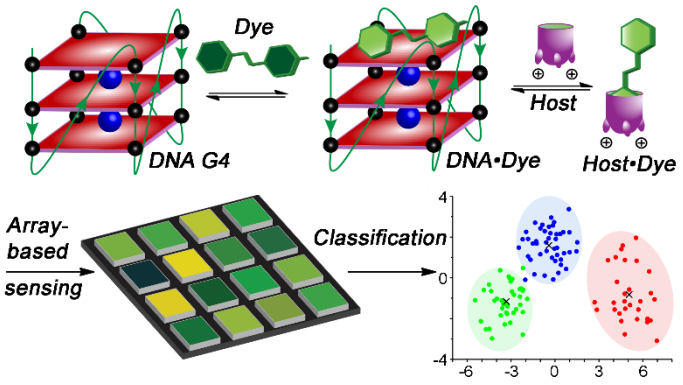
Graphical Abstract



An arrayed suite of synthetic hosts and dyes is capable of fluorescence detection of DNA G-quadruplex secondary structures. Multivariate analysis of different fluorescence enhancements — generated using cationic dyes that show affinity for both DNA and the hosts — enables discrimination between G4 structures of identical length and highly similar topological types.

Selective Discrimination and Classification of G-Quadruplex Structures with a Host–Guest Sensing Array

Junyi Chen,[b] Briana L. Hickey,[a] Linlin Wang,[a] Jiwon Lee,[a] Adam D. Gill,[c] Alessia Favero,[d] Roberta Pinalli,[d] Enrico Dalcanale,[d] Richard J. Hooley[a,c] and Wenwan Zhong[a,b]**

[a]Department of Chemistry, [b]Environmental Toxicology Program, [c]Department of Biochemistry and Molecular Biology, University of California, Riverside, CA. 92521, USA. [d] Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, 43124 Parma, and INSTM, UdR Parma, Italy.

ABSTRACT. *The secondary structures of nucleic acids have an important influence on their cellular functions, but can be difficult to identify and classify quickly. Here, we show that an arrayed suite of synthetic hosts and dyes is capable of fluorescence detection of oligonucleotide secondary structures. Multivariate analysis of different fluorescence enhancements — generated using cationic dyes that show affinity for both DNA G-quadruplexes and the synthetic hosts — enables discrimination between G4 structures of identical length and highly similar topological types. Different G4s that display the same folding topology can also be easily differentiated by the number of G-quartets and sequence differences at the 3' or 5' ends. The array is capable of both differentiation and classification of the G4 structures at the same time. This simple, non-invasive sensing method does not require the discovery and synthesis of specific G4 binding ligands, but employs a simple multicomponent approach to ensure wide applicability.*

INTRODUCTION

The cellular functions of nucleic acids are dependent not only on the nucleotide sequence, but also on the secondary structural architecture. Nucleotide strands can form complex 3D structures, expanding their function beyond encoding and transferring genetic information.[1] For example, uniquely folded DNA structures such as G-quadruplexes (G4s) have been found in cancer tissues and viral genomes,[2] and synthetic DNA and RNA nanostructures have a wide variety of applications in synthetic biology and nanomedicine.[3,4] As such, there is a need for methods that can quickly and easily assess the structure of nucleotides of various types, not just the sequence. Although DNA folding directed by Watson–Crick base-pairing can be predicted,[5] it remains challenging to analyze non-canonical structures from sequences alone. Determination of nucleic acid 3D structures can be achieved in different ways: complete structural analysis requires X-ray crystallography and/or multidimensional NMR spectroscopy,[6,7] whereas simple grouping into secondary structural types is possible with Circular Dichroism (CD) spectroscopy.[8] These techniques have their challenges, in that they are either too time-consuming and detailed for rapid analysis (X-Ray, NMR), or provide too little information (CD). A rapid, simple method that can selectively identify, differentiate and classify nucleotide secondary structures is highly desirable, but also very challenging, and applying this sensor in complex biological media is even more so.

Optical sensing of RNA secondary structures can be achieved with synthetic, derivatized nucleotides.[9] This requires differential sensing, which was popularized by Anslyn for small molecule analytes,[10] and has been used by others to classify the 2D and 3D structure of derivatized RNAs.[11] By incorporating dyes at multiple different positions on the RNA backbone and adding multiple small molecule effectors, array-based sensing can be exploited to classify different structural motifs and conformational changes, using multivariate analysis tools such as Principal

Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).[12] Some beautiful examples of RNA classification have been shown, but they are not an effective method for monitoring native, unmodified oligonucleotide structures. That requires an exogenous fluorophore that must show selective responses to nucleotides, and selectivity for different structural motifs, which is not simple. Some excellent work has been performed that uses fluorescence displacement assays and multivariate analysis to classify ligands that have selectivity for DNA strands with large structural differences (duplex, i-motif, G4, etc.).[13]

The greatest challenge for any sensor is to distinguish between highly similar structures in a complex mixture. DNA G4s are an excellent example of a challenging recognition target, as their overall size and the internal structure are relatively conserved. They are a common non-canonical structural motif,[14,15] involving stacked guanine quartets assembled *via* Hoogsteen base-pair contacts and intercalated Na^+ or K^+ ions linked by loop nucleotides.[16] G4s are generally classified into three different structural types, based on the directionality of the four G strands forming the G-quadruplex: parallel, antiparallel or hybrid (illustrated in Fig. 1b). As in dsDNA, G4 structures have grooves, the dimensions of which are determined by the overall topology and the nature of the loops. The variations lie in the strand orientation, glycosidic angles and the directionality of the external loops.[17] These changes are not easy to detect: they are, after all, just oligomers of phosphorylated sugars oriented in different directions. CD spectroscopy can disclose the types of topology, (i.e. parallel, antiparallel, or hybrid), but cannot recognize small structural differences in G4s of the same folding type. Although computational prediction and genomic mapping have revealed myriad sequence motifs that could form G4 structures,[18,19] conformational studies have only been conducted on a few of these sequences. This has involved the use of small molecule ligands, which recognize G4s mainly through π-π stacking with the G tetrads, plus binding to the

grooves or loops.[20,21] Some ligands are highly selective,[21] but they are not common, and not easily applied in sensing applications.

Here we show that small-molecule host–guest sensor arrays can be used to sense structural differences in unmodified DNA G-quadruplex structures, allowing both their differentiation and classification *via* simple fluorescence measurements and pattern recognition. Supramolecular probes are perfectly suited to array-based pattern recognition due to their simple synthesis, ease of use, and rapid access to multiple variables,[22] which is essential in detecting small changes in structure. Host–guest sensor arrays have been used to monitor enzyme reactions[23,24] and analyze peptide and protein modifications.[25,26] They have rarely been applied to nucleotide sensing, because synthetic hosts often show poor affinity for oligonucleotides.[27] The solution to this challenge lies in indirect sensing: by combining an arrayed suite of host molecules with multiple dye candidates that can bind both the hosts and the target oligonucleotides, small changes in target structure can be selectively detected.

RESULTS AND DISCUSSION

Design of host–guest sensor system and analysis of the sensing mechanism

Pattern recognition-based sensing requires multiple different components that show variable fluorescence responses to the target, in this case, oligonucleotide strands. The basic design principle is shown in Figure 1a, and exploits hosts and dye guests that have variable affinity for each other and with the target oligonucleotides. The sensor components consist of cationic dye molecules that can bind to DNA structures (Fig. 1d), as well as a series of synthetic host molecules that can also competitively bind these dyes while modulating their fluorescence (Fig. 1c). Extended Data Fig. 1 shows minimized structures of the host–guest complexes. The styrylpyridinium **DSMI** and **PSMI** fluorophores[28,29] are cationic, yet water-soluble, and similar in structure to known

oligonucleotide ligands.[30] They vary only in the size of headgroup and have similar fluorescence properties. The slightly varying size and shape should confer small differences in nucleotide affinity, while maintaining similar detection ranges.

The second component is a set of water-soluble host molecules (cavitands) that can bind to the fluorophores and modulate their emission. Cavitand hosts **1-5** (Fig. 1c) were used, all of which should show affinity for the dyes, allowing a competition between the G4 and the host. These sensor components can then be combined in an array for pattern recognition analysis. Tetraanionic host **1**[31] is known to bind **DSMI**[28] and **PSMI**[29], and has been extensively used in biosensing applications.[32,33] Cationic cavitand **2**[29] has a similar cavity size to **1**, but displays imidazolium groups at the lower rim for water solubility. Similarly, the benzimidazole cavitands **3** and **4**[34] are derivatives of **1** with cationic groups at the lower rim. Finally, the shallow phosphonate cavitand **5**[35] was used, as this also binds organic cations.[36]

The initial test was to determine whether the dyes could show variable response to G-quadruplexes in the presence of the different hosts. The two dyes were added to a suite of DNA structures, followed by increasing concentrations of hosts **1-5**, and the fluorescence response curves recorded. Eight different DNA structures were initially tested: four single-stranded DNAs (A20, G20, T20 and C20), and four G4 structures. *c-kit1* and *c-myc 2345* are parallel G4s, *bcl-2 2345* forms a hybrid structure, and *TBA* is antiparallel.[37] Gel electrophoresis confirmed that the G4 strands all formed unimolecular folded structures, and their folding topologies were all validated by CD (see Supplementary Figs. 1-6).

