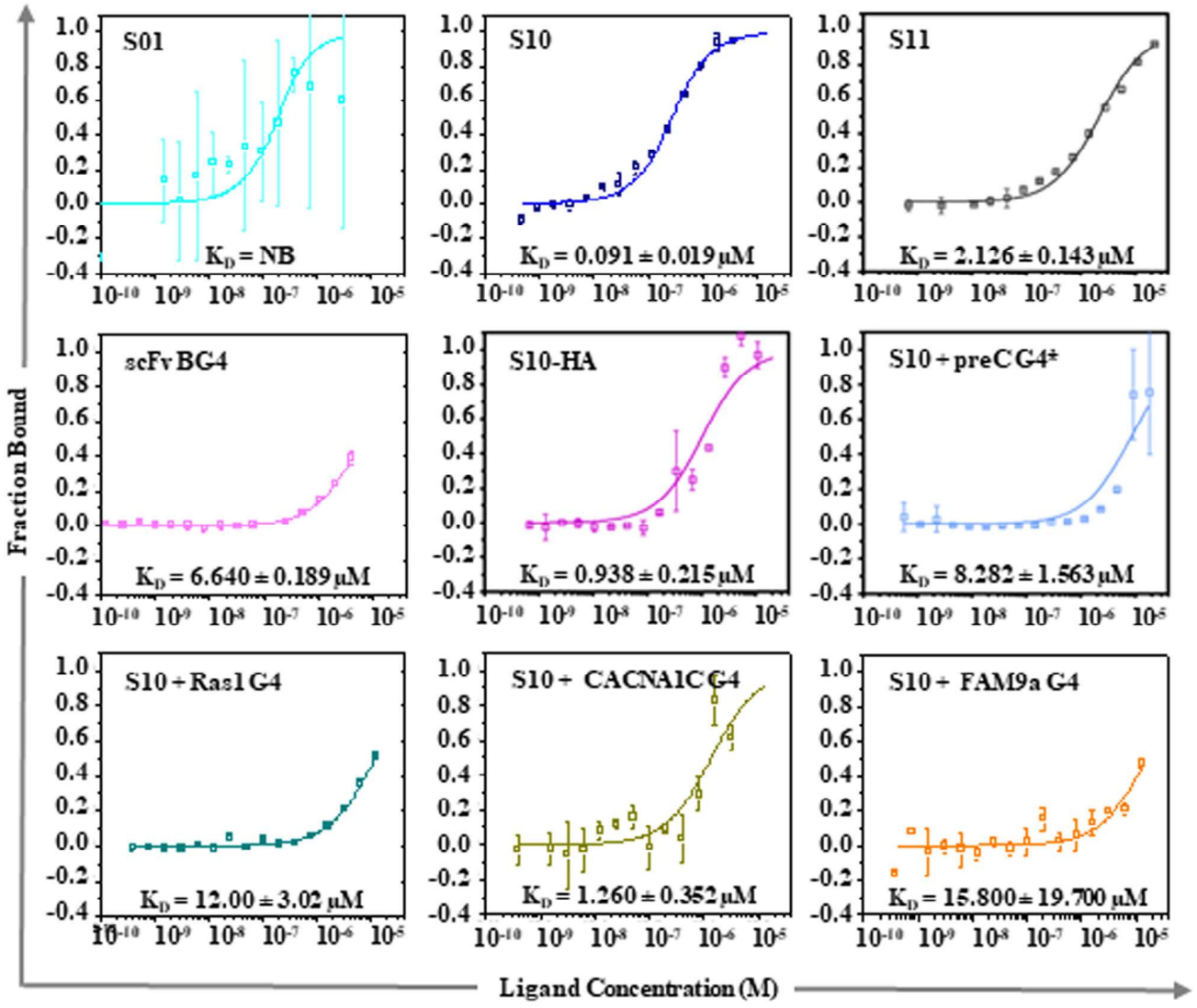


FIGURE 2 Binding interactions between G4-binding antibodies (S1, S10, S10-HA, S11, and BG4) and G4 sequences (PreC G4, Ras1 G4, CACNA1C G4, and FAM9a G4) as characterized by Microscale Thermophoresis (MST). The top two rows demonstrate protein binding interactions with HBV preC G4, and the bottom row displays the interactions between S10 and other G4 sequences found in the human genome. The Y-axis represents the fraction of DNA bound to the antibody, and the X-axis represents the concentration of antibody present in each capillary. S10 + preC G4* binding assay was characterized in lithium-based buffer. MST performed in biological triplicate, mean ± stdeva.

(YPYDVPDYA) on the N-terminus and are labeled as S1-HA and S10-HA. The addition of the tags negatively impacted the binding of S10-HA (K_D 0.938 ± 0.215 µM) to the PreC G4 by 10-fold, however, the binding strength was still stronger compared with S1 and S11 (Figure 2).

2.4 | Size exclusion chromatography-multi-angle light scattering-dynamic light scattering (SEC-MALS-DLS)

We performed SEC-MALS-DLS to determine the binding stoichiometry of the S10–PreC G4 complex. All species studied in Figure 3A were found to be monodispersed having a polydispersity ratio (Mw/Mn) of 1.00. As observed in Figure 3B, peak 1 displays the hydrodynamic radius (R_H) of S10 (27.48 ± 0.76 Å). Based on the R_H similarity between peak one and peak 3 (27.83 ± 0.88 Å), we infer that peak 3 is likely to be unbound S10. The R_H for peak 2 is 36.42 ± 0.127 Å, which is larger than the R_H calculated for S10, indicating a complex is formed between S10 and the PreC G4. The absolute molecular weight of S10 in Figure 3A (left panel) was calculated to be 17.97 ± 0.99 kDa when S10 was run by itself, which is close to the predicted value of 17.68 kDa by ExPasy Protparam. When S10 and PreC G4 (Figure 3A, right panel) were incubated together and then subject to SEC-MALS-DLS, two peaks were identified—20.33 ± 0.57 kDa and 29.57 ± 1.21 kDa. The difference in molecular weight between complex and S10 was

