

Article

# Custom G4 Microarrays Reveal Selective G-Quadruplex Recognition of Small Molecule BMVC: A Large-Scale Assessment of Ligand Binding Selectivity

Guanhui Wu <sup>1,†</sup>, Desiree Tillo <sup>2,†</sup>, Sreejana Ray <sup>2,†</sup>, Ta-Chau Chang <sup>3</sup>,  
John S. Schneekloth Jr. <sup>4</sup>, Charles Vinson <sup>2</sup> and Danzhou Yang <sup>1,5,6,\*</sup>

<sup>1</sup> Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, 575 W Stadium Ave, West Lafayette, IN 47907, USA; wu1109@purdue.edu

<sup>2</sup> Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA; desiree.tillo@nih.gov (D.T.); sreejana.ray@gmail.com (S.R.); vinsonc@mail.nih.gov (C.V.)

<sup>3</sup> Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei 106, Taiwan; tcchang@pub.iams.sinica.edu.tw

<sup>4</sup> Chemical Biology Laboratory, National Cancer Institute-Frederick, Frederick, MD 21702, USA; schneeklothjs@mail.nih.gov

<sup>5</sup> Purdue Center for Cancer Research, West Lafayette, IN 47906, USA

<sup>6</sup> Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA

\* Correspondence: yangdz@purdue.edu; Tel.: +1-765-494-8148

† These authors contributed equally.

Academic Editor: Aldo Galeone

Received: 29 June 2020; Accepted: 25 July 2020; Published: 30 July 2020



**Abstract:** G-quadruplexes (G4) are considered new drug targets for human diseases such as cancer. More than 10,000 G4s have been discovered in human chromatin, posing challenges for assessing the selectivity of a G4-interactive ligand. 3,6-bis(1-Methyl-4-vinylpyridinium) carbazole diiodide (BMVC) is the first fluorescent small molecule for G4 detection in vivo. Our previous structural study shows that BMVC binds to the MYC promoter G4 (MycG4) with high specificity. Here, we utilize high-throughput, large-scale custom DNA G4 microarrays to analyze the G4-binding selectivity of BMVC. BMVC preferentially binds to the parallel MycG4 and selectively recognizes flanking sequences of parallel G4s, especially the 3'-flanking thymine. Importantly, the microarray results are confirmed by orthogonal NMR and fluorescence binding analyses. Our study demonstrates the potential of custom G4 microarrays as a platform to broadly and unbiasedly assess the binding selectivity of G4-interactive ligands, and to help understand the properties that govern molecular recognition.

**Keywords:** G-quadruplex; G4; microarray; ligand selectivity; BMVC; MYC

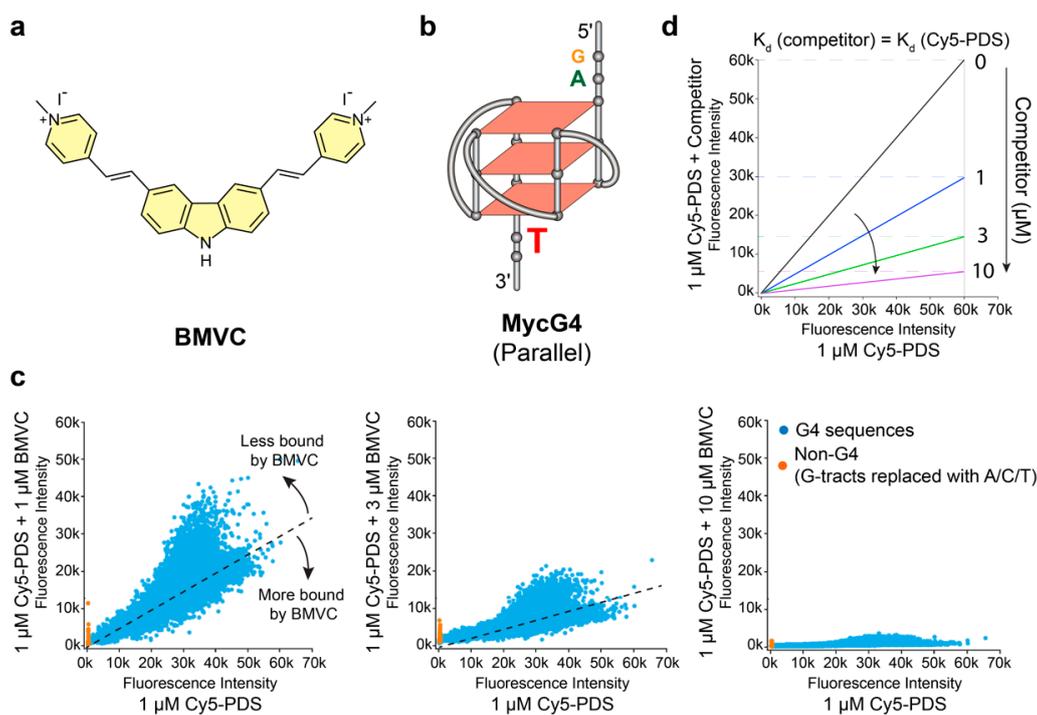
## 1. Introduction

G-quadruplexes (G4s) are four-stranded secondary structures formed in guanine-rich nucleic acids [1]. The building block of G4s is the G-tetrad, consisting of four guanines connected through Hoogsteen hydrogen bonds in a cyclic coplanar arrangement [2]. A G4 structure is formed when two or more G-tetrad planes stack on top of each other and is stabilized by physiological relevant monovalent cations, especially K<sup>+</sup> [3–5]. The biologically relevant intramolecular G4s are globular nucleic acid structures with unique folding and capping structures that provide an opportunity for selective targeting by small molecules [6–8].

G4 structures are involved in many cellular processes of DNA, including gene transcription [9,10], DNA replication [11], and genome stability [12,13]. In the human genome, G4 structures are prevalent

in the regulatory regions and enriched in the promoters of cancer-related genes [14,15]. In particular, MYC, one of the most deregulated oncogenes in human cancer, has a DNA-G4 forming motif (MycG4) in its promoter [9,16–20]. Compounds that bind and stabilize the MycG4 structure have been shown to repress MYC expression and lead to cancer cell death [8,9,16]. Therefore, the MycG4 is considered an attractive target for anticancer drugs. However, over 10,000 G4 structures have been discovered in human chromatin of precancerous cells [15,21]. It is thus important to determine the selectivity of a G4-targeting compound.