Both **PSMI** and **DSMI** showed fluorescence enhancements up to 3-6 fold upon addition to all the DNAs. The changes in the dyes by themselves with different G4 structures were not substantially different, so simple addition of the individual dyes to the DNAs does not provide

selectivity. However, upon addition of the cavitands to the dye–DNA mixtures, significant variations in emission properties can be immediately observed (see Fig. 2a-c). The responses are quite complex, and vary with host, in three broad types: the responses for anionic host **1**, shallow host **5**, and cationic hosts **2-4**. To account for the effects of the DNA on the emission profiles, two types of fluorescence plots are shown: the raw fluorescence counts (Extended Data Fig. 2a,c,e), and plots normalized to the response of cavitand–dye in the absence of DNA (Fig. 2a-c, and Extended Data Fig. 2 b, d, f). This normalization removes the effect of the host–dye emission on the signal, and focuses solely on the effects of changing DNA target. For full fluorescence plots for all host–dye combinations, see Supplementary Figs. 18, 19.

The normalized response for host **1** is quite simple (Fig. 2a), as the emission is lowered with increasing host concentration. The profiles for the cationic hosts are different, however (Fig. 2b, c, Extended Data Fig. 2c-f and Supplementary Figs. 18, 19). When hosts **2-5** are added to the DNA•dye complexes, an initial *increase* in emission is observed at very low concentration, followed by a fluorescence decrease. After 1 μM host is added, no effects on the normalized emission are seen with hosts **2-4**. In contrast, phosphonate host **5** causes a slight quenching of the dye upon **host•dye** complex formation (Extended Data Fig. 2e).[38] The fluorescence responses for the two dyes are slightly different, which provides a simple array variable: small changes in dye can cause small, reproducible changes in emission.

To rationalize these effects, we determined the binding affinities[39] between **PSMI/DSMI**, the hosts, and for the *c-myc-2345* G4 (see Table 1 and Supplementary Figs. 14-16). Both dyes bind strongly, yet variably to each host, with dissociation constants K_d from 2.8 μM (**DSMI•5**) to 72.3 μM (**DSMI•2**). The dyes also bound to the DNA G4, with K_d on the order of 100-250 μM for both

DSMI and **PSMI**. In each case, the G4–dye affinity was lower than the host–dye affinity, corroborating the theory that the hosts extract the dyes from the DNA.

These results suggest multiple sensing mechanisms, as shown in abbreviated form in Figure 2d. Cavitand **1** competitively binds the fluorophore, removing it from the DNA and causing an initial decrease in normalized fluorescence. Cationic hosts **2-5** are slightly different: the hosts competitively bind the dyes, and at high concentrations this eliminates any **dye•DNA** binding. At low concentrations, the arced fluorescence plots indicate that a second mechanism is present, presumably due to the cationic hosts also interacting with the DNA. Three states can be in equilibrium: **DNA•dye**, **host•dye** and **DNA•dye•host**, whereby the host causes an additional enhancement in fluorescence of the dye. This is consistent for all cationic hosts **2-5**, and only occurs at low [**host**]. It is not completely clear why the "arcing" in the fluorescence titrations occurs for the cationic hosts. Some plausible options are that either the **dye•host** complex associates with the DNA ("state 2", Fig. 2d), causing an increase in emission, or the added host interacts with the **dye•DNA** complex ("state 1"), or both.

Discriminating different DNA structures with the sensor array

While there are multiple mechanistic possibilities for the emission differences with variable hosts and guests, the important conclusion for differential sensing purposes is that the responses are small, but highly sensitive to DNA structure. As such, we constructed four different **host•dye** arrays of varying sizes. The **host•dye** concentrations were chosen to give the optimal response differences. The **Type 1** array (6 component) uses **DSMI** dye and hosts **1-5**, along with no host, and the **Type 2** array (6 component) uses **PSMI** dye with hosts **1-5**, along with no host. Two larger arrays were also used - the **10-component array** uses **DSMI** and **PSMI** dyes with hosts **1-5**, and the **12-component array** consists of both **Type 1** and **Type 2** arrays. For full details, see Methods.

Exposing the eight different DNA targets to the full 12-component array gives the response plot in Figure 3a, which illustrates the highly variable responses possible. These responses were subjected to PCA, and the scores plot is shown in Figure 3b. This shows that the array can fully discriminate all 8 strands tested, with all the repeated measurements included within the 95% confidence ellipses, and no overlap detected between these ellipses. Importantly, as well as distinguishing large differences between structured/unstructured DNA strands, the sensor array could discriminate two parallel G4 strands of identical length, *c-kit1* and *c-myc 2345*, indicating the array is sensitive to small changes in sequence and structure. Another notable phenomenon is that the G20 strand is well separated from the unstructured A/T/C20 strands, which is presumably due to the formation of higher-order structures *in situ*.[40] The limit of detection was excellent, and the array can work within a wide DNA concentration range of 3 - 500 nM (see Supplementary Tables 2,3 and Supplementary Fig. 20, 21).

In addition to discriminating between DNAs with large structural differences, much more subtle selectivity is possible. The hybrid (*bcl-2 2345*) and parallel (*c-myc 2345, c-kit1*) G4 structures were well separated from each other, but resided in the same general area of the scores plot. This clustering matches their similar sizes: these three G4s are all 22 or 23 nucleotides (nt) long and contain 3 stacks of G-quartets. The most notable difference was the 2-stack antiparallel strand *TBA*, which is far shorter than the other G4s (15 nt), and was located far apart from the other 3-stack antiparallel G4s in the scores plot. This shows that the array can discriminate between G4s with different sizes as well as G4s with different topology types, despite their similar structures.

The combination of dyes and hosts in the array is essential - if the two dyes are used in the absence of host, minimal nucleotide discrimination is seen (see Supplementary Fig. 25). To determine which hosts had the greatest effect on G4 differentiation, we systematically removed

array components and examined the grouping effect (see Supplementary Figs. 22 – 27). For the simple sensing shown in Figure 3, good discrimination can be achieved with 6 component arrays, using a single dye (either **PSMI** or **DSMI**, although **PSMI** is more effective) with the 5 hosts. Reducing the number of hosts decreased the discrimination power, although the combination of **DSMI/PSMI** and cavitand **2** was surprisingly effective (see Supplementary Fig. 26).

Discrimination and classification of structurally similar DNA targets

This simple screen illustrates the abilities of the host–guest array: the different molecular recognition events, combined with varying fluorescence responses, allow differentiation of nucleotide structures, even those with similar structural characteristics. The next, far more stringent test was to see if the sensor could discriminate between a series of different yet structurally similar G4 structures, and whether it could also provide classification. A suite of 23 different G4s in three family types with strand lengths from 15 to 26 nt were tested (for full sequences see Supplementary Table 1). This included six parallel (*c-kit1, c-kit2, pu22, c-myc 2345, EAD4 and PS5.M*), eight antiparallel (*TBA, Bom19, 2KF8, 442A3, TTT-L13, 2KKA, TA2, and 6FTU*) and nine hybrid structures (*AG22, bcl-2 2345, TP3, wtTel23, wtTel24, wtTel26, Tel 26, Telo24 and H24*). The topologies of these strands were confirmed by CD analysis, and each strand was confirmed to be a unimolecular G4 by gel electrophoresis (Supplementary Figs. 5, 6).

As most of these G4 targets are highly similar in size and global structure, we employed the full 12-component array to deliver the most powerful discrimination possible (see Supplementary Fig. 28 for full fluorescence response plots). The fluorescence data was collected as above, and the responses analyzed by a series of multivariate analysis tools (Fig. 4). For clarity, Figure 4a shows the average fluorescence change from 5 repeated measurements on each G4 as a single plot point. The full suite of data points is shown in Extended Data Fig. 3, including error ellipses, but the

simplified plot in Figure 4a illustrates the results nicely. The clear outcome of the PCA plots is that the sensor can easily discriminate between the parallel G4 structures (six different G4s, illustrated in red) from all the other entries. The discrimination is remarkable, as three of the G4s (*c-kit1, pu22, c-myc 2345*) have exactly the same strand length (22 nts) and the same topology, yet are fully separated on the scores plot. The discrimination between the hybrid and antiparallel G4 structures is at first glance less impressive, but closer inspection leads to some interesting observations. The grouping of the hybrid structures is quite clear, and shown in blue in Figure 4a, grouped around the centerpoint of the plot. However, there is significant variation in the positions of the antiparallel strands, with four G4s (*TTT-L3, 2KKA, 442 A3* and *2KF8*) co-locating with the hybrid structures, and the other four structures in two separate groups (*6FTU/TA2* and *TBA/Bom19*). The challenges in differentiating the hybrid and antiparallel structures can be easily explained by their structural similarities. Hybrid and antiparallel G4s both contain one or two of the four G strands in an orientation opposite the others, and both display *anti* and *syn* guanines. They are even, in certain literature cases, considered as a single topology group.[16,18] The discrimination shown between certain hybrid and antiparallel strands in our screen is dependent on their length. The antiparallel and hybrid G4s that co-locate close to each other are all of similar lengths (22-26 nts) and all display three G-quartet stacks. The outlier antiparallel structures vary in the number of G-quartet stacks, either 2 (*TBA* and *Bom19*) or 4 (*6FTU* and *TA2*), and these larger variations in structure are fully separated from the 3-stack G-quartets (either antiparallel or hybrid). In addition, some impressive discrimination effects can be seen for these G4s. Although *2KKA* has high sequence similarity with *2KF8, AG22*, *wtTel23* and *wtTel24* (varying only by the absence of one T or one A on the 3' or 5' end), it is well separated from the other strands in the PC 2 axis. Similar differentiation is also observed between *Tel26* and *wtTel26*, as well as between