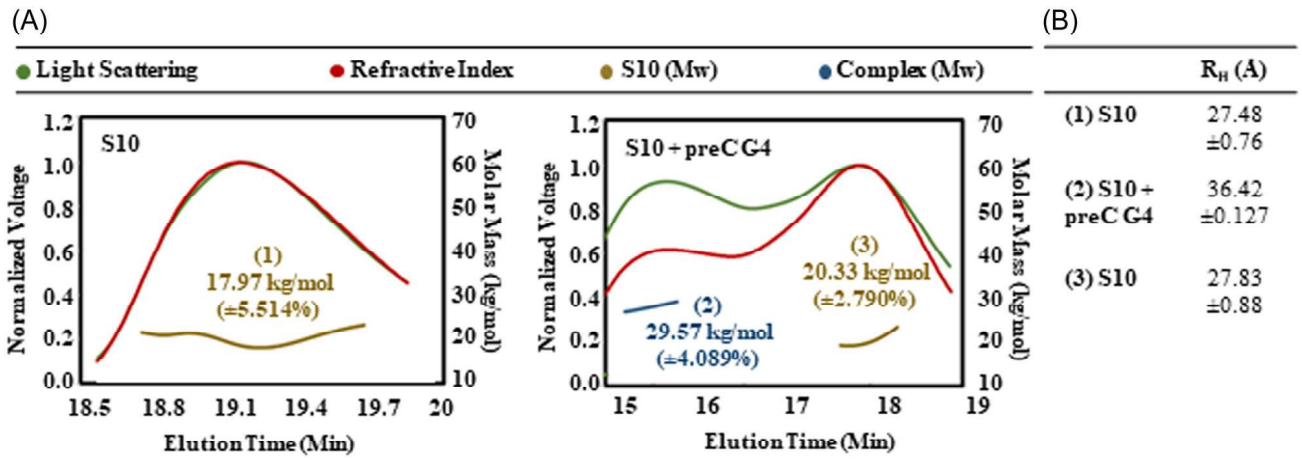


FIGURE 3 (A) SEC-MALS performed with S10 alone (left panel) and S10 incubated with PreC G4 (right panel). Refractive index and light scattering were used to determine the molecular weight and monodispersity of S10 and the S10-preC G4 complex. (B) SEC-DLS was used to determine the hydrodynamic radius (R_H) of each species found in each peak found in panel (A).

estimated to be 9.24 kDa. This closely approximates the expected size of PreC G4 (7.4 kDa) and suggests a complex with 1:1 stoichiometry.

2.5 | Small-angle X-ray scattering (SAXS)

From MST and SEC-MALS-DLS we demonstrated that a complex between S10 and PreC G4 is formed, and it can be purified via size exclusion chromatography. To determine how the complex appears in solution, we performed HPLC-SAXS (high-performance liquid chromatography coupled small angle X-ray scattering) to gather information on size, shape, and electron density envelope for S10 and S10-PreC G4 complex. HPLC-SAXS was collected by the B21 BioSAXS beamline at the Diamond Light Source (Didcot) synchrotron facility, as previously described.[31]

The S10 and the S10-PreC G4 complex had enough light scattering intensity to proceed with the SAXS analysis. The homogeneity and presence of aggregates in the samples were assessed by analyzing the data points in the Guinier region of the scattering profile that fall within the regression line (Figure 4A,B). The scattering pattern of all samples (S10, PreC G4, and complex) are indicative of having high homogeneity and being free of aggregates since their scattering falls within the regression line.[32] S10's Kratky plot is consistent with a tightly folded protein (Figure 4A) since it follows a bell-shaped trend as characteristic from folded proteins.[33] Although the plot for the S10-PreC G4 complex follows a similar trend, it also depicts an increase in flexibility and a decrease in foldedness as the data pattern increases in the maximum scattering angle detected. Both data sets have an electron distance pair-distribution function plot (Figure 4A,B) where a relatively globular concentration of electrons is found from 0 to 40 Å for S10 with a tail sticking out at one end of the plot (right trailing shoulder). The complex's trailing tail extends past S10's by 33.05 Å. Following DAMMIN and DAMAVER modeling, the relative shape of S10 (Figure 4C) resulted in a relatively tightly folded protein with a smaller tail.

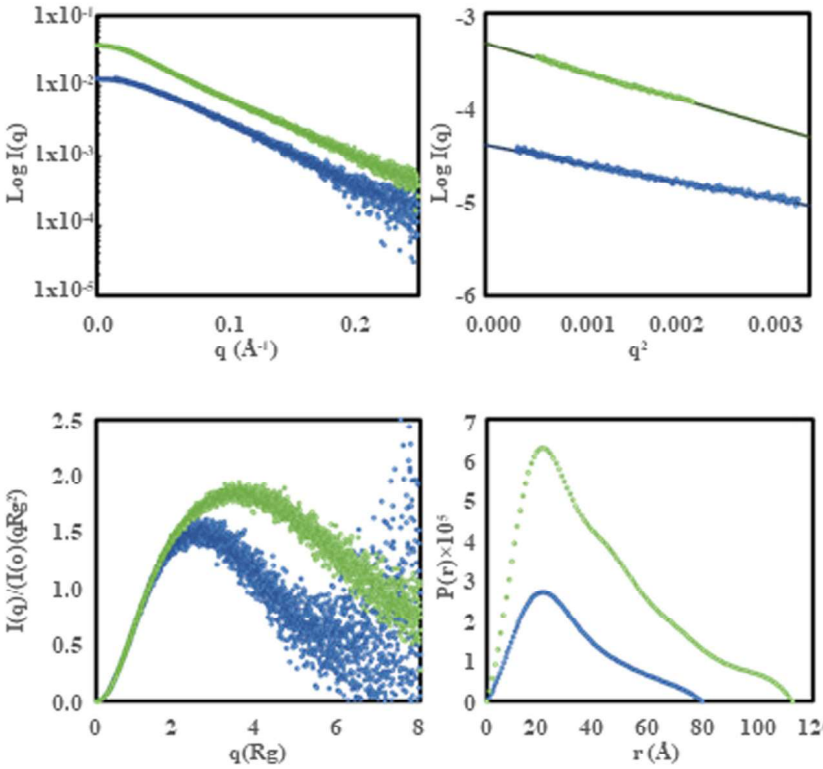
2.6 | Generation of cell lines expressing single-domain antibodies

Following the biochemical and biophysical characterization of S10 and the S10-PreC G4 complex, cell culture experiments were employed to determine if the complex can be identified intracellularly during HBV infection. HepG2-NTCP-A3 cells are susceptible to HBV infection, and as such were transduced to create stable cell lines expressing S1-HA (weak sequence specific PreC G4 binding control), S10-HA (strong sequence and G4 structure specific PreC G4 binder), and an empty vector (negative control). Intracellular expression of S1-HA and S10-HA was confirmed by Western blot (Figure 5A) and immunofluorescence (Figure 5B) against the HA tag. The Western blot (Figure 5A) shows that S1-HA and S10-HA can be produced inside HepG2-NTCP-A3 cells and match the expected molecular weight of approximately 15 kDa. Immunofluorescence demonstrates that S1-HA and S10-HA can localize into the nucleus (Figure 5B), thus confirming the potential ability to target the intranuclear HBV cccDNA reservoir.