3,6-bis(1-Methyl-4-vinylpyridinium) carbazole diiodide (BMVC, Figure 1a) is a G4-interactive compound and the first fluorescent probe ( $\lambda_{ex,max} = 435$ ,  $\lambda_{em,max} = 580$ ) to detect G4 structures in human cells [22–24]. BMVC has also been developed as a potential fluorescent marker for cancer cells [25,26]. Whereas BMVC was first developed to detect G4 structures in human telomeres, our recent study shows that BMVC binds the MYC promoter G4 (MycG4, Figure 1b) with higher selectivity and affinity [27]. We have determined the solution structures of BMVC-MycG4 complexes, which show that BMVC binds to the MycG4 via multiple interactions, including stacking external G-tetrads, recognition of the MycG4-flanking bases, and conformational adjustment of the BMVC molecule. Moreover, our results show BMVC represses MYC expression in a human breast cancer cell line. However, the binding selectivity of BMVC to potential G4s formed in the human chromatin has not been broadly examined.



**Figure 1.** The binding of 3,6-bis(1-Methyl-4-vinylpyridinium) carbazole diiodide (BMVC) to various G4 structures differs from pyridostatin (PDS). (a) Chemical structure of BMVC. (b) MycG4, the major G-quadruplex formed in the MYC promoter NHE III<sub>1</sub> in K<sup>+</sup> solution, a parallel-stranded G4 structure. (c) Competition experiments of DNA microarrays with thousands of G4 sequences showing the differential binding of BMVC to various G4s as compared to Cy5-fluorophore ( $\lambda_{ex,max} = 647$ ,  $\lambda_{em,max} = 665$ ) labeled small molecule pyridostatin (Cy5-PDS), as shown by Cy5-PDS fluorescence intensity. The competition experiments were performed in the presence of 1, 3, and 10  $\mu$ M BMVC. The black dashed lines represent predicted linear relationships when the binding affinities of BMVC and PDS are the same. G4-containing sequences are shown in blue spots. Non-G4 forming sequences are shown in orange spots and serve as negative controls. Each spot represents the average of two independent measurements (Figure S1). (d) Schematic diagram showing the predicted linear relationships when the competitor has the same binding affinities as Cy5-PDS. The competition effects can be revealed by a dose-dependent slope reduction.

Microarray glass slides with hundred thousands of DNA sequences are a fast, straightforward, and high-throughput platform that has been employed to screen, profile, and quantify ligand and protein interactions with DNA and RNA molecules [28–30]. We have very recently designed custom DNA microarrays that can assess the binding selectivities of proteins, small molecules, and antibodies across over 15,000 potential G4 structures [31].

Herein, we report a binding-selectivity analysis of BMVC to the MycG4 and other G4 structures using custom G4 microarrays and competition experiments between Cy5-fluorophore ( $\lambda_{\text{ex,max}} = 647$ ,  $\lambda_{\text{em,max}} = 665$ ) labeled small molecule pyridostatin [32] (Cy5-PDS) and unlabeled BMVC. Our results show that BMVC differentially binds to various G4 structures and has a different G4 selectivity profile from Cy5-PDS. BMVC shows preferential binding to the MycG4 among the known G4 structures. Moreover, the microarray data reveals the sequence selectivity of BMVC to the flanking residues of the MycG4, especially at the 3'-end. The large-scale microarray results are confirmed by orthogonal small-scale NMR and fluorescence binding analyses. This is the first large-scale study of a G4-interactive ligand that shows a high-throughput evaluation of G4-binding selectivity and sequence specificity with unbiased selection of G4 sequences. It demonstrates the potential of custom DNA microarrays in the development of drugs targeting DNA or RNA structures.

## 2. Results

### 2.1. BMVC Binds G4 Sequences Differently from PDS

We have designed custom G4 microarrays that contain a total of 19,249 G4 DNA sequences [31]. The G4 microarrays were created by covalently attaching thousands of unique G4-forming DNA 60-mers to a glass surface. Pyridostatin (PDS) is a known G4-interactive compound. Measured by the fluorescence intensity of Cy5-PDS bound to each sequence in potassium-containing solution, Cy5-PDS was shown to preferentially bind G4-forming sequences on the G4 microarrays [31]. To test the binding selectivity of BMVC, we performed competition experiments using custom G4 microarrays (Figure 1 and Figure S1). The addition of potassium-containing solution to G4-forming oligonucleotides induced G4 formation. Subsequently, the microarrays were incubated with 1  $\mu\text{M}$  Cy5-PDS in the absence or presence of 1  $\mu\text{M}$ , 3  $\mu\text{M}$ , or 10  $\mu\text{M}$  of the unlabeled BMVC molecule. After washing to remove the unbound Cy5-PDS and BMVC, the fluorescence intensities of Cy5-PDS bound to DNA oligonucleotides were detected using a fluorescence scanner. The binding selectivity of BMVC to different G4 structures was assessed by measuring the relative fluorescence intensity reduction of Cy5-PDS as BMVC concentration increased.

The fluorescence intensities of 1  $\mu\text{M}$  Cy5-PDS in the presence of various concentrations of unlabeled BMVC were plotted against the fluorescence intensities in the absence of BMVC (Figure 1c). The competition experiment of 1  $\mu\text{M}$  Cy5-PDS with 1  $\mu\text{M}$  of unlabeled PDS was performed as the positive control (Figure S2). For a compound that competitively binds all sequences with the same affinity as Cy5-PDS, the competition experiments of 1  $\mu\text{M}$  Cy5-PDS with various concentrations of the unlabeled compound will follow the predicted linear relationships (Figure 1d). Furthermore, fluorescence intensities of Cy5-PDS bound to various G4 sequences will uniformly decrease in a dose-dependent manner, as presented by decreased slopes (Figure 1d). In the competition experiment, the unlabeled BMVC could compete with the Cy5-PDS binding to G4 sequences in a dose-dependent manner (Figure 1c). However, the binding profile of BMVC was different from unlabeled PDS. Selectivity can be better assessed at equimolar concentrations of unlabeled ligand and Cy5-PDS (both 1  $\mu\text{M}$ ) (Figure 1c, left and Figure S2). BMVC displays a more pronounced binding selectivity to different G4 sequences, as shown by a larger deviation from linear relationships, particularly with the stable-G4 forming sequences (at higher fluorescence intensities, Figure 1c, left). Unlabeled PDS appears to bind less selectively to the G4 sequences than BMVC, as shown by the stronger competition at the weaker Cy5-PDS-bound sequences (non-G4 sequences) (at lower fluorescence intensities, Figure S2).