Telo24 and *H24*. When an additional component was added to the PCA (i.e. PC 3, accounting for 4.46% of the overall variation in the dataset) even greater separation between the 3-stack hybrid and antiparallel structures was seen: *442 A3* was well separated from *bcl-2 2345* (Fig. 4b). The array can differentiate these G4 structures based on multiple secondary structural features, including topology, number of G-quartet stacks, and the underlying base sequence. Dissecting the contribution of individual array elements shows that changing the host is most important for differentiating the structures by their topology, while the different dyes contribute more towards revealing differences in stack numbers and sequences (see Supplementary Figs. 29, 30 for PCA plots with varying array elements). This illustrates the power of the array concept: by introducing (or removing) individual array components, the sensor can be tailored for specific targets, and is not restricted to one type of DNA structure.

The strong differentiation between different G4 topologies shown by the array suggests that it could also classify the G4s by their topology. To validate this, the responses were also subjected to Canonical Discrimination Analysis (CDA) (Fig. 4c).[12] Whereas PCA is an unsupervised classification tool that finds the greatest variance between samples and clusters the samples with a smaller variance, CDA is supervised in classification: the class of each sample (i.e. the topology of each G4) is included to maximize class discrimination. As such, the unsupervised PCA confirms differentiation of different G4 sequences, but the supervised CDA analyzes whether the array can classify the G4s by topology. In this case, the raw data with the full 12-element array (including all 5 repeats) for all 23 G4 structures were used as the input (a total of 115 samples classified by 12 characteristics), together with their topology information. The results are extremely impressive: the Canonical Scores plot (Fig. 4c) clearly shows that the targets are robustly classified into the three expected topology groups, parallel, hybrid and antiparallel (shown in red, blue and green

respectively in Fig. 4c), with 114 out of 115 samples assigned to the correct topology. Only 1 sample in the hybrid group was wrongly assigned to antiparallel (Fig. 4d), an error rate in classification of 0.7%. In addition, k-fold (k = 8) cross-validation analysis, in which 1/8 of the observations were removed from the data set and treated as unknowns, led to an average precision rate of 98.21%. This analysis corroborates the PCA results well, and validates the power of the sensor array in classifying highly similar G4 sequences by topology type.

Broadening the sensing scope

The discrimination tests were then performed in more complex media (Supplementary Figs. 32 – 39). The sensor was completely tolerant to the presence of small saccharides, with no loss of performance in the presence of 20 μM lactose. Most impressively, the sensor remained effective in the presence of 5% fetal bovine serum (FBS), and full discrimination between the initial 8 targets (A/G/T/C20, *c-kit1*, *c-myc 2345*, *bcl-2 2345, and TBA*) was possible with the **Type 1** array (Supplementary Fig. 33). Greater serum concentrations reduced the effectiveness of the sensing, but this is to be expected, considering the variety of interfering molecules in FBS. Cationic proteins such as lysozyme were the greatest interferents (Supplementary Figs. 34 – 37), and the array could only tolerate 0.1 μM lysozyme before losing selectivity.

As G-quadruplexes use alkali metal cations as structural components, changing the nature of the cation (from K^+ to Na^+) can sometimes cause a change in G4 topology. Notably, *AG22* displays a hybrid quadruplex structure in potassium phosphate buffer, but switches to an antiparallel structure in sodium phosphate.[16, 41] To determine whether the sensor array could detect these changes, we repeated the fluorescence measurements for a series of G4s in either Na^+ or K^+ phosphate buffer, using the compressed **Type 1** array with **DSMI** as dye. In this case, five G4 structures were chosen: *AG22*, with *TBA*, *c-myc 2345*, *c-kit1*, *bcl-2 2345* as controls. The sensor results are shown in Figure

5a: in the PCA scores plot in K^+ buffer, the data points corresponding to *AG22* are in close proximity to those of the hybrid G4 control of *bcl-2 2345*, with the 95% confidence ellipses overlapping significantly. The parallel and antiparallel controls are fully distinguished. When the same measurements are made in buffer containing Na^+ ions, the changes become obvious: no longer is *AG22* co-located with the hybrid *bcl-2 2345*, but moves into close proximity with the antiparallel *TBA*. These observations were corroborated by CD spectra (Supplementary Fig. 4). While the sensor is sensitive to changes in buffer, it is still fully capable of sensing the hybrid→antiparallel structural change of *AG22*, which was one of the most challenging conformational variations to detect among our 23 G4 candidates in identical buffer conditions.

The sensor can also detect changes in G4 concentration in a mixture of oligonucleotides. The *c-myc 2345* G4 was combined with the A20 ssDNA strand in varying proportions, and added to a solution of the compressed **Type 1** array with **DSMI** as the dye. The fluorescence responses (Supplementary Fig. 40) were subjected to PCA, and the scores plot shown in Fig. 5b. As the proportion of A20/*c-myc 2345* varies, the signals on the scores plot move towards the relevant signals for 100% A20/*c-myc 2345* (shown in blue and red respectively), with the changes dominated by the position on PC 1. Even in the presence of competing nucleotides, the selectivity of the sensor for the folded structure is retained.

CONCLUSIONS

Here we have shown that an arrayed suite of synthetic hosts and dyes is capable of sensing oligonucleotide secondary structures. Multiple recognition mechanisms can be exploited to create a unique sensing fingerprint consisting of variable fluorescence enhancements in the presence of different DNA strands. Multivariate analysis of these sensing fingerprints enables discrimination

between DNA structures, which can be as different as single stranded A20 and G20, or as similar as G-quadruplex strands that vary only slightly in structure and size.

Most importantly, the array is able to both differentiate and classify the G4 structures at the same time. Simply differentiating between different G4 types has limited utility by itself: the ability to differentiate between different G4 structures while also enabling their classification is essential for detecting the presence of specific secondary structural motifs as disease markers, for example. Whereas CD spectroscopy can quickly identify folding topology, it is incapable of distinguishing between different G4 structures of the same folding type. Mass spectrometry and sequencing techniques can easily detect differences in size, but are poor at determining the secondary structure in solution. Our array is capable of both types of analysis at the same time, in a simple, non-invasive manner, which requires little sample and does not require discovery and synthesis of specific ligand targets for each oligonucleotide type. The selectivity is excellent: different G4s that display the same folding topology can be easily differentiated by the number of G-quartets and sequence differences. It can even detect changes in G4 folding pattern in different types of complex media, in the presence of interfering small molecules and in mixtures of nucleotides. This method could potentially be employed in computational modeling to reveal close correlation between sequence and structure, paving the way to accurate prediction of nucleic acid folding based on its sequence and their fluorescence responses in our array.

REFERENCES

1. Belmont, P., Constant, J.-F. & Demeunynck, M. Nucleic acid conformation diversity: from structure to function and regulation. *Chem. Soc. Rev.* **30**, 70-81 (2001).
2. Balasubramanian, S., Hurley, L. H., & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10**, 261-275 (2011).

3. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).
4. Pinheiro, A. V., Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **6**, 763-772 (2011).
5. Jaeger, L. & Chworos, A. The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.* **16**, 531-543 (2006).
6. Winnerdy, F.R., *et al.* NMR solution and X-ray crystal structures of a DNA containing both right-and left-handed parallel-stranded G-quadruplexes *Nucleic Acids Res.* **47**, 8272-8281 (2019).
7. Salgado, G. F., Cazenave, C., Kerkour, A. & Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **6**, 3314-3320 (2015).
8. del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex secondary structure obtained from Circular Dichroism spectroscopy. *Angew. Chem. Int. Ed.* **57**, 7171-7175 (2018).
9. Eubanks, C. S., Forte, J. E., Kapral, G. J. & Hargrove, A. E. Small molecule-based pattern recognition to classify RNA structure. *J. Am. Chem. Soc.* **139**, 409-416 (2017).
10. You, L. Zha, D. & Anslyn, E.V. Recent advances in supramolecular analytical chemistry using optical sensing. *Chem. Rev.* **115**, 7840-7892 (2015).
11. Eubanks, C. S. et al. Visualizing RNA conformational changes via pattern recognition of RNA by small molecules. *J. Am. Chem. Soc.* **141**, 5692-5698 (2019).
12. Stewart, S., Ivy, M. A. & Anslyn, E. V. The use of principal component analysis and discriminant analysis in differential sensing routines. *Chem. Soc. Rev.* **43**, 70-84 (2014).
13. del Villar-Guerra, R., Gray, R.D. Trent, J. O. & Chaires, J. B. A rapid fluorescent indicator displacement assay and principal component/cluster data analysis for determination of ligand–nucleic acid structural selectivity. *Nucleic Acids Res.* **46**, e41 (2018).
14. Huppert, J. L. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37**, 1375-1384 (2008).
15. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770-780 (2012).
16. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402-5415 (2006).