2.7 | cccDNA cross-linked chromatin immunoprecipitation (X-ChIP)

To investigate whether intracellularly expressed S10-HA binds to the HBV cccDNA PreC G4, we performed cross-linked cccDNA chromatin immunoprecipitation. Tape station DNA quality analysis was performed to ensure successful fragmentation of intranuclear DNA (Supporting Information S1: Table 1), and the average

(A)



(B)

Data collection parameters	preC G4[1]	S10	S10-preC G4
I(0) from P(r)	0.0034 ±0.7x10⁻⁵	0.01245 ±1.1x10⁻⁵	0.03804 ±5.3x10⁻⁵
I(0) from Guinier plot	0.0035 ±1.1x10⁻⁵	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
D_{max} (Å)	46.85	79.22	86±5
Molecular Weight (kDa)	7.3*	21.93	28.85
χ2	~1.80	~1.09	~1.13
NSD	0.55 ±0.01	0.58 ±0.02	0.59 ±0.01

(C)

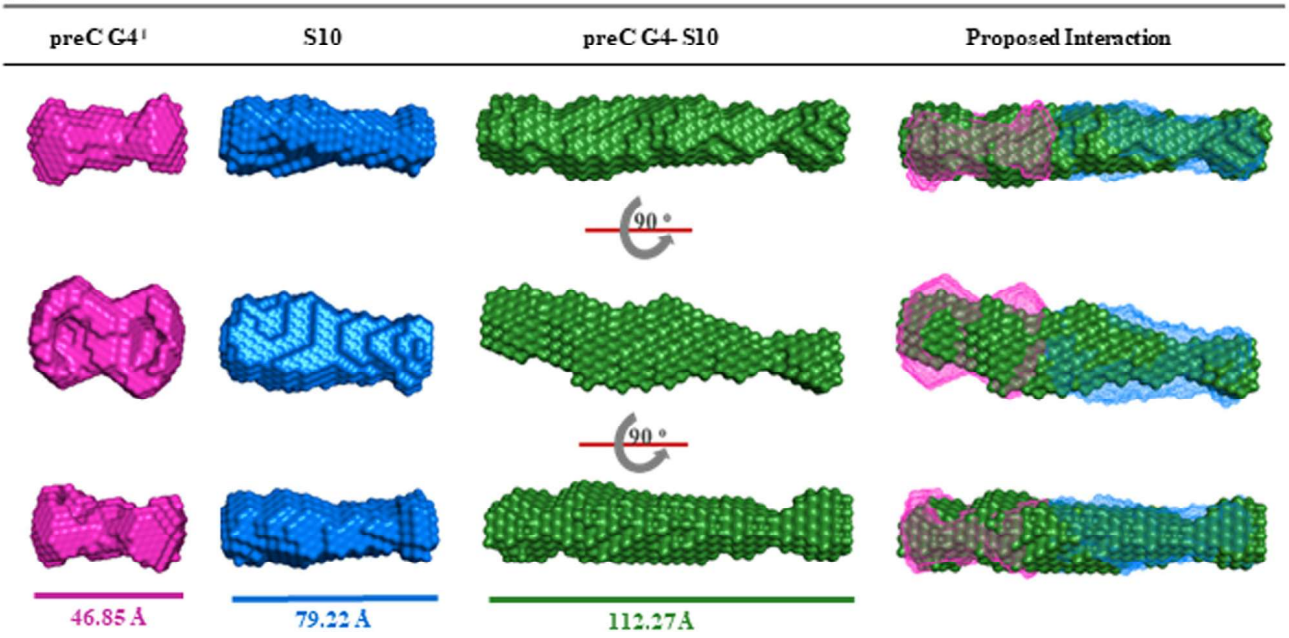


FIGURE 4 Small Angle X-ray scattering of data of S10 and S10 + preC G4. (A) The data is shown (from left to right) as raw scattering data, in a Guinier plot, in a dimensionless Kratky plot, and an electron pair-distance distribution function plot. (B) Low-resolution models were calculated in DAMAVER and DAMMIN software. The overall shape and dimensions of the models were used to propose the nature of the S10–PreC G4 interaction. (C) Summary of the analyzed scattering data for each model used. X^2 and NSD were derived from DAMAVER and DAMAVER analysis. (B–C)[1] preC G4 is a previously calculated SAXS model as seen in Meier-Stephenson et al.,[16] where the same 23nt G4 DNA used in the complex was studied.