## 2.2. BMVC Shows Different Binding Selectivity to Various G4 Structures as Compared to PDS

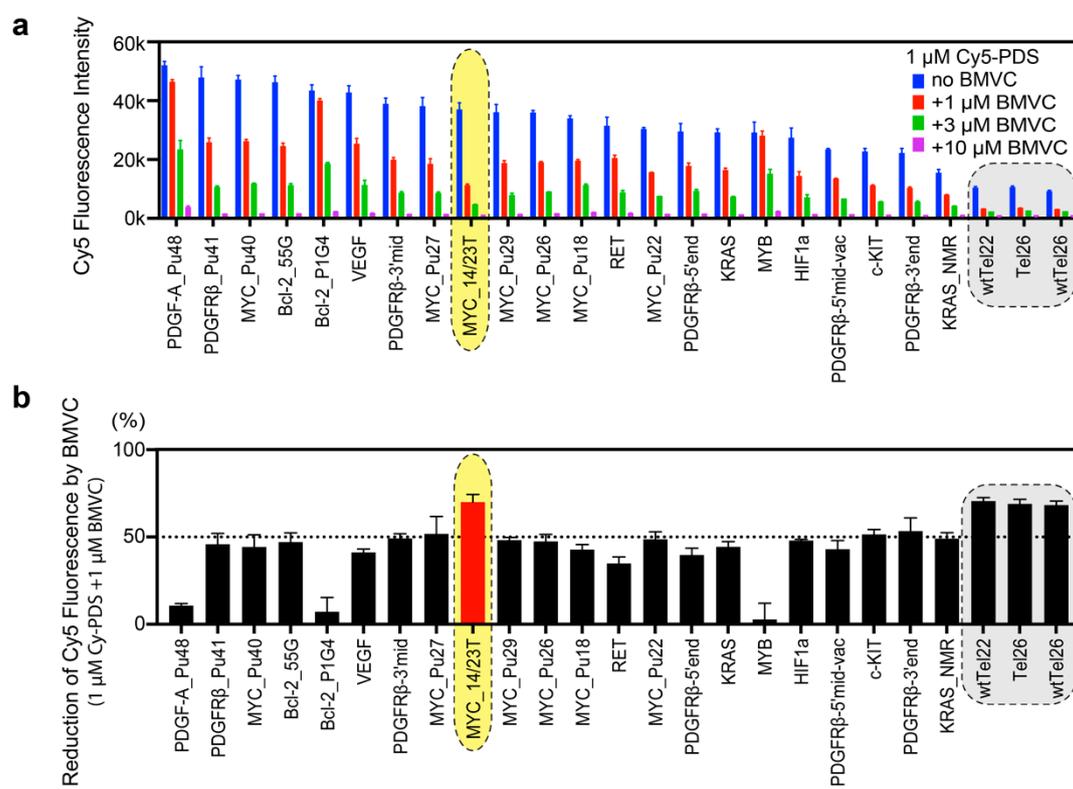
To determine the G4-binding selectivity of BMVC, we examined the BMVC binding to known G4 structures, including 7 well-studied MYC promoter G4 sequences, 15 other oncogene promoter G4 sequences, and 3 human telomeric G4 sequences (Table 1). BMVC competes with the binding of Cy5-PDS to most G4 sequences in a dose-dependent manner as indicated by reduced fluorescence intensities (Figure 2a).

**Table 1.** G4 sequences analyzed in Figure 2.

Name	G4 Sequence (5'→3')
MYC_Pu40 [9]	TTATGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGGAGGAG
MYC_Pu29 [9]	TTGGGGAGGGTGGGGAGGGTGGGGAAGGT
MYC_Pu27 [9]	TGGGGAGGGTGGGGAGGGTGGGGAAGG
MYC_Pu26 [33,34]	TTGGGGAGGGTGGGGAGGGTGGGGAA
MYC_Pu22 [35,36]	TGAGGGTGGGGAGGGTGGGGAA
MYC_14/23T [35,36]	TGAGGGTGGGTAGGGTGGGTAA
MYC_Pu18 [37]	AGGGTGGGGAGGGTGGGG
PDGFRβ_Pu41 [38]	GCTGGGAGAAGGGGGGGCGGCGGGCAGGGAGGGTGGACGC
PDGFRβ-5'end [38]	TTGGGAGAAGGGGGGGCGGCGGGGCA
PDGFRβ-5'mid-vac [39]	AAGGGAGGGCGGCGGGGCA
PDGFRβ-3'mid [40]	AAGGGGGGGCGGCGGGCAGGGAGGGT
PDGFRβ-3'end [41]	CGGCGGGCAGGGAGGGTGGACG
wtTel22 [42]	AGGGTTAGGGTTAGGGTTAGGG
Tel26 [43–45]	TTAGGGTTAGGGTTAGGGTTAGGGAAA
wtTel26 [45,46]	TTAGGGTTAGGGTTAGGGTTAGGGTTA
Bcl-2_55G [47]	AGGGGCGGGCGGCGGGAGGAAGGGGGCGGGA
Bcl-2_P1G4 [48]	CGGGCGGGAGCGGCGGGCGGGCGGGCGGGC
PDGF-A_Pu48 [49]	GGAGGCGGGGGGGGGGGGGCGGGGGCGGGGGCGGGGGAGGGGGCGGGC
KRAS [50]	AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGGCAG
KRAS_NMR [51]	AGGGCGGTGTGGGAATAGGGAA
VEGF [52]	CGGGGCGGGCCGGGGGGCGGGGT
RET [53]	GGGTAGGGGCGGGGCGGGGCGGGGGC
MYB [54]	GGAGGAGGAGGTCACGAGGAGGAGGAGAAGGAGGAGGAGGA
HIF1a [55]	GGGAGGGAGAGGGGGCGGG
c-KIT [56]	AGGGAGGGCGCTGGGAGGAGGG