17. Dolinnaya, N. G., Ogloblina, A. M. & Yakubovskaya, M. G. Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry* **81**, 1602-1649 (2016).
18. Kwok, C. K. & Merrick, C. J. G-Quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* **35**, 997-1013 (2017).
19. Puig Lombardi, E. & Londoño-Vallejo, A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* **48**, 1-15 (2020).
20. Zuffo, M. *et al.* More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.* **46**, e115 (2018).
21. Felsenstein, K. M., *et al.* Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression. *ACS Chem. Biol.* **11**, 139-148 (2016).
22. Pinalli, R., Pedrini, A. & Dalcanale, E. Biochemical sensing with macrocyclic receptors. *Chem. Soc. Rev.* **47**, 7006-7026 (2018).
23. Dsouza, R. N., Hennig, A. & Nau, W. M. Supramolecular tandem enzyme assays. *Chem.-Eur. J.* **18**, 3444-3459 (2012).
24. Hennig, A., Bakirci, H. & Nau, W. M. Label-free continuous enzyme assays with macrocycle-fluorescent dye complexes. *Nat. Methods* **4**, 629-632 (2007).
25. Peacor, B. C., Ramsay, C. M. & Waters, M. L. Fluorogenic sensor platform for the histone code using receptors from dynamic combinatorial libraries. *Chem. Sci.* **8**, 1422-1428 (2017).
26. Minaker, S. A., Daze, K. D., Ma, M. C. F. & Hof, F. Antibody-free reading of the histone code using a simple chemical sensor array. *J. Am. Chem. Soc.* **134**, 11674-11680 (2012).
27. Florea, M. & Nau, W. M. Implementation of anion-receptor macrocycles in supramolecular tandem assays for enzymes involving nucleotides as substrates, products, and cofactors. *Org. Biomol. Chem.* **8**, 1033-1039 (2010).
28. Liu, Y. *et al.* Selective heavy element sensing with a simple host–guest fluorescent array. *Anal. Chem.* **89**, 11113-11121 (2017).
29. Gill, A. D. *et al.* Sensing of citrulline modifications in histone peptides by deep cavitand hosts. *Chem. Commun.* **55**, 13259-13262 (2019).
30. Murat, P., Singh, Y. & Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev.* **40**, 5293-5307 (2011).

31. Biros, S. M., Ullrich, E. C., Hof, F., Trembleau, L. & Rebek, J. Kinetically stable complexes in water: the role of hydration and hydrophobicity. *J. Am. Chem. Soc.* **126**, 2870-2876 (2004).
32. Liu, Y. *et al.* Site-selective sensing of histone methylation enzyme activity via an arrayed supramolecular tandem assay. *J. Am. Chem. Soc.* **139**, 10964-10967 (2017).
33. Liu, Y. *et al.* Selective sensing of phosphorylated peptides and monitoring kinase and phosphatase activity with a supramolecular tandem assay. *J. Am. Chem. Soc.* **140**, 13869-13877 (2018).
34. Mosca, S., Yu, Y. & Rebek, J. Preparative scale and convenient synthesis of a water-soluble, deep cavitand. *Nat. Protoc.* **11**, 1371-1387 (2016).
35. Pinalli, R., *et al.* The origin of selectivity in the complexation of N-methyl amino acids by tetraphosphonate cavitands. *J. Am. Chem. Soc.* **138**, 8569-8580 (2016).
36. Menozzi, D., *et al.* Thermodynamics of host–guest interactions between methylpyridinium salts and phosphonate cavitands. *Supramol. Chem.* **22**, 768-775 (2010).
37. Yang, Q. *et al.* Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: II. The binding characterization with specific intramolecular G-quadruplex and the recognizing mechanism. *Nucleic Acids Res.* **38**, 1022-1033 (2010).
38. Früh, A. E., Artoni, F., Brighenti, R. & Dalcanale, E. Strain Field Self-Diagnostic Poly(dimethylsiloxane) Elastomers. *Chem. Mater.* **29**, 7450-7467 (2017).
39. Thordarson, P. Determining association constants from titration experiments in supramolecular chemistry. *Chem. Soc. Rev.* **40**, 1305–1323 (2011).
40. Sengar, A., Heddi, B. & Phan, A. T. Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G15 stretch. *Biochemistry* **53**, 7718−7723 (2014).
41. Ambrus, A. *et al.* Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **34**, 2723-2735 (2006).

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (CHE-1707347 to W.Z and R.J.H.), and MIUR (PRIN 20179BJNA2 to E.D.) for funding.

AUTHOR CONTRIBUTIONS

R.J.H. and W.Z. conceived and designed the experiments and wrote the paper. J.C. performed the arrayed sensing experiments and statistical analysis with help from L.W. and J. L. Chemical synthesis and optical analysis of hosts and dyes was performed by B.L.H., A.D.G, A.F., R.P. and E.D. All authors contributed to manuscript creation and proofing.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. **A host–guest sensor system for label-free classification and differentiation of G-quadruplex structures.** a) Illustration of the supramolecular sensing concept: the cationic dyes can be bound by both the DNA targets and the various host molecules, and their emission is modulated by each recognition event in a complex equilibrium. When multiple hosts and dyes are used, an array-based sensor system can be created for pattern recognition *via* multivariate analysis. b) Representation of the common DNA G-quadruplex (G4) structural motifs targeted here; c) Structures of hosts **1-5** and d) dyes. Cartoon illustrations of each host/dye used in later Figures are adjacent to their structures, showing the relative position of their charged groups at the upper or lower rim. **DSMI** = *trans*-4-[4-(dimethylamino)-styryl]-1-methyl-pyridinium iodide, **PSMI** = including *trans*-4-[4-(piperidino)-styryl]-1-methyl-pyridinium iodide.

Figure 2. Effect of DNA strands on the emission profile of various host–dye complexes. Normalized fluorescence response curves corresponding to the emission of **DSMI** dye in the presence of different DNA strands upon titration of hosts a) **1**; b) **2**; c) **5**, illustrating the effect of the DNA structure on the emission of the various host–guest complexes. [**DSMI**] = 0.625 μM, [DNA] = 0.1 μM, [**host**] = 0-2 μM, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. The normalization process defines F_0 as the emission at [DNA] = 0. The raw, unnormalized plots corresponding to this data can be found in Extended Data Fig. 2, along with expanded versions of the normalized plots. For the full suite of titration plots with all hosts **1-5** and both dyes, see Supplementary Figs. 18 and 19. Error bars represent the standard deviation of 3 repeated measurements. d) Illustration of the various equilibria present when the hosts and dyes are exposed to DNA G-quadruplexes. These multiple recognition events allow application of this system in an array-based format for differentiation of different G4s (see Figs. 3-5).

Figure 3. Selective array-based sensing of variable DNA structures. a) Fluorescence (F) responses upon addition of the eight DNA strands to the **host•dye** sensor components, F_0 = emission at [DNA] = 0. Each array component shows a different fluorescence response, and this array of responses can be subjected to multivariate analysis to differentiate the DNA structures. b)

Principal Component Analysis (PCA) scores plot generated from the data in part a), using the 12-component array (the combination of **Type 1** and **Type 2** arrays). The scores plot of the first two principal components, in total summarizing more than 87% of the variation contained in the data, provides a visualization of how the 8 DNA strands can be differentiated by our array. Different strands are well separated from each other, with the 5 repeats of the same strand clustered tightly together. Error bars indicate standard deviation of 5 repeated measurements. [**Dye**] = 0.625 μM in both arrays, with **Type 1 Array** using **DSMI** and [**1**, **4**, or **5**] = 0.25 μM or [**2** or **3**] = 0.50 μM, or no cavitand, and **Type 2 Array** using **PSMI** and [**1**, **3**, **or 5**] = 0.25 μM, [**2**] = 1.0 μM, or [**4**] = 0.50 μM, or no cavitand. [DNA] = 0.1 μM, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. Ellipses indicate 95% confidence.