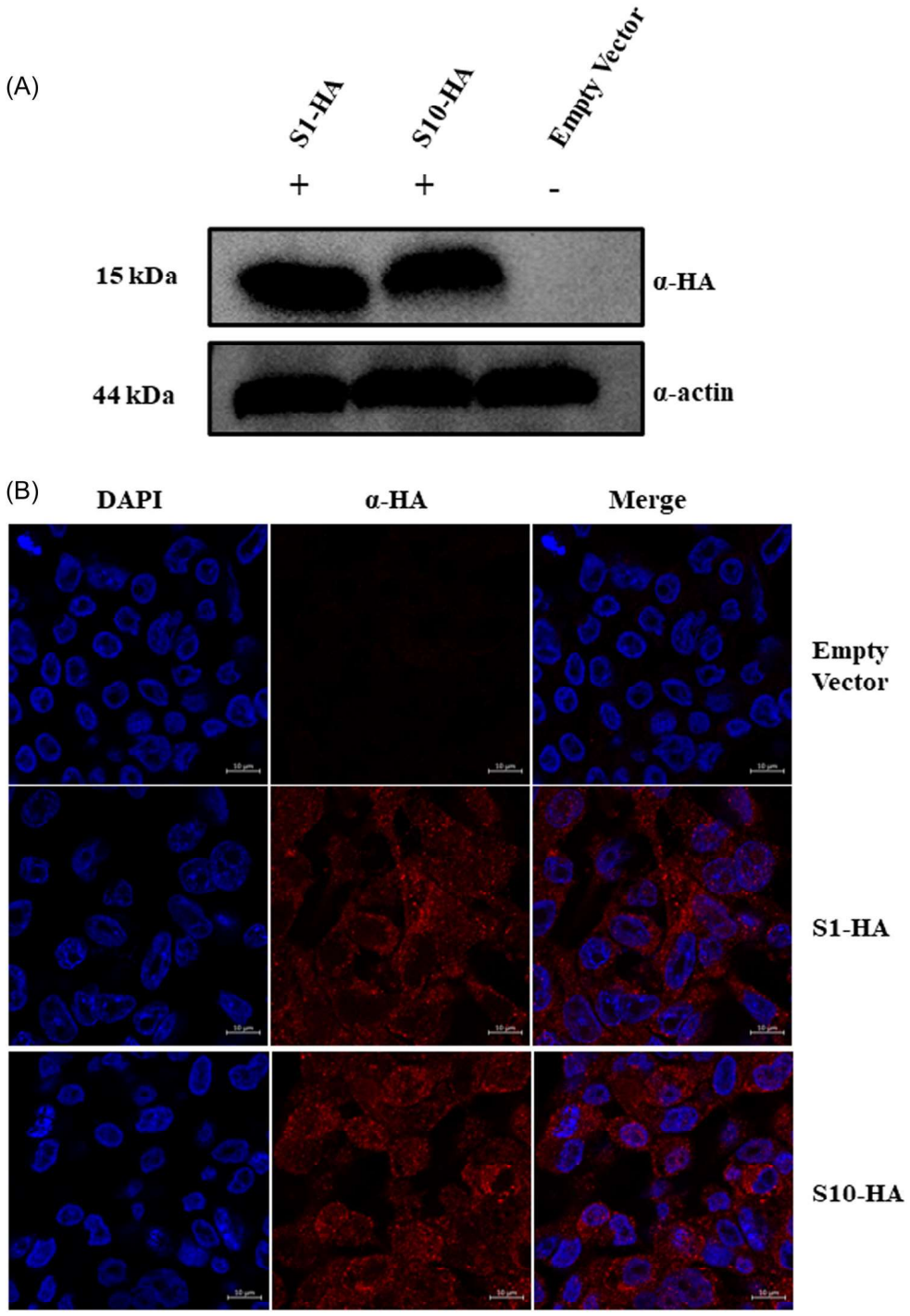


FIGURE 5 (A) Western blot showing intracellular protein expression of S1-HA and S10-HA in stably transduced HepG2-NTCP-A3 cells. (B) Immunofluorescence staining of S1-HA and S10-HA was performed using an antibody against the HA tag (red) and nuclei were stained with 4′, 6-diamidino-2-phenylindole (DAPI) (blue). Confocal microscopy (63x) imaging shows colocalization of S1-HA and S10-HA within the cell nuclei.

fragmentation size for the samples was approximately 1400 bp. Following immunoprecipitation, quantitative real-time polymerase chain reaction (qPCR) was performed on three loci: (i) the actin transcription start site (TSS)[28] to determine if the immunoprecipitation was successful; (ii) the 735 bp product spanning the gap region of cccDNA to determine if the HBV cccDNA is targeted and precipitated,[29,30] and (iii) a 160 bp product spanning the PreC/C promoter (nt position 1671–1831) containing the PreC G4 sequence of interest.[31]

Since S1-HA and S10-HA were designed against the PreC G4 sequence on the HBV genome, the actin TSS locus served as a negative control to ensure that these single-domain antibodies have some target specificity. Through ChIP-qPCR (Figure 6A) we show target specificity through extremely low enrichment on the

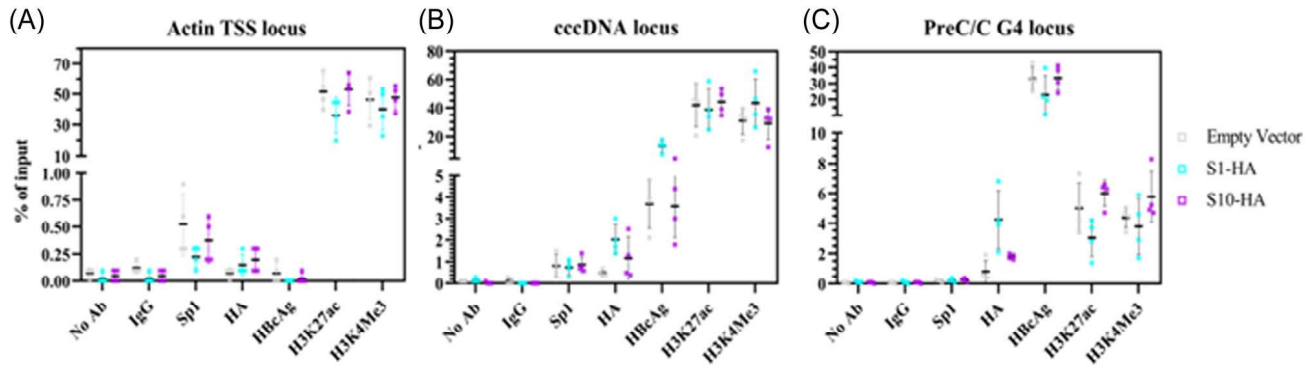


FIGURE 6 Cross-linked chromatin immunoprecipitation (ChIP) coupled with quantitative real-time polymerase chain reaction (qPCR) in HepG2-NTCP-A3 cells stable transduced with S1-HA, S10-HA, or vector only. ChIP-qPCR has been performed over the (A) actin transcription start site (TSS) locus, (B) covalently closed circular DNA (cccDNA) locus, and (C) PreC C G4 locus to determine target enrichment from each immunoprecipitation (four biological replicates).

human Actin TSS locus for the following proteins: HBcAg, S1-HA, and S10-HA. However, there is a slight target enrichment for Sp1 and definite binding of H3K27ac and H3K4Me3 on the actin TSS locus—consistent with literature findings.[28,32]

To determine if cccDNA can be immunoprecipitated, qPCR spanning the gap region of HBV (Figure 6B) was performed to selectively enrich nuclear cccDNA instead of the partially double-stranded relaxed circular HBV DNA predominantly found in the cytoplasm of cells.[30,33] ChIP-qPCR has been previously performed to immunoprecipitate HBV cccDNA by targeting HBcAg, H3K27Ac, and H3K4Me3, and results in Figure 6B are consistent with these findings.[17,34,35] Since H3K27Ac and H3k4Me3 are histone markers that correlate with active transcription, we can infer that actively transcribed HBV cccDNA was immunoprecipitated.[28] G-quadruplexes are typically found in transcriptionally active states within double-stranded DNA, thus the presence of actively transcribed cccDNA increases the likelihood of G4 complex identification.[34] In line with previous studies, we demonstrate that Sp1 can bind and immunoprecipitate HBV cccDNA, verifying that Sp1 does interact with the HBV cccDNA genome.[10,17,36–38] Furthermore, a comparison of the empty vector cell line control to both S1-HA and S10-HA, showed that both single-domain antibodies can bind to HBV cccDNA.

After demonstrating that the targeted proteins bind to the HBV cccDNA, we investigated whether these proteins recognize the PreC region. To do so, we performed qPCR (Figure 6C) covering 160nt of the PreC/Core promoter of HBV. As expected, due to the chromatinized state of the HBV cccDNA genome, we observed enrichment in H3K27Ac and H3K4Me3 histone markers.[28] HBcAg can also bind and immunoprecipitate the HBV genome in this location and is consistent with previous studies showing that HBV core strongly associates with HBV cccDNA.[30] For Sp1, the level of enrichment (determined by qPCR) in the empty vector control cell line, S1-HA, and S10-HA is minimal and comparable with the background rabbit IgG isotype control. There are other Sp1 binding sites located on the HBV cccDNA genome, detected by qPCR[10]; however, there is extremely low enrichment for the site containing the HBV PreC G4. Interestingly, S1-HA, and S10-HA were also able to bind to the PreC G4 locus. The binding of S1-HA and S10-HA on the PreC G4 locus indicates that there is a binding site in the 160 bp region recognized by these single-domain antibodies and that the G4 structure is likely to be present in transcriptionally active HBV cccDNA.