Comparison of the inhibitory effects for the known G4 structures revealed differential G4 binding selectivity of BMVC vs. Cy5-PDS (Figure 2b). We first ranked G4 sequences based on the fluorescence intensity of bound Cy5-PDS. As illustrated in Figure 2a (blue bars), Cy5-PDS prefers long and highly G-rich sequences, such as PDGF-A\_Pu48, PDGFRb\_Pu41, and MYC\_Pu40. In addition, it also binds well to parallel G4s, such as Bcl-2\_55G, Bcl-2\_P1G4, VEGF, and various MYC G4s. For most G4s, the fluorescence intensity of 1 μM Cy5-PDS was reduced by 50% upon equimolar addition of BMVC (Figure 2b), suggesting a similar binding affinity of BMVC and Cy5-PDS to these G4s. However, the binding of BMVC was much weaker for the PDGF-A\_Pu48 and MYB sequences than the binding of Cy5-PDS (Figure 2b). Both PDGF-A\_Pu48 and MYB sequences have no-5'-flanking, while MYB forms a tetrad-heptad structure [54], whereas the optimal binding of BMVC requires a flanking base at both

the 5'-end and 3'-end, as shown by NMR solution structural study of the BMVC-MycG4 complex [27] (Figure S3).



**Figure 2.** The binding preference of BMVC to known G4 structures. Among the known G4 structures, BMVC prefers to bind to MYC\_14/23T, 5'-TGAGGGTGGGTAGGGTGGGTAA-3' (highlighted by yellow shade). Telomeric sequences are known to form nonparallel structures [42–46] and are poorly bound by Cy5-PDS (shaded in gray). (a) The competition microarray experiments showing dose-dependent inhibitory effects of BMVC on the binding of Cy5-PDS to various known G4 structures. The G4 sequences are shown in Table 1.  $n = 2$  to 20 independent measurements. Error bars represent mean  $\pm$  SD. (b) BMVC has different inhibitory effects on the binding of Cy5-PDS to the known G4 structures at the equal molar concentration (1  $\mu\text{M}$ ).  $n = 2$  to 20 independent measurements. Error bars represent mean  $\pm$  SD.

Cy5-PDS binds strikingly poor to nonparallel G4s, such as human telomeric G4s, which show less than 25% fluorescence intensity as compared to parallel-stranded G4s (Figure 2b). BMVC significantly inhibited the binding of Cy5-PDS to human telomeric G4s (Figure 2b), indicating a stronger binding of BMVC to the human telomeric G4s as compared to PDS. However, fluorescence measurements showed that BMVC binds parallel G4s, such as MYC and VEGF G4s, much stronger than the human telomeric G4s (Figure S4). Therefore, the microarray competition result indicates that Cy5-PDS binds the human telomeric G4s even weaker than BMVC.

### 2.3. BMVC Preferentially Binds to MYC\_14/23T among the Known G4 Structures

In general, Cy5-PDS and BMVC both strongly bind to parallel G4s (Figure 2b). Intriguingly, among all parallel G4s, BMVC induced largest reduction of the Cy5-PDS binding to the MYC\_14/23T G4. It is important to note that Cy5-PDS also binds the MYC\_14/23T G4 sequence very well (Figure 2a), therefore the strongest competition effect demonstrates that BMVC selectively recognizes the MYC\_14/23T G4. The MYC promoter G4 is the best-studied promoter G-quadruplex structure and a prototype of parallel G4s [8]. Notably, MYC\_14/23T and MYC\_Pu22 form the same parallel G4 (Figure 1b) except for the 3'-end flanking residue, which is a T in MYC\_14/23T and a G in MYC\_Pu22 [35]. The strikingly

stronger binding of BMVC to MYC\_14/23T than MYC\_Pu22 (Figure 2b) indicates that BMVC selectively recognizes the 3'-flanking T of MYC\_14/23T G4.

#### 2.4. BMVC Selectively Recognizes the Flanking Sequences of Parallel G4s, Especially the 3'-Flanking T

To examine the preference of BMVC for specific flanking sequences, we analyzed the binding of BMVC to MYC G4-derived sequence variants of the two flanking bases at both ends (5'-NNGGGTGGGGAGGGTGGGNN-3') using the competition microarray experiments. The differential reduction of Cy5-PDS binding to variants in the flanking sequences induced by BMVC addition reveals the binding selectivity for specific MYC G4 flanking sequences (Figure 3a). In the absence of BMVC, Cy5-PDS exhibits a slight preference for the 3'-flanking C and T, as shown by the most-bound (top 10%) and least-bound (bottom 10%) flanking variants (Figure 3b, top panel). The addition of BMVC significantly altered the most and least Cy5-PDS-bound flanking variants, with a clearly stronger selectivity at the 3'-end than at the 5'-end (Figure 3b, middle and bottom panels). The most and least Cy5-PDS-bound flanking variants in the presence of equimolar BMVC reveal the binding selectivity of BMVC. Particularly, thymine became markedly less enriched in the top 10% Cy5-PDS most-bound 3'-flanking variants but significantly enriched in the bottom 10% Cy5-PDS least-bound variants, indicating that BMVC strongly prefers the MYC G4 with the 3'-flanking T. On the other hand, C is the least-favored flanking base for BMVC binding at both the 3'- and 5'- ends, as shown by the greater enrichment in the top 10% Cy5-PDS most-bound flanking variants.

We further analyzed the effects of BMVC on the Cy5-PDS-binding to MYC G4 loop and single flanking-base sequence variants (5'-NGGGNNGGGNNGGGNNGGGN-3'), which include all possible loop and flanking variants (Figure 3c). Consistent with the two-base-flanking variants, the results showed BMVC strongly preferred the 3'-end flanking T but disfavored the flanking C at both ends. In contrast, Cy5-PDS preferred C for all three loops and the 3'-end flanking. It is noted that the MYC G4 single flanking-base variants all contain additional 3'-flanking bases for linking the G4 oligos to the microarray plates.