Figure 4. **Classification and discrimination of a suite of 23 G-quadruplex structures.** a) 2D PCA scores plot of the first two principal components (PC) generated from the fluorescence responses of 23 G4 strands to the sensing array. Each point is the average of 5 repeated measurements (see Extended Data Fig. 3. for the individual repeats). PC1 and PC2 in total summarize close to 90% of the variation contained in the data. Their scores plot illustrates that all the parallel G4s (red dots) locate close to each other, but distant from the hybrid and antiparallel ones. The 3-stack antiparallel strands cannot be separated from some of the 3-stack hybrid G4s, but the 4- and 2-stack antiparallel G4s can be differentiated from the 3-stack strands on this plot. b) 3D PCA scores plot generated from the same data as in part a), illustrating only the hybrid (blue) and antiparallel (green) G4 strands. The third PC axis better illustrates the differentiation of the 2-stack and 3-stack antiparallel G4s. c) Canonical scores plot resulted from Canonical Discriminant Analysis of the individual responses from the 12 array elements to the 115 samples grouped by topology. The scores plot illustrates clear grouping of most of the 115 samples into the correct topology class, except for one hybrid G4. d) Tabulation of the CDA results to illustrate the classification error rate of 1/115 samples. Sensor conditions identical to those described in Figure 3; 12-component array used. Parallel G4s labeled in red, hybrid G4s labeled in blue, antiparallel G4s labeled in green.

Figure 5. **More complex sensing with the array, which can detect structural topology switching and changing concentration of specific G4s in a mixture.** a) Sensing the structural change of the *AG22* G4 from hybrid to antiparallel with different structural cations (K^+ (left) to Na^+ (right)). PCA scores plot from the fluorescence responses of 5 G4 strands, *AG22* and 4 controls ([DNA] = 0.1 μM) in either 10 mM KH_2PO_4/K_2HPO_4 or NaH_2PO_4/Na_2HPO_4 buffer, Ex/Em = 485nm/605nm. In the PCA scores plot generated from DNAs in a K^+ buffer, *AG22* co-locates with the hybrid G4 of bcl-2 2345, but the plot from DNAs in the Na^+ buffer shows *AG22* next to the antiparallel G4 of TBA, confirming the topology switch. Both scores plots show good discrimination of the G4s with different topology, with the two first principal components summarizing more than 93% of the total variance of the data. b) Detection of varying concentration of a specific G4 in a mixture of different DNAs. PCA scores plot from the fluorescence responses of solutions containing various molar ratios of the *c-myc 2345* G4 : polyA20 strand ([total DNA] = 0.5 μM) in 10 mM KH_2PO_4/K_2HPO_4 buffer, Ex/Em = 485nm/600nm. Both plots were generated using the **Type 1** sensing array; all buffers with 1 mM EDTA, at pH = 7.4. The first two principal components that summarize close to 99% of the overall data variance. The scores plot shows the

mixtures containing higher proportions of G4 moving to the right of the plot with higher PC1 values, while the ones having lower G4 contents moving to the left, with lower PC1 values. Fluorescence data for these plots can be found in Supplementary Fig. 40.

TABLES

Table 1. Binding Affinities Between Dyes and Hosts/DNA[a]

Dye	K_d (**1**), μM[b]	K_d (**2**), μM[b]	K_d (**3**), μM[b]	K_d (**4**), μM[b]	K_d (**5**), μM[b]	K_d (***c-myc 2345***), μM[c]
DSMI	5.3±0.4	72.3±2.4	6.1±1.3	20.3±1.5	2.8±1.4	247±68
PSMI	8.2±0.9	51.7±2.8	8.6±1.3	5.8±1.22	3.3±1.8	112±27

[a] 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4. [b] Measured by fluorescence titrations,[39] with the [dye] = 0.625 μM for hosts **1-4**, and 10 μM for host **5**. [c] Measured by ITC at 25 °C, with increasing amounts of **DSMI** added to 10 μM *c-myc 2345* or **PSMI** to 40 μM *c-myc 2345*.

METHODS

General Details

Cavitands **1**,[31] **2**,[29] **3**,[34] and **4**,[34] as well as phosphonate cavitand **5**[35] and **PSMI** fluorophore[29] were synthesized according to literature procedures. NMR spectra were recorded on a Bruker Avance NEO 400 MHz. Deuterated NMR solvents were obtained from Cambridge Isotope Laboratories, Inc., Andover, MA, and used without further purification. All other materials, including *trans*-4-[4-(dimethylamino)-styryl]-1-methyl-pyridinium iodide (**DSMI**) were obtained from Aldrich Chemical Company (St. Louis, MO), or Fisher Scientific (Fairlawn, NJ), and were used as received. Solvents were dried through a commercial solvent purification system (Pure Process Technologies, Inc.). Oligonucleotides were purchased from Integrated DNA Technologies (IDT) with standard desalting and no further purification, the sequence and topology information of which were given in Supplementary Table 1. All DNA solutions were prepared in 10 mM K_2HPO_4/KH_2PO_4 or Na_2HPO_4/NaH_2PO_4 buffer at pH 7.4, both containing 1 mM EDTA (referred

as the K^+ or Na^+ buffer in the text); and their concentrations were determined by NanoDrop 2000 (Thermo Fisher Scientific) using the corresponding molar extinction coefficients provided by IDT after background subtraction. Before fluorescence or CD measurement, the DNAs were diluted to 1 μM with the K^+ or Na^+ buffer; and re-annealed to form the most stable folding topology, in which the DNA solution was heated at 95 °C for 5 min and then kept at 4°C for overnight. Fluorescence measurements were performed with a Perkin Elmer Wallac 1420 Victor 2 Microplate Reader (PerkinElmer) with the Ex/Em wavelengths at 485/605 nm. Principal Component Analysis (PCA) and confidence ellipses were performed with RStudio (Version 1.2.5019), an integrated development environment (IDE) for R (version 3.6.1). Canonical Discriminant Analysis was conducted with OriginPro 2018. Classification performance was evaluated by k-fold (k = 8) cross validation using Python 3, with Linear Discriminant Analysis (LDA) as the supervised classification model.

Experimental Procedures

Fluorescence measurements. In general, the fluorescence assay was carried out by mixing 10 μL of the fluorescent guest (6.25 μM **DSMI** or **PSMI** in water), 10 μL of the cavitand (in water) or water, 70 μL of the incubation buffer, and 10 μL of 1 μM DNA in the 96-well plate, resulting in a final total volume around 100 μL in 10 mM K_2HPO_4/KH_2PO_4 (or Na_2HPO_4/NaH_2PO_4) and 1 mM EDTA at pH 7.4. The mixture was incubated with mild shaking for 15 min at room temperature, before the fluorescence signal (F) was recorded in a Perkin Elmer Wallac 1420 Victor 2 Microplate Reader (PerkinElmer) with the Ex/Em wavelengths at 485/605 nm.

Array constituents. Type 1 Array: hosts **1-5** or no host, **DSMI** dye. [**DSMI**] = 0.625 μM, [**1**, **4**, or **5**] = 0.25 μM or [**2** or **3**] = 0.50 μM. **Type 2 Array:** hosts **1-5** or no host, **PSMI** dye. [**PSMI**] = 0.625 μM, [**1**, **3**, **or 5**] = 0.25 μM, [**2**] = 1.0 μM, or [**4**] = 0.50 μM. **10-component array:**

Combination of **Type 1** and **Type 2** arrays, not including "no host". **12-component array:** Combination of **Type 1** and **Type 2** arrays.

Circular Dichroism (CD). CD spectra were recorded on a Jasco J-815 CD spectrophotometer over a wavelength range of 200 nm–350 nm at room temperature, with a band width of 1 nm and a data pitch of 1 nm. The instrument scanning speed was set at 100 nm/min, with a response time of 1 s. 10 μM of 200 μL oligonucleotide solution prepared in the K^+ buffer or Na^+ buffer then was pipetted into a quartz cell with a path length of 0.1 cm. The CD spectra were presented with baseline-correction in which the background signal from the buffer was subtracted.

Gel Electrophoresis. The quality of the DNA solution was inspected by native gel electrophoresis using a precast gradient (4%-20%) PAGE gel. 10 μL of a 1 μM DNA solution was loaded to the gel, after being denatured at 95°C for 5 min, cooled on ice for 10 min and then at room temperature for 30 min. The gel was run at 120 V for 90 min at room temperature in 1×TBE buffer, and stained with SYBR Gold (1:10000 dilution) before imaged using the UV transilluminator (SPECTROLINE).

Binding Affinity Measurements: Fluorescence Titrations. For Hosts **1-4**, the fluorescence titration curves were obtained by adding 0-45 μM host **1-4** into the solution that contained 0.625 μM **DSMI** or **PSMI** in the K^+ buffer. For host **5**, the titration was carried in the same manner, by using 0-10 μM host **5** and 10 μM **DSMI** or **PSMI**, because the dye fluorescence was only slightly quenched by the host. Fluorescence was recorded after 15 min of mixing in the plate reader using the instrument setting described above. The binding affinity calculations were performed using the fitting program provided at http://supramolecular.org.[39] The UV 1:1 filter within the Bindfit function was used along with the Nelder-Mead fitting method,[39] with the "subtract initial value" option selected.