2.8 | cccDNA ChIP-seq

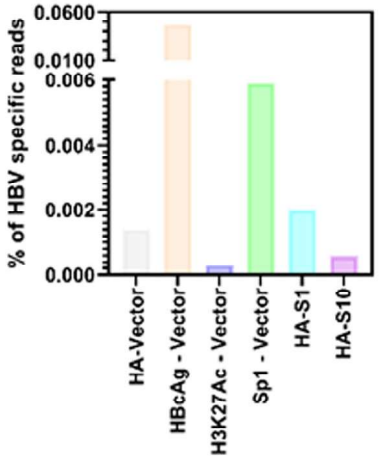
The data obtained from ChIP-seq (Figure 7) was used to assess the binding distributions on the HBV and human genome for HBcAg, Sp1, S1-HA, S10-HA, and H3K27ac. The blasticidin transduced vector-only cell line pulled down with the HA antibody was used as the background and negative control since this cell line does not express any HA-tagged proteins. Moreover, immunoprecipitations performed from the vector-only cell line for HBcAg, Sp1, and H3K27ac were used for ChIP-seq as this cell line was used as a transduction control. Immunoprecipitations against the HA tag for S1-HA and S10-HA cell lines were used to evaluate the binding distribution of S1 and S10, respectively.

There is limited data describing ChIP-seq for HBV cccDNA,[28] however one single study reported that higher sequencing depth is required to obtain significant HBV-specific reads in cccDNA-ChIP-seq. Despite performing 20 million reads in the current study, a higher number of human genome-specific reads compared to HBV-specific reads was found in all assessed samples (Figure 7A,B). Density plots (Figure 7C) were prepared in IGV and aligned to the annotated HBV genotype D (accession number V01460.1). Detailed.bam read coverage plots which provide a comprehensive binding distribution of each IP protein are available in the supplemental (Supporting Information S1: Figures 3–8). The anti-HA immunoprecipitation background control in the vector-only cell line displays a relatively large number of reads corresponding to the small HBsAg region (Supporting Information S1: Figure S3). Moreover, the HBc protein appears to have diffuse binding throughout the HBV genome (Supporting Information S1: Figure S4), which is consistent with what

(A)

Target	Total Reads	Hg19 Alignment	HBV Alignment
HA - Vector	68195659	50844635	979
HBcAg - Vector	49137117	31727819	22935
H3K27Ac - Vector	73057051	46726538	199
Sp1 - Vector	16136876	5169977	954
HA-S1	66872427	28532763	1331
HA-S10	68829980	46311944	389

(B)



(C)

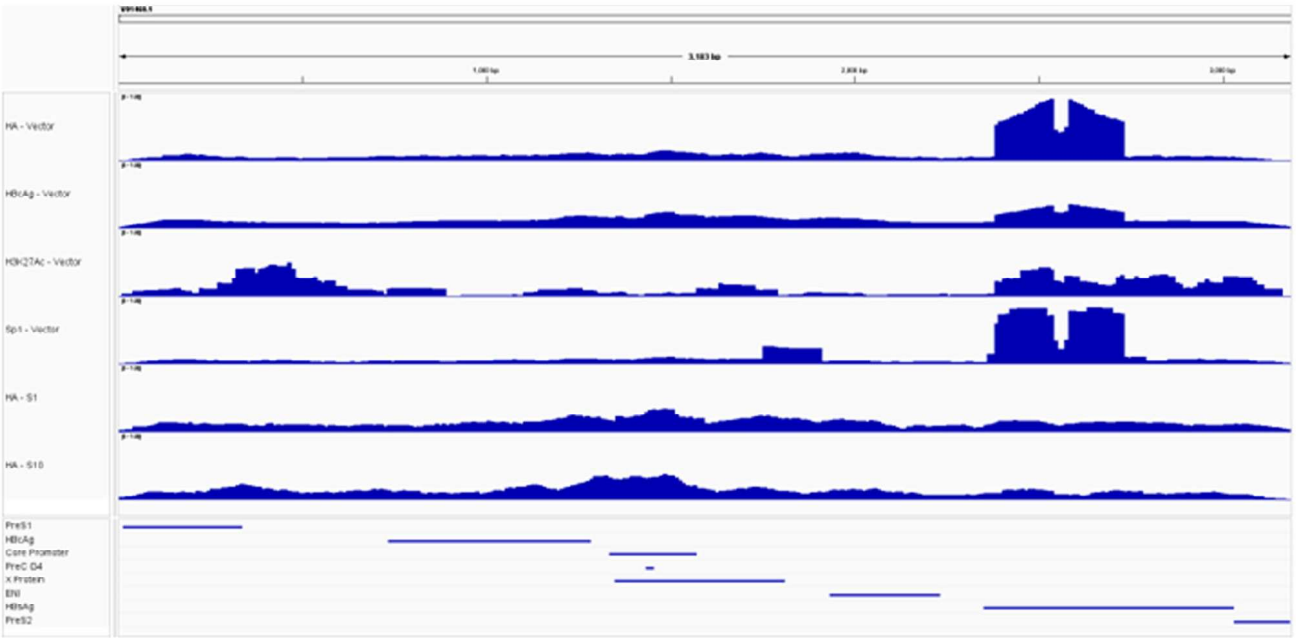


FIGURE 7 Cross-linked chromatin immunoprecipitation (ChIP) coupled with next-generation sequencing in HepG2-NTCP-A3 cells stable transduced with S1-HA, S10-HA, or vector only. ChIP-seq was performed from one biological replicate of S1-HA, S10-HA, and Vector control-SP1, HBcAg, H3K27Ac. Sequencing was performed on the Illumina NovaSeq. 6000 platform with 2 × 150 Paired End (PE) reads. (A) Distribution of reads aligning to human genome 19 and HBV (Genbank ID V01460.1) for each IP. (B) Percentage of hepatitis B virus (HBV)-specific reads for each IP compared to the total number of reads in each sequencing run. (C) Density plots displaying the binding distribution of each protein on the HBV genome. The density plots were generated from.bw files in IGV browser to show the distribution of protein binding to the HBV genome.

is observed in the literature.[34,35,39] Nucleosome binding of H3K27ac (Supporting Information S1: Figure S5) to cccDNA appears to be also consistent with literature findings—observed enrichment within the transcription start site of X between ENHI and ENHII, as well as PreS1, PreS2, and HBsAg regions.[35,39] Sp1 (Supporting Information S1: Figure S6) binding appears to localize to the HBsAg region and the X-gene regions. To corroborate our experiment, we determined the fraction of reads mapped to the PreC G4 region are more significant in the case of S10-HA than S1-HA, thus confirming the chromatin enrichment for the S10-HA specific library. S1-HA (Supporting Information S1: Figure S7) appears to have diffuse binding throughout the HBV genome, with slightly stronger enrichment in the PreC region. S10-HA (Supporting Information S1: Figure S8) has a larger proportion of reads mapping to the PreC region when compared to S1-HA, indicating that there is a stronger sequence specificity associated with the S10 single-domain antibody. Moreover, from the IGV snapshot (Figure 7C and Supporting Information S1: Figures 3–8), we show that S10-HA has a higher specificity for the PreC G4 compared to any other protein that was immunoprecipitated. Overall, the cccDNA-ChIP-seq assay validates

the biophysical data, and indicates that there is a sequence and structural specificity in S10s recognition of the HBV cccDNA genome.