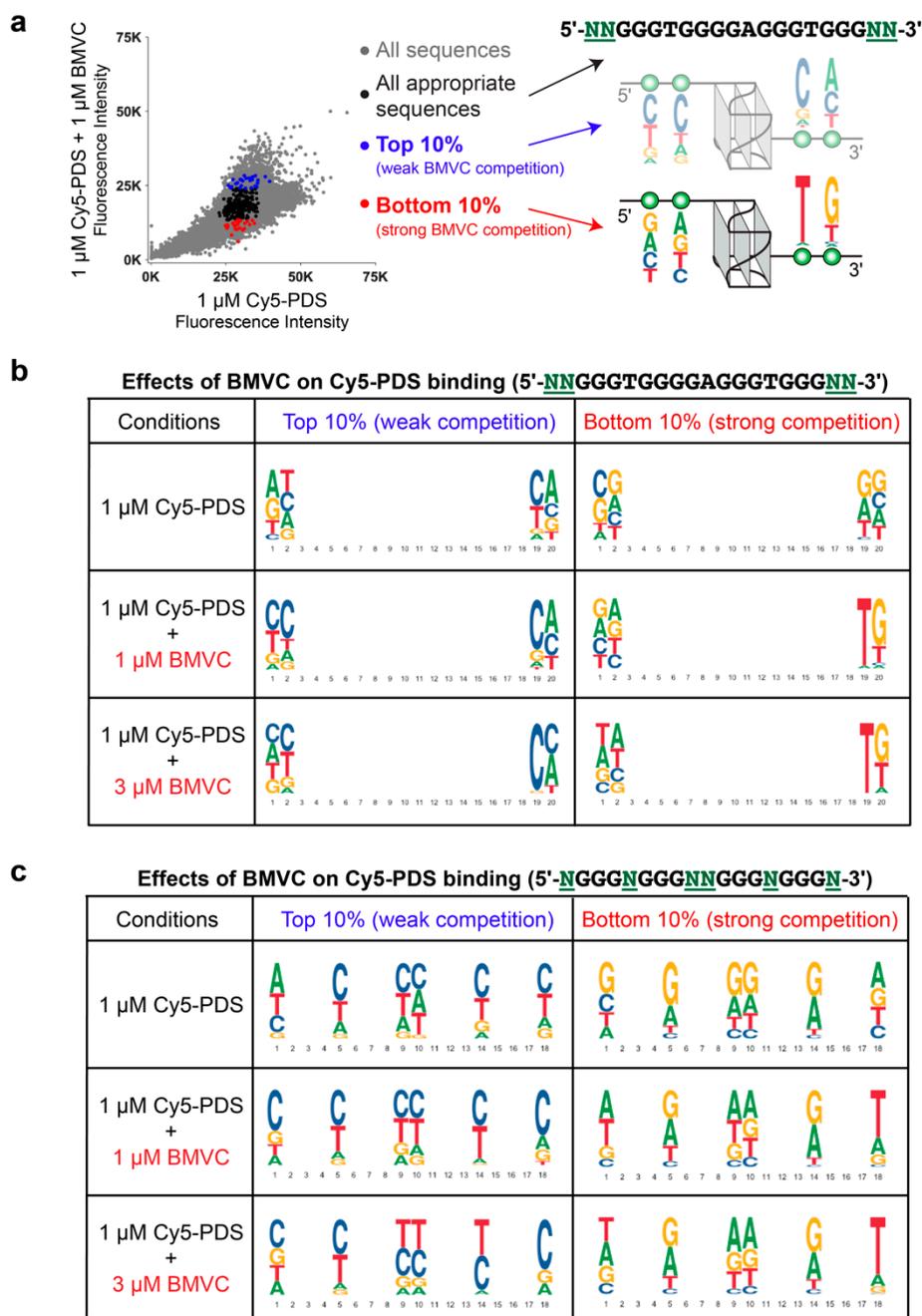
The sequence selectivity shown by the flanking variants explains the markedly weaker binding of BMVC to Bcl-2\_P1G4 (Figure 2). Bcl2\_55G and Bcl-2\_P1G4 both form parallel G4s with a long central loop (13-nt long in Bcl2\_55G and 12-nt long in Bcl-2\_P1G4) but different flanking sequences [47,48]. However, whereas BMVC showed good binding to Bcl2\_55G similar to other parallel G4s, the binding to Bcl-2\_P1G4 was markedly weaker (Figure 2a,b). Bcl-2\_P1G4 has a flanking C at both the 5'- and 3'- ends and only contains a short 1-nt flanking at the 5'-end, suggesting that BMVC disfavors the flanking C and short flanking.

#### 2.5. NMR Binding Experiments Confirm the Binding Selectivity of BMVC to G4 Structures and Flanking Sequences

The binding selectivity of BMVC to G4 structures and flanking sequences was confirmed by NMR titration experiments of BMVC to different G4 sequences, including parallel-stranded MYC\_14/23T G4 and its 5'- and 3'-flanking variants, VEGF and MYC1234 G4s, basket-type human telomeric G4 (wtTel22 in Na<sup>+</sup>), and hybrid type human telomeric G4 (Tel26 in K<sup>+</sup>) (Figures S5 and S6). BMVC binds best to the MYC\_14/23T G4, as indicated by well-resolved imino proton peaks for BMVC complexes (Figure S5a). Our previous NMR solution structural study shows that BMVC binds at both ends of the MYC\_14/23T G4 to form a 2:1 complex [27] (Figure S3). Mutations at the 5'-flanking sequence do not affect the binding of BMVC at the 5'-end (Figure S6c–e). In contrast, the 3'-end binding of BMVC is sensitive to the mutations at the 3'-flanking sequence, with a clear preference for the 3'-flanking T (Figure S6b). In addition, BMVC prefers at least two flanking bases for a specific binding (Figure S6). These results are in good agreement with the DNA microarray data (Figure 3).

While BMVC binds the MYC\_14/23T G4 with the highest affinity (Figure S4), BMVC can bind well to other parallel G4s, such as MYC1234 and VEGF G4 (Figure S5b,d). Additionally, BMVC favors the 5'-flanking A of parallel G4s, as indicated in the NMR titration data of the VEGF G4 flanking variants

(Figure S5c,d). However, BMVC did not show specific binding to the basket-type or hybrid-type human telomeric G4s (Figure S5e,f). These results are consistent with the G4 microarray data.



**Figure 3.** The binding selectivity of BMVC for the flanking sequences of MYC G4. (a) An illustration of binding logo analysis for Cy5-PDS signal with the competition of 1 μM BMVC on MYC G4-derived sequences 5'-NNGGGTGGGGAGGGTGGGNN-3' (256 total sequences). (b) The inhibitory effects of BMVC on the binding of Cy5-PDS to MYC G4-derived sequences with variant 5'- and 3'- flanking segments (5'-NNGGGTGGGGAGGGTGGGNN-3'). (c) The inhibitory effects of BMVC on the binding of Cy5-PDS to MYC G4-derived sequences 5'-NCGGNNGGGNNGGGNCGGN-3', (4096 total sequences), which include all possible loop and flanking variants.