Binding Affinity Measurements: ITC Titrations. All ITC experiments were performed using a MicroCal iTC200 (GE Healthcare, Freiburg, Germany) with a stirring rate of 800 rpm. The baseline was stabilized prior to the experiment, and a pre-injection delay was set to 60 s. A stock solution of **DSMI** or **PSMI** at 3 mM in K^+ buffer was added in 2 μL aliquots to the *c-myc 2345* solution of 10 and 40 μM (also in K^+ buffer), respectively. **DSMI** binds about 2 times more weakly than **PSMI**, and thus requires a higher dye:DNA ratio to reach a plateau. All experiments were conducted at 25 °C. The heat of dilution, measured by the injection of titrant into the buffer solution, was subtracted for each titration to obtain the net reaction heat value. Curve fitting was performed by the MicroCal program using the One Set of Sites model.

Methods Discussion/Additional Points

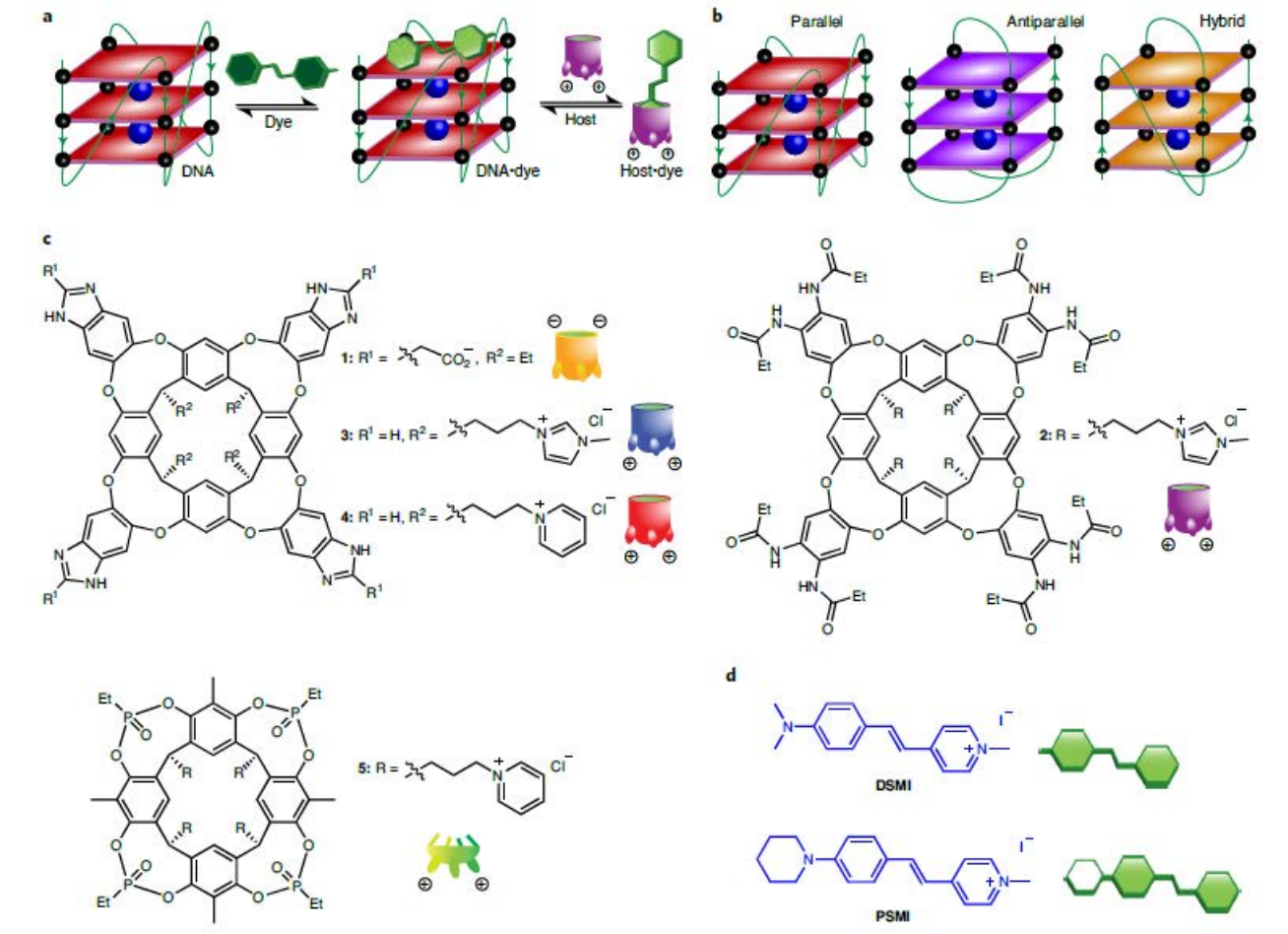
Analysis of Fluorescence Responses. The fluorescence response titrations shown in Fig. 1a-c are normalized (y axis = F/F_0) to the response of cavitand–dye in the absence of DNA, i.e. F_0 is defined as the fluorescence recorded for that concentration of host and guest when [DNA] = 0 (purple lines in Extended Data Fig 2a,c,e). As hosts **1-4** cause an enhancement in emission when binding **DSMI/PSMI**, the raw fluorescence increases variably, depending on host, in each case. Host **5** causes a slight quenching of the dyes upon binding. The unnormalized plots (y axis = fluorescence counts) are shown in Extended Data Fig 2a,c,e (for **DSMI** and hosts **1**, **2**, **5**) and Supplementary Figs. 18 and 19 (all dye/host combinations). The normalized emission plots allow simple visualization of the effect of the target DNA on the emission profile, by removing the effect of the host–dye emission on the signal. The effect of the various DNAs and the mechanism of their sensing is the important point of our study, as the effect of hosts on the dyes is well-known.[29,31,34,38]

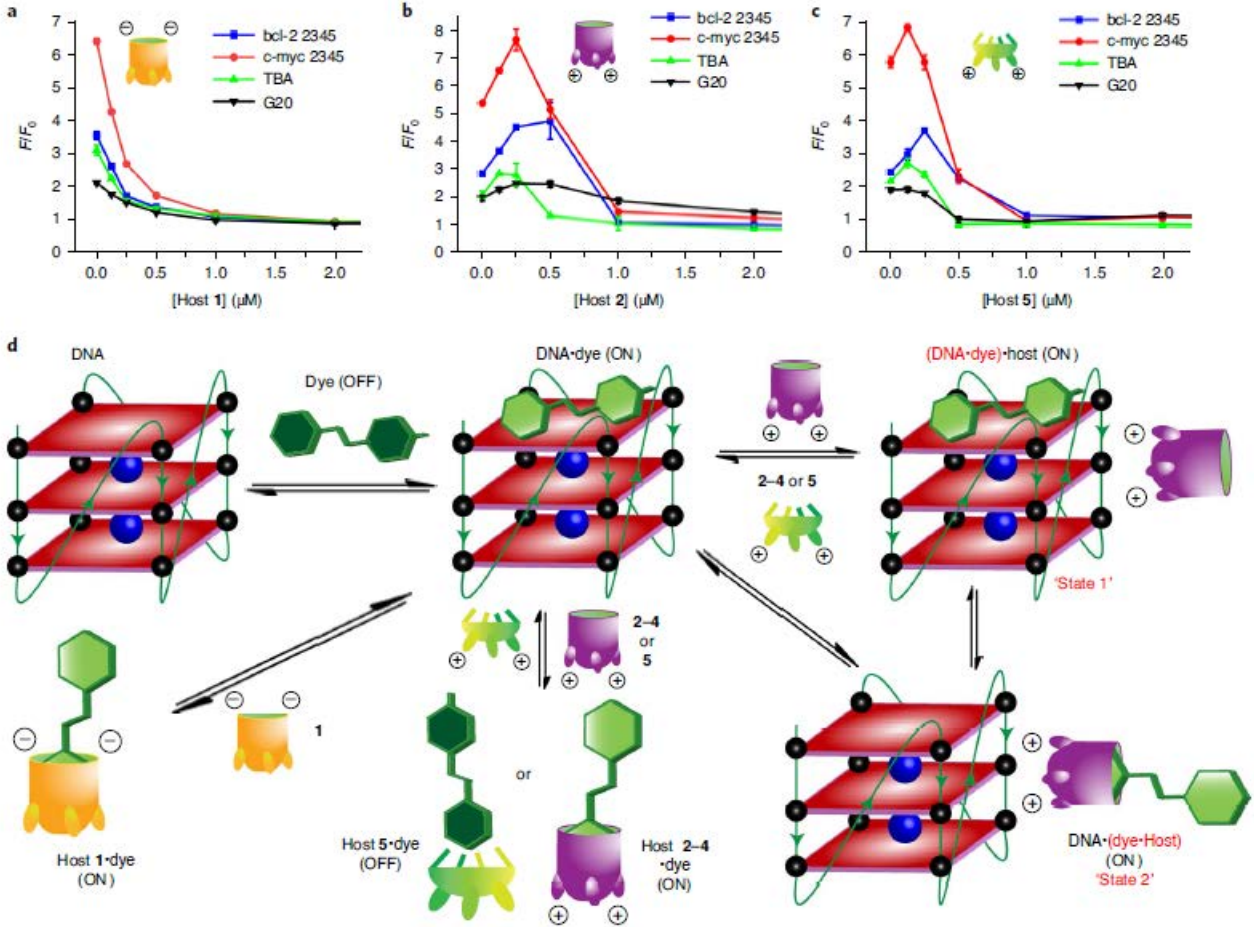
Methods used to Calculate Binding Affinities. The **dye•host** affinities (Table 1, Supplementary Fig. 14) were measured by fluorescence emission titrations and calculated *via* the BindFit fitting