3 | DISCUSSION

The existence and importance of G4 structures in the human genome have been extensively described and have implications for a variety of diseases.[39] Recently there has been increasing interest in identifying G4s in viruses to validate their existence and understand their importance during viral infection.[40] Previously, researchers have developed a variety of small molecules like TMPyP4 and Thioflavin which can target G-quadruplexes.[40] These have provided valuable tools to validate the existence of these structures and characterize their function. Macromolecules have also been developed to broadly target G-quadruplex structures as is the case with antibodies such as BG4 scFv and SG4 sdAb[16,19] or L-RNA aptamers such as L-Apt12-6.[41] However, in the context of cellular work or higher-order organisms, the use of these G4 ligands presents challenges with off-target interactions. Discriminating between different G4s is complicated and in the case of the human genome, the number of described and undescribed G4 structures is high.[42] Furthermore, the complete biochemical pathways affected by G4-structures have not been fully described, but it has been shown that they occupy various jobs that include transcription repressing and promoting.[43] Phage display technology has been previously exploited where negative and competitive selection through ssDNA, dsDNA, yeast tRNA, salmon sperm DNA, and random primer ssDNA.[22] This negative screening procedure selectively removed sdAb clones that would nonspecifically bind to other noncanonical DNA structures, allowing the development of a broad G4 binding sdAb. In our work, we introduced two base substitutions (G–A) to the Pr eC G4 sequence to destabilize the G4 structure[20] and increase the stringency of negative selection during phage display (Figure 1). As a result, we could identify sdAbs with a high sequence and structure specificity to HBV preC G4.

Our biochemical interaction studies confirm that S10 requires a stable G4 structure to be present for binding to occur (Figure 2). As expected from phage display screening, upon exposure to the destabilized G4 structure of DM preC G4, under the parameters of MST, no binding activity by S10 was detected. We tested the G4 structure detection and increased the stringency by subjecting the PreC G4 sequence to a G4 structure-destabilizing environment (LiCl-based buffer). Our findings confirm that the use of DM PreC G4 as a negative selection agent in phage display successfully allowed us to identify a G4-binding sdAb. Furthermore, S10 exhibited a preference for PreC G4 sequence where at least a 14-fold decrease in binding strength was shown when tested against human G4 sequences. In the case of SG4, a variety of parallel G4s can be bound with a nanomolar affinity between approximately 2–80 nM.[22] Interestingly, S10 also exhibits a low nanomolar range in its binding affinity to PreC G4, and unlike the case of SG4, the G4 structure and sequence are integral for target binding. S10's performance in our binding assays suggests that sdAb technology can be further exploited to target other viral G4 sequences of interest.

Our results from SEC-MALS-DLS (Figure 3) and SAXS (Figure 4) confirmed the stoichiometry (1:1) of the interaction between S10 and PreC G4 based on the experimental molecular weights, where the addition of the individual molecules closely approximated the expected weight. The Guinier plots and the polydispersity index in SEC-MALS-DLS suggests that the interaction between S10 and PreC G4 does not promote the formation of aggregates and the bound or unbound molecules exist monodispersed. Interestingly, S10 and the complex share observable trends in their Kratky and the electron pair-distribution plots where the main differences are in terms of their elongation and level of folding. Our group has previously published a low-resolution SAXS model of PreC G4.[16] As seen in Figure 4C, we have shown here that the D_{max} of the PreC G4 model (46.85 Å) closely approximates the remainder of the complex's D_{max} once S10's D_{max} has been subtracted. Based on this knowledge, we overlayed the complexed model and individual models of S10 and PreC G4 to propose a head-to-head binding configuration (Figure 4C). The sdAb structure and sequence are highly conserved, apart from the three hypervariable loops known as complementary determining regions (CDRs) that are responsible for the sdAb's antibody properties (specificity and affinity).[44] Given that the sdAb CDRs face one side of the molecule and are required for an interaction to occur, we suggest that the left side of the sdAb model (Figure 4C) contains these regions as this is the contact site for PreC G4. When it comes to the structure and orientation of PreC G4 once bound, it is difficult to draw conclusion due to the low-resolution modeling approach used. However, based on the dumbbell-like shape of the G4, it can be inferred that the interaction between S10 and PreC G4 occurs at one of these two ends of the molecule and structural changes are not expected based on the interaction studies which show the G4 structure is required for identification. Our findings from the biophysical characterization provide further evidence that S10 can form stable complexes with PreC G4 and alternative structural or monodispersity changes are unlikely.

Developing and identifying agents that target cccDNA has been a major goal in the field. The majority of research has focused on identifying host and viral proteins that interact with cccDNA, to better understand potential protein targets involved in repair and maintenance.[44] To target the intranuclear cccDNA, we re-engineered the single-domain antibodies to contain an HA tag (for detection) and an NLS tag for nuclear localization. The addition of these tags validated the intranuclear expression of S1 and S10 (Figure 5). Since cccDNA is a difficult molecule to study due to the low abundance in hepatocytes (1–10 copies),[45] there is a limited amount of research involved in chromatin immunoprecipitation assays coupled with either qPCR or next-generation

sequencing.[17,34,35,39] From the X-ChIP-qPCR experiments (Figure 6) we show that both S1-HA and S10-HA can bind to the PreC G4 locus and cccDNA locus. As a complementary tool to ChIP-qPCR, next-generation sequencing was performed to identify the exact binding regions on the HBV cccDNA genome that were enriched during immunoprecipitation. S10-HA appears to have a strong binding to the PreC G4 region (Figure 7C and Supporting Information S1: Figure S8), which is in line with our biochemical analysis and phage-display screening. Moreover, S1-HA also appears to have more diffuse binding to the HBV genome but can bind to the PreC G4 region. Within the parameters of the biophysical MST experiments (Figure 2), we did not observe S1 binding to the PreC G4 region. However, the phage display ELISA (Figure 1) shows that S1 indeed does recognize the PreC G4 sequence but is unable to differentiate with high specificity from the mutant G4 sequence. In line with the data from the phage display experiments and ChIP-seq, we infer that S1 can likely recognize the G4 structures, albeit with lower sequence specificity to the PreC G4 when compared to S10. In both S1 and S10, we see prominent binding within the PreC G4 region; however, there is more diffuse binding throughout the HBV genome for S1 (Figure S7 and Supporting Information S1: Figure S8). A recent study identified 12 potential quadruplex-forming sequences on the HBV genome, providing a potential site for S1 to recognize these other sequences (Supporting Information S1: Figure S9).[17] Although these single-domain antibodies have been engineered to study G4s in HBV cccDNA, many of the sequences can also be found in pregenomic RNA, and future studies will investigate if these antibodies are specific to cccDNA or can recognize RNA G4s in HBV (Supporting Information S1: Figures S9 and S10).