### 3. Conclusions

We have established a high-throughput, large-scale custom G4 DNA microarray to assess the binding selectivities of proteins and small molecules across ~20,000 potential G4 structures simultaneously. Competition binding experiments of the Cy5 labeled PDS and the unlabeled G4-interactive small molecule BMVC demonstrate that the custom G4 microarray platform can assess the binding selectivity of BMVC to various G4 structures and flanking sequences, as well as differential G4 binding selectivity between BMVC and PDS. Our results reveal that BMVC selectively binds parallel G4s, in particular the MYC\_14/23T G4. Moreover, the G4 microarray data shows BMVC selectively recognizes the flanking sequences of parallel G4s, especially the 3'-flanking T. Importantly, the binding and sequence selectivity revealed by the large-scale DNA microarray data is in good agreement with the individual binding data by NMR and fluorescence. Our study demonstrates that the G4 DNA microarray provides a high-throughput and unbiased platform to assess the binding selectivity of G4-targeting molecules on a large scale and can help understand the properties that govern molecular recognition.

### 4. Materials and Methods

#### 4.1. Custom G4 DNA Microarray Design

We designed a custom microarray that contains four identical sectors that contain ~177,440 ssDNA 60-mers to examine G4 binding selectivity (NCBI GEO Platform GPL28372). The microarray contains different sets of G4 variants designed to examine several sequence parameters that affect G4 formation and binding selectivity such as loop length, loop sequence, flanking tail sequence, and single nucleotide variants of known G4s [31]. Briefly, the array includes a set of sequences from human telomeres and oncogene promoters known to form G4s with various topologies as positive controls (Table 1) as well as a set of 295 additional G4-forming sequences from the literature [57]. Loop and flanking tail sequences were varied using A, T, G, and C polynucleotide stretches and a subset of combinations, described in [31]. For the flanking variants, we generated 256 versions of the major MYC G4 with all possible dinucleotide flanking sequences (5'-NNGGGTGGGGAGGGTGGGNN-3'). For the loop sequence variants, we generated 4,096 sequences of the form 5'-NNGGNGGGNNGGGNNGGGN-3'. Negative controls include 19 oncogene G4s in which all G-tracts are replaced with either A, T, or C, reverse complements of G4 sequences, as well as a set of 86 published non-G4 sequences [57].

#### 4.2. DNA Microarray Binding Experiments

DNA microarray experiments were performed and analyzed as described previously [31]. Microarrays were preincubated with a pH 7.4 phosphate buffer solution with 100 mM potassium for 1 h at room temperature to induce G4 formation. Arrays then were blocked with 4% nonfat dry-milk in a potassium phosphate buffer before incubation with small molecules (Cy5-PDS, Cy5-PDS+BMVC, or Cy5-PDS+PDS) for 1 h at room temperature.

#### 4.3. Data Processing and Analysis

Molecule-bound microarrays were scanned with an Agilent G5761A SureScan Dx Microarray Scanner System to detect Cy5 signal at two laser settings (30 and 100 PMT). Spot intensities from microarray images were extracted using Agilent Feature Extraction Software and are reported as raw fluorescence intensities. All binding assays were performed twice with high agreement between replicates ( $R > 0.8$ ). Microarrays with the fewest number of saturated spots were used for further analysis. Median intensity was then computed for probes containing identical sequence on each microarray. Sequence logos were generated from a position frequency matrix generated from selected sequences using ggseqlogo [58].

#### 4.4. NMR Spectroscopy Experiments

G4 DNA oligonucleotides were synthesized using  $\beta$ -cyanoethylphosphoramidite solid-phase chemistry (Applied Biosystem Expedite 8909), as described previously [36]. NMR experiments were performed on a Bruker AV-III-500-HD equipped with a BBFO Z-gradient cryoprobe. DNA samples were heated to 95 °C for 5 min, then cooled slowly for G4 formation. For the 1D <sup>1</sup>H NMR experiments, samples contained 100–250  $\mu$ M DNA in an appropriate buffer solution with 10% D<sub>2</sub>O for the lock. The titrations were performed by adding increasing amounts of the compounds to the DNA samples in solution.

**Supplementary Materials:** The following are available online. Figure S1: Comparison of replicate fluorescence intensity of Cy5-PDS in the absence and presence of 1, 3, and 10  $\mu$ M of BMVC; Figure S2: Competition microarray experiments of Cy5-PDS with unlabeled PDS; Figure S3: NMR solution structure of the 2:1 complex of BMVC and MYC\_14/23T G4 (PDB ID: 6O2L); Figure S4: Apparent dissociation constant ( $K_d$ , app) of BMVC binding to various G4s; Figure S5: 1D <sup>1</sup>H NMR titration spectra of BMVC with various G4 sequences; Figure S6: 1D <sup>1</sup>H NMR titration spectra of BMVC with MYC\_14/23T flanking variants.

**Author Contributions:** Conceptualization, G.W., C.V., and D.Y.; Data curation, G.W. and S.R.; Formal analysis, G.W., D.T., and D.Y.; Funding acquisition, C.V. and D.Y.; Resources, T.-C.C., J.S.S.J., C.V., and D.Y.; Supervision, C.V. and D.Y.; Writing—original draft, G.W. and D.Y.; Writing—review and editing, G.W., D.T., T.-C.C., J.S.S.J., C.V., and D.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Institutes of Health R01CA177585 (DY) and P30CA023168 (Purdue Center for Cancer Research).

**Acknowledgments:** We are grateful to Jonathan Dickerhoff for his comments and proofreading the manuscript. We thank Kaibo Wang for his help in  $K_d$  fluorescence measurements.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Notes:** Data (raw probe intensities and processed median intensities) are available at the NCBI GEO database under accession GSE148419.