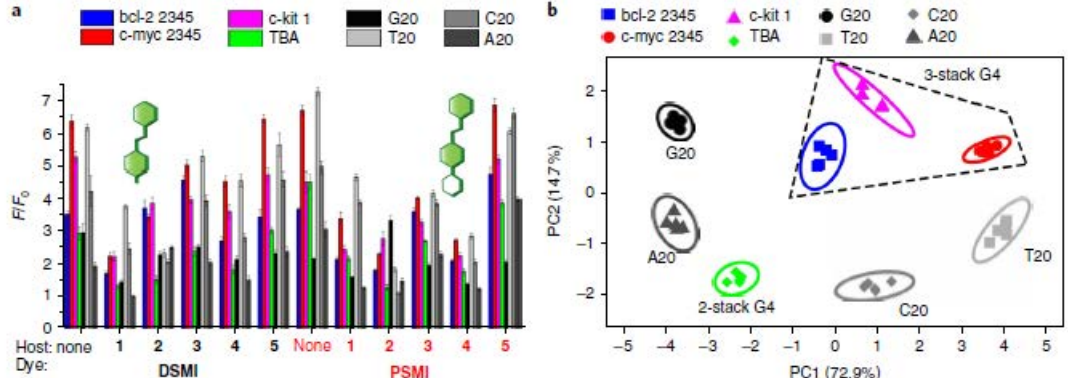
program at www.supramolecular.org.[39] The affinity between the DNA G4 *c-myc-2345* and the two dyes was measured by ITC (see Supplementary Figs. 15, 16), as the emission changes were too small for accurate analysis *via* fluorescence titrations. Determining accurate affinity measurements and the specific location of binding were complicated by the low, multivalent affinity of both dyes for the *c-myc-2345* G4, but simple fitting using the OneSites model gave with approximate binding affinities on the order of K_d = 100-250 μM for both **DSMI** and **PSMI**. The smaller **DSMI** showed a greater multivalency than the larger **PSMI**, albeit with a lower overall affinity. The competitive binding of the dyes to both DNA and hosts is also shown by adding increasing [DNA] to **host•dye** complexes (see Supplementary Figs. 20, 21). In this case, the fluorescence increases due to greater [**DNA•dye**]. The signal changes vary depending on the nature of host, dye and DNA as before.

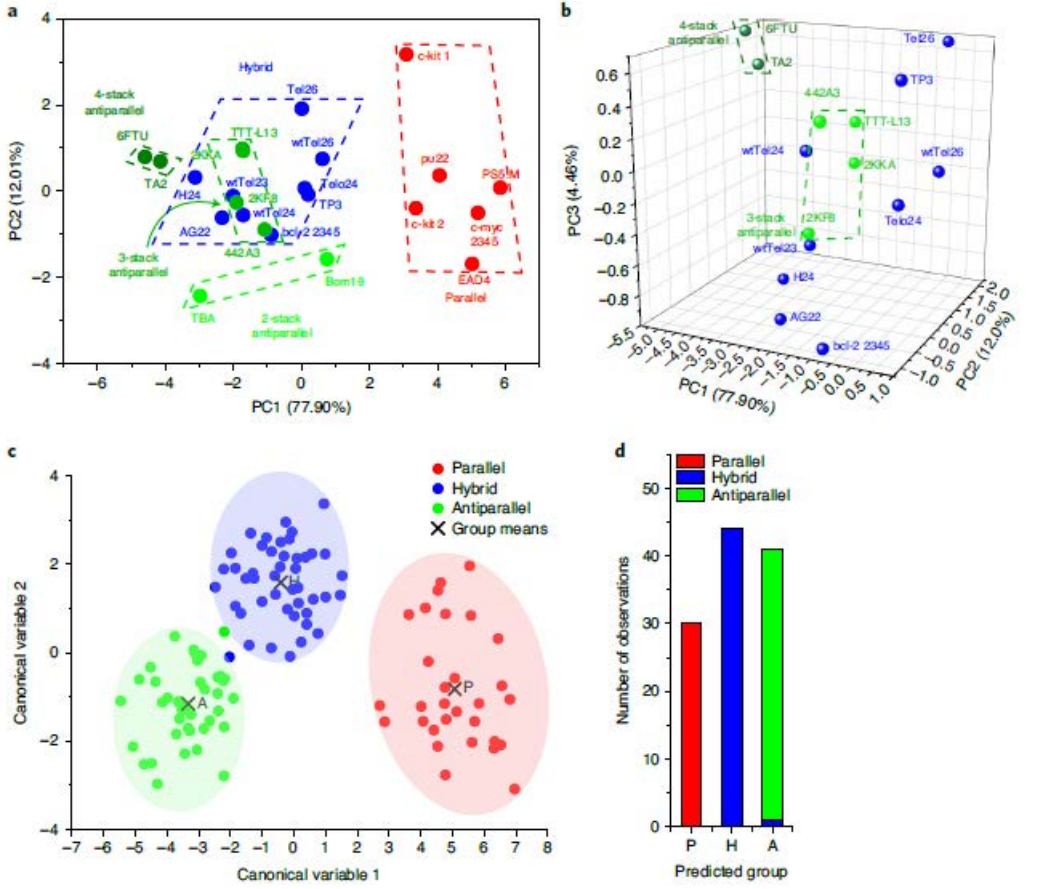
DATA AVAILABILITY STATEMENT

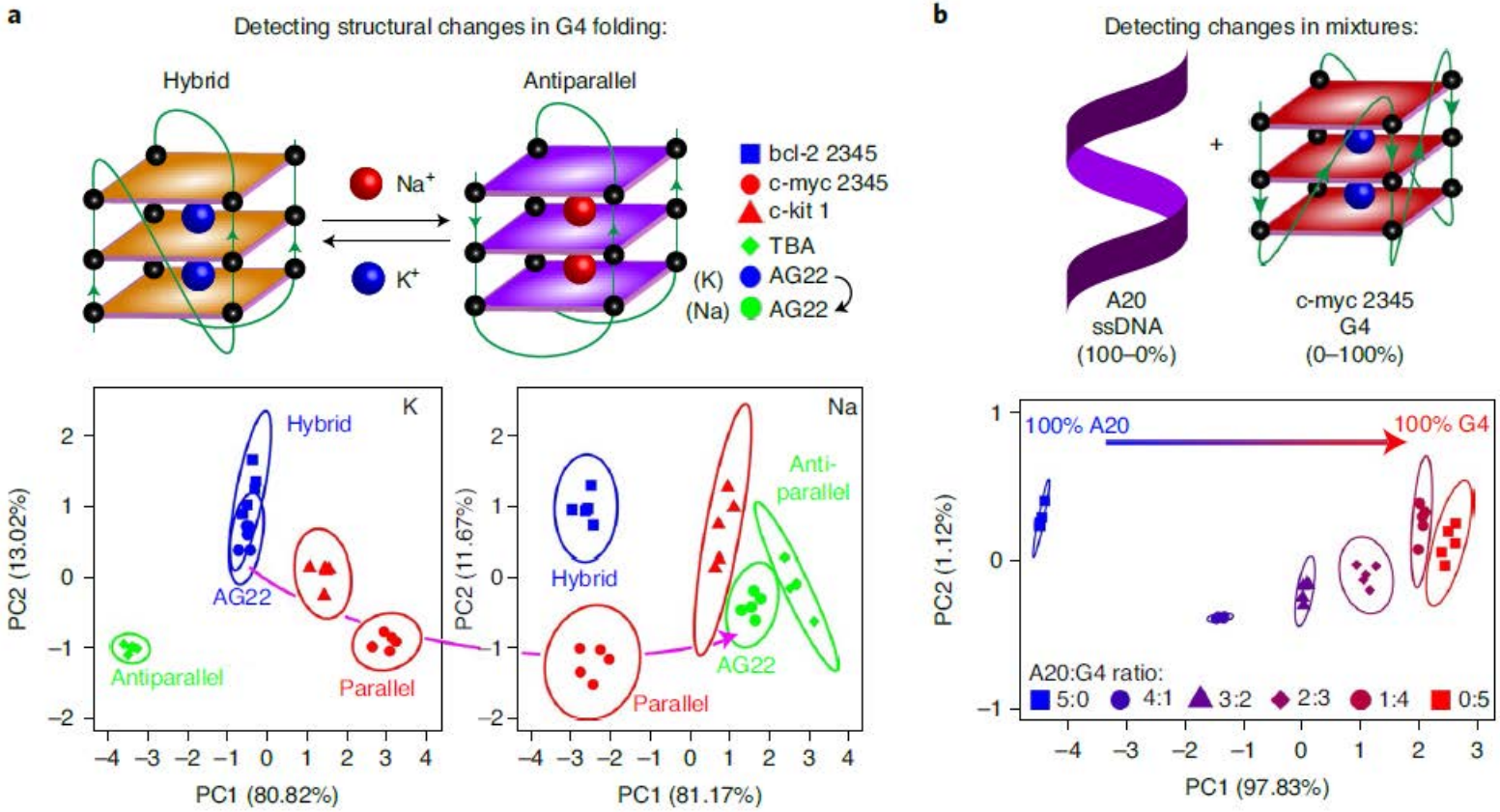
The fluorescence datasets, and multivariate analysis data generated in this study are available as Source data, linked to the corresponding data figures in the online version of this manuscript. All other datasets used in background experiments shown in the Supplementary Information are available from the corresponding authors upon reasonable request.