4 | CONCLUSION

In summary, we show for the first time that a sdAb can be developed to target a viral G-quadruplex structure on the hepatitis B virus cccDNA through phage display technology. We have generated a pipeline for bacterial overexpression and purification of high levels of these sdAbs. Thorough biochemical and biophysical characterization of S10, we demonstrated S10's ability to stably interact with preC G4 DNA. Other sdAbs (SG4) have been identified as G4-structure binding molecules;[22] however, this is the first account where a sdAb has been demonstrated to have a high degree of sequence specificity and structural recognition. Through the addition of NLS and HA tags to S10, we have validated the intracellular expression of S10-HA and its localization to the nucleus within the HepG2-NTCP-A3 human hepatocyte cell line. Consistent with biochemical assays, ChIP-qPCR and NGS studies demonstrated that S10-HA can immunoprecipitate the PreC/C promoter locus containing the PreC G4 and HBV cccDNA. These studies experimentally confirm our previous work identifying the presence of G-quadruplex structures within the HBV PreC/Core promoter[16] and provide an approach for G4 targeting. The methodology used in this work exhibits a valuable platform to develop single-domain antibodies that can be adopted to investigate the presence and relevance of specific G-quadruplexes intracellularly during viral infections. Our pipeline is advantageous from other G4 targeting approaches as its high specificity allows for the study of key G4 structures of interest, whereas other biomolecules (SG4 and BG4)[22,25] and small molecules (i.e., Braco-19 and Pyridostatin)[40] indiscriminately bind to all G4 structures they encounter. Lastly, our work is relevant for future studies targeting the stable hepatitis B virus intranuclear reservoir, cccDNA, and the development of a sterilizing cure for HBV infection.

AUTHOR CONTRIBUTIONS

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Gerardo Balderas Figueroa and Simmone D'souza contributed equally. Trushar R. Patel and Carla S. Coffin conceptualized the study. Gerardo Balderas Figueroa and Simmone D'souza conceptualized the study, designed, and completed the experiments, analyzed the data, and drafted the article. Higor Sette Pereira and Gunjan Vasudeva performed data analysis for SAXS and ChIP-seq, respectively. Sara Balderas Figueroa and Zachary E. Robinson assisted in biophysical characterization experiments. Maulik D. Badmalia, Vanessa Meier-Stephenson, Jennifer A. Corcoran, Guido van Marle, Yi Ni, and Stephan Urban helped with study design. All authors contributed to the article preparation. Carla S. Coffin and Trushar R. Patel were responsible for the conduct and coordination of the study and served as guarantors of the work.

ACKNOWLEDGMENTS

We would like to thank Dr. Sabine Gilch for providing access to the lsm 700 confocal microscope for imaging. We also thank Dr. Julie Lucifora and Mariel Kleer for their helpful discussions in executing ChIP experiments. We would also like to thank Milan Heck for helping with revisions. The graphical abstract was created with "biorender.com." Carla S. Coffin and Trushar R. Patel would like to acknowledge the Alberta Innovates and University of Calgary Cumming School of Medicine, Clinical Research Fund. Trushar R. Patel acknowledges Canada Research Chair, Canada Foundation for Innovation, NSERC RTI, DIAMOND Light Source, and Univerity Health Foundation Support. Simmone D'souza acknowledges funding from The University of Calgary Cumming School of Medicine Graduate Student Scholarship and Eyes High Doctoral Recruitment Scholarship, Canadian Institute of Health Research Canada CGS-M, CGS-D, and Michael Smith Foreign Study Supplement, The Canadian Network on Hepatitis C (CanHepC) Masters, Doctoral, and TRR179 Fellowships. Vanessa Meier-Stephenson acknowledges CanHepC Post-doctoral fellowship funding. CanHepC is funded by a joint initiative of the Canadian Institutes of Health Research (NPC-178912) and the Public Health Agency of Canada. Higor Sette Pereira is funded by the

NSERC-CREATE PDF Fellowship. Zachary E. Robinson received an NSERC USRA Fellowship.

CONFLICT OF INTEREST STATEMENT

Dr. Trushar Patel, Dr. Carla Coffin, Dr. Maulik Badmalia, and Dr. Vanessa Meier-Stephenson hold a patent on the sequences of the single-domain antibodies (US patent application number 17425791). Other authors who have taken part in this study have declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study will be shared on reasonable request to the corresponding author. The raw reads of ChIP-seq data have been submitted to NCBI under accession number PRJNA1067084.

ETHICS STATEMENT

Ethics approval was obtained from the University of Calgary for the use of cell lines by Conjoint Health Research Ethics Board, CHREB (Ethics ID: REB20-0933).

ORCID

Gerardo B. Figueroa https://orcid.org/0009-0008-3382-642X
Simmone D'souza http://orcid.org/0000-0003-2750-8675
Higor S. Pereira https://orcid.org/0000-0003-4761-8361
Gunjan Vasudeva http://orcid.org/0000-0002-1404-0498
Sara B. Figueroa https://orcid.org/0009-0003-9510-8136
Zachary E. Robinson http://orcid.org/0009-0009-2719-5796
Maulik D. Badmalia http://orcid.org/0000-0002-3384-6550
Vanessa Meier-Stephenson http://orcid.org/0000-0003-4004-4940
Jennifer A. Corcoran http://orcid.org/0000-0002-3764-0218
Guido van Marle http://orcid.org/0000-0002-5148-6229
Yi Ni http://orcid.org/0000-0003-1197-4161
Stephan Urban http://orcid.org/0000-0003-1560-8038
Carla S. Coffin http://orcid.org/0000-0002-1472-0901
Trushar R. Patel http://orcid.org/0000-0003-0627-2923