#### References

1. Yang, D. G-Quadruplex DNA and RNA. *Methods Mol. Biol.* **2019**, *2035*, 1–24. [[CrossRef](#)] [[PubMed](#)]
2. Gellert, M.; Lipsett, M.N.; Davies, D.R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018. [[CrossRef](#)] [[PubMed](#)]
3. Williamson, J.R.; Raghuraman, M.K.; Cech, T.R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **1989**, *59*, 871–880. [[CrossRef](#)]
4. Sen, D.; Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **1990**, *344*, 410–414. [[CrossRef](#)] [[PubMed](#)]
5. Hud, N.V.; Smith, F.W.; Anet, F.A.L.; Feigon, J. The selectivity for  $K^+$  versus  $Na^+$  in DNA quadruplexes is dominated by relative free energies of hydration: A thermodynamic analysis by <sup>1</sup>H NMR. *Biochemistry* **1996**, *35*, 15383–15390. [[CrossRef](#)]
6. Neidle, S. Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **2016**, *59*, 5987–6011. [[CrossRef](#)]
7. Yang, D.; Okamoto, K. Structural insights into G-quadruplexes: Towards new anticancer drugs. *Future Med. Chem.* **2010**, *2*, 619–646. [[CrossRef](#)]
8. Chen, Y.; Yang, D. Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Protoc. Nucleic Acid Chem.* **2012**, *50*, 17.5.1–17.5.17. [[CrossRef](#)]
9. Siddiqui-Jain, A.; Grand, C.L.; Bearss, D.J.; Hurley, L.H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598. [[CrossRef](#)]
10. Gray, L.T.; Vallur, A.C.; Eddy, J.; Maizels, N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **2014**, *10*, 313–318. [[CrossRef](#)]
11. Bochman, M.L.; Paeschke, K.; Zakian, V.A. DNA secondary structures: Stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **2012**, *13*, 770–780. [[CrossRef](#)] [[PubMed](#)]

12. Piazza, A.; Boule, J.B.; Lopes, J.; Mingo, K.; Largy, E.; Teulade-Fichou, M.P.; Nicolas, A. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2010**, *38*, 4337–4348. [[CrossRef](#)]
13. Ribeyre, C.; Lopes, J.; Boule, J.B.; Piazza, A.; Guedin, A.; Zakian, V.A.; Mergny, J.L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet* **2009**, *5*, e1000475. [[CrossRef](#)]
14. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [[CrossRef](#)]
15. Hansel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48*, 1267–1272. [[CrossRef](#)]
16. Brooks, T.A.; Hurley, L.H. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **2009**, *9*, 849–861. [[CrossRef](#)] [[PubMed](#)]
17. Simonsson, T.; Pecinka, P.; Kubista, M. DNA tetraplex formation in the control region of *c-myc*. *Nucleic Acids Res.* **1998**, *26*, 1167–1172. [[CrossRef](#)] [[PubMed](#)]
18. DesJardins, E.; Hay, N. Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human *c-myc* promoters. *Mol. Cell. Biol.* **1993**, *13*, 5710–5724. [[CrossRef](#)] [[PubMed](#)]
19. Michelotti, E.F.; Tomonaga, T.; Krutzsch, H.; Levens, D. Cellular nucleic acid binding protein regulates the CT element of the human *c-myc* protooncogene. *J. Biol. Chem.* **1995**, *270*, 9494–9499. [[CrossRef](#)] [[PubMed](#)]
20. Wu, G.; Xing, Z.; Tran, E.J.; Yang, D. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20453–20461. [[CrossRef](#)] [[PubMed](#)]
21. Kouzine, F.; Wojtowicz, D.; Baranello, L.; Yamane, A.; Nelson, S.; Resch, W.; Kieffer-Kwon, K.R.; Benham, C.J.; Casellas, R.; Przytycka, T.M.; et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **2017**, *4*, 344–356. [[CrossRef](#)] [[PubMed](#)]
22. Chang, C.C.; Wu, J.Y.; Chien, C.W.; Wu, W.S.; Liu, H.; Kang, C.C.; Yu, L.J.; Chang, T.C. A fluorescent carbazole derivative: High sensitivity for quadruplex DNA. *Anal. Chem.* **2003**, *75*, 6177–6183. [[CrossRef](#)] [[PubMed](#)]
23. Chang, C.C.; Kuo, I.C.; Ling, I.F.; Chen, C.T.; Chen, H.C.; Lou, P.J.; Lin, J.J.; Chang, T.C. Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. *Anal. Chem.* **2004**, *76*, 4490–4494. [[CrossRef](#)] [[PubMed](#)]
24. Chang, C.C.; Chu, J.F.; Kao, F.J.; Chiu, Y.C.; Lou, P.J.; Chen, H.C.; Chang, T.C. Verification of antiparallel G-quadruplex structure in human telomeres by using two-photon excitation fluorescence lifetime imaging microscopy of the 3,6-bis(1-methyl-4-vinylpyridinium)carbazole diiodide molecule. *Anal. Chem.* **2006**, *78*, 2810–2815. [[CrossRef](#)] [[PubMed](#)]
25. Kang, C.C.; Chang, C.C.; Cheng, J.Y.; Chang, T.C. Simple method in diagnosing cancer cells by a novel fluorescence probe BMVC. *J. Chin. Chem. Soc.* **2005**, *52*, 1069–1072. [[CrossRef](#)]
26. Chang, C.C.; Kuo, I.C.; Lin, J.J.; Lu, Y.C.; Chen, C.T.; Back, H.T.; Lou, P.J.; Chang, T.C. A novel carbazole derivative, BMVC: A potential antitumor agent and fluorescence marker of cancer cells. *Chem. Biodivers.* **2004**, *1*, 1377–1384. [[CrossRef](#)]
27. Liu, W.; Lin, C.; Wu, G.; Dai, J.; Chang, T.C.; Yang, D. Structures of 1:1 and 2:1 complexes of BMVC and MYC promoter G-quadruplex reveal a mechanism of ligand conformation adjustment for G4-recognition. *Nucleic Acids Res.* **2019**, *47*, 11931–11942. [[CrossRef](#)]
28. Berger, M.F.; Bulyk, M.L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* **2009**, *4*, 393–411. [[CrossRef](#)]
29. Badis, G.; Berger, M.F.; Philippakis, A.A.; Talukder, S.; Gehrke, A.R.; Jaeger, S.A.; Chan, E.T.; Metzler, G.; Vedenko, A.; Chen, X.; et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **2009**, *324*, 1720–1723. [[CrossRef](#)]
30. Iida, K.; Nakamura, T.; Yoshida, W.; Tera, M.; Nakabayashi, K.; Hata, K.; Ikebukuro, K.; Nagasawa, K. Fluorescent-ligand-mediated screening of G-quadruplex structures using a DNA microarray. *Angew. Chem. Int. Ed.* **2013**, *52*, 12052–12055. [[CrossRef](#)] [[PubMed](#)]
31. Ray, S.; Tillo, D.; Boer, R.E.; Assad, N.; Barshai, M.; Wu, G.; Orenstein, Y.; Yang, D.; Schneekloth, J.S., Jr.; Vinson, C. Custom DNA microarrays reveal diverse binding preferences of proteins and small molecules to thousands of G-quadruplexes. *ACS Chem. Biol.* **2020**, *15*, 925–935. [[CrossRef](#)] [[PubMed](#)]

32. Muller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. *Nat. Chem.* **2010**, *2*, 1095–1098. [[CrossRef](#)] [[PubMed](#)]
33. Phan, A.T.; Modi, Y.S.; Patel, D.J. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716. [[CrossRef](#)] [[PubMed](#)]
34. Dickerhoff, J.; Onel, B.; Chen, L.; Chen, Y.; Yang, D. Solution structure of a MYC promoter G-quadruplex with 1:6:1 loop length. *ACS Omega* **2019**, *4*, 2533–2539. [[CrossRef](#)]
35. Ambrus, A.; Chen, D.; Dai, J.; Jones, R.A.; Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **2005**, *44*, 2048–2058. [[CrossRef](#)]
36. Dai, J.; Carver, M.; Hurley, L.H.; Yang, D. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: Insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680. [[CrossRef](#)]
37. Seenisamy, J.; Rezler, E.M.; Powell, T.J.; Tye, D.; Gokhale, V.; Joshi, C.S.; Siddiqui-Jain, A.; Hurley, L.H. The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709. [[CrossRef](#)]
38. Qin, Y.; Fortin, J.S.; Tye, D.; Gleason-Guzman, M.; Brooks, T.A.; Hurley, L.H. Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. *Biochemistry* **2010**, *49*, 4208–4219. [[CrossRef](#)]
39. Wang, K.B.; Dickerhoff, J.; Wu, G.; Yang, D. PDGFR-beta Promoter Forms a Vacancy G-Quadruplex that Can Be Filled in by dGMP: Solution Structure and Molecular Recognition of Guanine Metabolites and Drugs. *J. Am. Chem. Soc.* **2020**, *142*, 5204–5211. [[CrossRef](#)]
40. Chen, Y.; Agrawal, P.; Brown, R.V.; Hatzakis, E.; Hurley, L.H.; Yang, D. The major G-quadruplex formed in the human platelet-derived growth factor receptor beta promoter adopts a novel broken-strand structure in K<sup>+</sup> solution. *J. Am. Chem. Soc.* **2012**, *134*, 13220–13223. [[CrossRef](#)]
41. Onel, B.; Carver, M.; Agrawal, P.; Hurley, L.H.; Yang, D. The 3'-end region of the human PDGFR-beta core promoter nuclease hypersensitive element forms a mixture of two unique end-insertion G-quadruplexes. *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862*, 846–854. [[CrossRef](#)] [[PubMed](#)]
42. Wang, Y.; Patel, D.J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1993**, *1*, 263–282. [[CrossRef](#)]
43. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R.A.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006**, *34*, 2723–2735. [[CrossRef](#)] [[PubMed](#)]
44. Luu, K.N.; Phan, A.T.; Kuryavyi, V.; Lacroix, L.; Patel, D.J. Structure of the human telomere in K<sup>+</sup> solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. [[CrossRef](#)] [[PubMed](#)]
45. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R.A.; Yang, D. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.* **2007**, *35*, 2440–2450. [[CrossRef](#)]
46. Phan, A.T.; Luu, K.N.; Patel, D.J. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K<sup>+</sup> solution. *Nucleic Acids Res.* **2006**, *34*, 5715–5719. [[CrossRef](#)]
47. Agrawal, P.; Lin, C.; Mathad, R.I.; Carver, M.; Yang, D. The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K<sup>+</sup> solution. *J. Am. Chem. Soc.* **2014**, *136*, 1750–1753. [[CrossRef](#)]
48. Onel, B.; Carver, M.; Wu, G.; Timonina, D.; Kalarn, S.; Larriva, M.; Yang, D. A New G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *J. Am. Chem. Soc.* **2016**, *138*, 2563–2570. [[CrossRef](#)]
49. Qin, Y.; Rezler, E.M.; Gokhale, V.; Sun, D.; Hurley, L.H. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* **2007**, *35*, 7698–7713. [[CrossRef](#)]
50. Morgan, R.K.; Batra, H.; Gaerig, V.C.; Hockings, J.; Brooks, T.A. Identification and characterization of a new G-quadruplex forming region within the KRAS promoter as a transcriptional regulator. *Biochim. Biophys. Acta* **2016**, *1859*, 235–245. [[CrossRef](#)]

51. Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L.A.; Mergny, J.L.; Salgado, G.F. High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem.* **2017**, *292*, 8082–8091. [[CrossRef](#)]
52. Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. Solution structure of the major G-quadruplex formed in the human VEGF promoter in K<sup>+</sup>: Insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.* **2013**, *41*, 10584–10592. [[CrossRef](#)] [[PubMed](#)]
53. Tong, X.; Lan, W.; Zhang, X.; Wu, H.; Liu, M.; Cao, C. Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Res.* **2011**, *39*, 6753–6763. [[CrossRef](#)] [[PubMed](#)]
54. Palumbo, S.L.; Memmott, R.M.; Uribe, D.J.; Krotova-Khan, Y.; Hurley, L.H.; Ebbinghaus, S.W. A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res.* **2008**, *36*, 1755–1769. [[CrossRef](#)] [[PubMed](#)]
55. De Armond, R.; Wood, S.; Sun, D.; Hurley, L.H.; Ebbinghaus, S.W. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* **2005**, *44*, 16341–16350. [[CrossRef](#)]
56. Wei, D.; Husby, J.; Neidle, S. Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic Acids Res.* **2015**, *43*, 629–644. [[CrossRef](#)]
57. Bedrat, A.; Lacroix, L.; Mergny, J.L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [[CrossRef](#)]
58. Wagih, O. ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* **2017**, *33*, 3645–3647. [[CrossRef](#)]

**Sample Availability:** Samples of BMVC and Cy5-PDS are available from the authors upon request.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).