REFERENCES

1. World Health Organization. Hepatitis B. 2024. https://www.who.int/news-room/fact-sheets/detail/hepatitis-b
2. D'Souza S, Lau KC, Coffin CS, Patel TR. Molecular mechanisms of viral hepatitis induced hepatocellular carcinoma. *World J Gastroenterol*. 2020;26(38):5759-5783.
3. Nassal M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut*. 2015;64(12):1972-1984.
4. Chuang Y-C, Tsai K-N, Ou J-HJ. Pathogenicity and virulence of hepatitis B virus. *Virulence*. 2022;13(1):258-296.
5. Guo J-T, Guo H. Metabolism and function of hepatitis B virus cccDNA: implications for the development of cccDNA-targeting antiviral therapeutics. *Antiviral Res*. 2015;122:91-100.
6. Ye J, Chen J. Interferon and hepatitis B: current and future perspectives. *Front Immunol*. 2021;12:733364.
7. Tsounis EP, Tourkochristou E, Mouzaki A, Triantos C. Toward a new era of hepatitis B virus therapeutics: the pursuit of a functional cure. *World J Gastroenterol*. 2021;27(21):2727-2757.
8. Bhattacharya D, Thio CL. Review of hepatitis B therapeutics. *Clin Infect Dis*. 2010;51(10):1201-1208.
9. Tawada A. Current and future directions for treating hepatitis B virus infection. *World J Hepatol*. 2015;7(11):1541-1552.
10. Li J, Ou J. Differential regulation of hepatitis B virus gene expression by the Sp1 transcription factor. *J Virol*. 2001;75(18):8400-8406.
11. Briggs MR, Kadonaga JT, Bell SP, Tjian R. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science*. 1986;234(4772):47-52.
12. Moon J, Han JH, Kim DY, Jung M, Kim SK. Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin. *Biochem Biophys Rep*. 2015;2:29-35.
13. Huppert JL. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem Soc Rev*. 2008;37(7):1375-1384.
14. Brázda V, Valková N, Dobrovolná M, Mergny JL. Abundance of G-Quadruplex forming sequences in the hepatitis delta virus genomes. *ACS Omega*. 2024;9(3):4096-4101.
15. Pathak R. G-Quadruplexes in the viral genome: unlocking targets for therapeutic interventions and antiviral strategies. *Viruses*. 2023;15(11):2216.
16. Meier-Stephenson V, Badmalia MD, Mrozowich T, et al. Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA. *J Biol Chem*. 2021;296:100589.
17. Giraud G, Rodà M, Huchon P, et al. G-quadruplexes control hepatitis B virus replication by promoting cccDNA transcription and phase separation in hepatocytes. *Nucleic Acids Res*. 2023;52(5):2290-2305.
18. Biswas B, Kandpal M, Vivekanandan P. A G-quadruplex motif in an envelope gene promoter regulates transcription and virion secretion in HBV genotype B. *Nucleic Acids Res*. 2017;45(19):11268-11280.
19. Wang J, Huang H, Zhao K, et al. G-quadruplex in hepatitis B virus pregenomic RNA promotes its translation. *J Biol Chem*. 2023;299(9):105151.
20. Abiri A, Lavigne M, Rezaei M, et al. Unlocking G-Quadruplexes as antiviral targets. *Pharmacol Rev*. 2021;73(3):897-923.
21. Böldicke T. Single domain antibodies for the knockdown of cytosolic and nuclear proteins. *Prot Sci*. 2017;26(5):925-945.
22. Galli S, Melidis L, Flynn SM, et al. DNA G-Quadruplex recognition in vitro and in live cells by a structure-specific nanobody. *J Am Chem Soc*. 2022;144(50):23096-23103.
23. Bazan J, Całkosiński I, Gamian A. Phage display—a powerful technique for immunotherapy. *Hum Vaccines Immunother*. 2012;8(12):1817-1828.
24. Jovčevska I, Muyldermans S. The therapeutic potential of nanobodies. *BioDrugs*. 2020;34(1):11-26.
25. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*. 2013;5(3):182-186.
26. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. *Proteins: Struct, Funct, Bioinf*. 2018;86(7):697-706.
27. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 2000;7(1-2):203-214.
28. 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):669-677.

29. Bay DH, Busch A, Lisdat F, et al. Identification of G-quadruplex structures that possess transcriptional regulating functions in the Dele and Cdc6 CpG islands. *BMC Mol Biol*. 2017;18(1):17.
30. del Villar-Guerra R, Trent JO, Chaires JB. G-Quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew Chem Int Ed*. 2018;57(24):7171-7175.
31. Reuten R, Patel TR, McDougall M, et al. Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nat Commun*. 2016;7(1):13515.
32. Grant TD, Luft JR, Carter LG, et al. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr D*. 2015;71(1):45-56.
33. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. 2011;95(8):559-571.
34. Lucifora J, Pastor F, Charles É, et al. Evidence for long-term association of virion-delivered HBV core protein with cccDNA independently of viral protein production. *JHEP Reports*. 2021;3(5):100330.
35. Tropberger P, Mercier A, Robinson M, Zhong W, Ganem DE, Holdorf M. Mapping of histone modifications in episomal HBV cccDNA uncovers an unusual chromatin organization amenable to epigenetic manipulation. *Proc Natl Acad Sci USA*. 2015;112(42):E5715-E5724.
36. Turton KL, Meier-Stephenson V, Badmalia MD, Coffin CS, Patel TR. Host transcription factors in hepatitis B virus RNA synthesis. *Viruses*. 2020;12(2):160.
37. Quasdorff M, Protzer U. Control of hepatitis B virus at the level of transcription. *J Viral Hepatitis*. 2010;17(8):527-536.
38. Raney AK, McLachlan A. Characterization of the hepatitis B virus large surface antigen promoter Sp1 binding site. *Virology*. 1995;208(1):399-404.
39. Zhong Y, Wu C, Xu Z, et al. Hepatitis B virus core protein is not required for covalently closed circular DNA transcriptional regulation. *J Virol*. 2022;96(21):e0136222.
40. Mendes E, Aljnadi IM, Bahls B, Victor BL, Paulo A. Major achievements in the design of Quadruplex-Interactive small molecules. *Pharmaceuticals*. 2022;15(3):300.
41. Cadoni E, De Paepe L, Manicardi A, Madder A. Beyond small molecules: targeting G-quadruplex structures with oligonucleotides and their analogues. *Nucleic Acids Res*. 2021;49(12):6638-6659.
42. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*. 2005;33(9):2908-2916.
43. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol*. 2020;21(8):459-474.
44. Asaadi Y, Jouneghani FF, Janani S, Rahbarizadeh F. A comprehensive comparison between camelid nanobodies and single chain variable fragments. *Biomark Res*. 2021;9(1):87.
45. Zhang YY, Zhang BH, Theele D, Litwin S, Toll E, Summers J. Single-cell analysis of covalently closed circular DNA copy numbers in a hepadnavirus-infected liver. *Proc Natl Acad Sci USA*. 2003;100(21):12372-12377.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Figueroa GB, D'souza S, Pereira HS, et al. Development of a single-domain antibody to target a G-quadruplex located on the hepatitis B virus covalently closed circular DNA genome. *J Med Virol*. 2024;96:e29692. doi:10.1002/jmv.29692