a
DNA
Dye
DNA-dye
Host
Host-dye
b
Parallel
Antiparallel
Hybrid
c
1: R1 = , R2 = Et
3: R1 = H, R2 =
4: R1 = H, R2 =
2: R =
5: R =
d
DSMI
PSMI


a
b
c
bcl-2 2345
c-myc 2345
TBA
G20
F/F0
[Host 1] (μM)
[Host 2] (μM)
[Host 5] (μM)
d
DNA
Dye (OFF)
DNA·dye (ON)
(DNA·dye)·host (ON)
2–4 or 5
'State 1'
1
2–4 or 5
Host 1·dye (ON)
Host 5·dye (OFF)
or
Host 2–4 ·dye (ON)
DNA·(dye·Host) (ON)
'State 2'


a
bcl-2 2345
c-myc 2345
c-kit 1
TBA
G20
T20
C20
A20
F/F0
Host: none
Dye:
DSMI
PSMI
b
3-stack G4
G20
A20
2-stack G4
C20
T20
PC1 (72.9%)
PC2 (14.7%)


a
PC2 (12.01%)
PC1 (77.90%)
Hybrid
4-stack antiparallel
6FTU
TA2
TTT-L13
2KKA
wtTel26
Tel26
wtTel23
2KF8
wtTel24
Tel24
TP3
H24
AG22
bcl-2 2345
442A3
3-stack antiparallel
Bom19
TBA
2-stack antiparallel
c-kit 1
pu22
PS5.M
c-kit 2
c-myc 2345
EA04
Parallel
b
PC3 (4.46%)
PC1 (77.90%)
PC2 (12.0%)
4-stack antiparallel
6FTU
TA2
Tel26
TP3
442A3
TTT-L13
wtTel24
2KKA
wtTel26
3-stack antiparallel
2KF8
Tel24
wtTel23
H24
AG22
bcl-2 2345
c
Parallel
Hybrid
Antiparallel
Group means
Canonical variable 2
Canonical variable 1
H
P
A
d
Parallel
Hybrid
Antiparallel
Number of observations
Predicted group
P
H
A


a
Detecting structural changes in G4 folding:
Hybrid
Antiparallel
Na+
K+
bcl-2 2345
c-myc 2345
c-kit 1
TBA
AG22
AG22
(K)
(Na)
K
Hybrid
AG22
Antiparallel
Parallel
PC2 (13.02%)
PC1 (80.82%)
Na
Hybrid
Anti-
parallel
AG22
Parallel
PC2 (11.67%)
PC1 (81.17%)
b
Detecting changes in mixtures:
A20
ssDNA
(100–0%)
c-myc 2345
G4
(0–100%)
100% A20
100% G4
A20:G4 ratio:
5:0
4:1
3:2
2:3
1:4
0:5
PC2 (1.12%)
PC1 (97.83%)

Selective Discrimination and Classification of G-Quadruplex Structures with a Host-Guest Sensing Array

Junyi Chen,[a,b] Briana L. Hickey,[a] Linlin Wang,[a] Jiwon Lee,[a] Adam D. Gill,[c] Alessia Favero,[d] Roberta Pinalli,[d] Enrico Dalcanale,[d] Richard J. Hooley[a,c]* and Wenwan Zhong[a,b]*

[a]*Department of Chemistry,* [b]*Environmental Toxicology Program,* [c]*Department of Biochemistry and Molecular Biology, University of California, Riverside, CA 92521, USA.*

[d] *Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, 43124 Parma, and INSTM, UdR Parma, Italy.*

E-mail: richard.hooley@ucr.edu; wenwan.zhong@ucr.edu

Supplementary Information

Table of Contents

1. DNA Sequences and Characterization

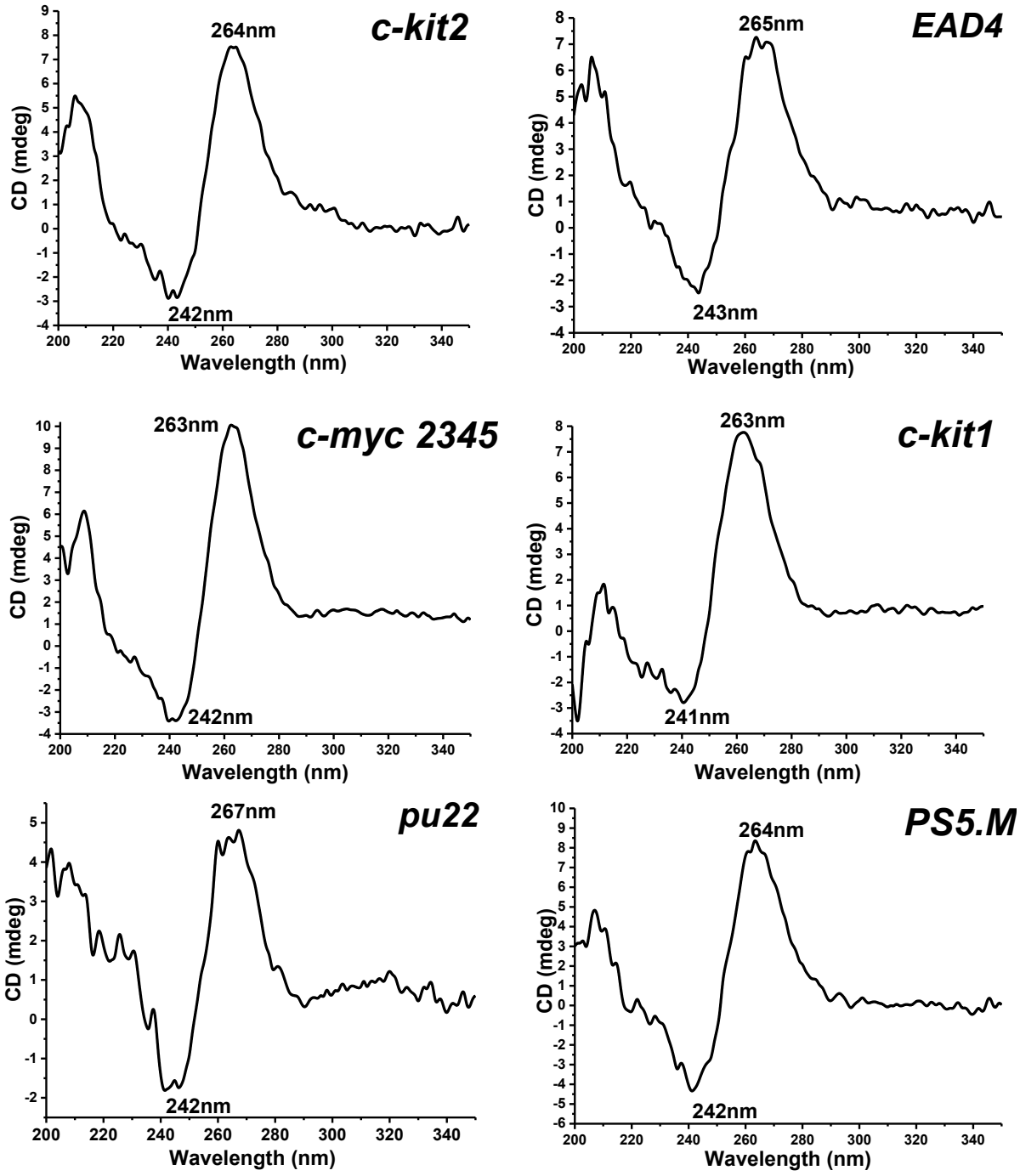
1.1 DNA sequences

Supplementary Table 1 G4 sequences and topology types as reported in K^+ buffer (10mM K_2HPO_4/KH_2PO_4, 1mM EDTA, pH 7.4), except for AG22, the topology of which in both K^+ and Na^+ buffer (10 mM Na_2HPO_4/NaH_2PO_4) was reported.

Name	Sequence	G4 type	Bases	Ref.
c-kit2	CGGGCGGGCGCGAGGGAGGG	Parallel	20	1
EAD4	CTGGGTTGGGTTGGGTTGGGA	Parallel	21	2
c-myc 2345	TGAGGGTGGGGAGGGTGGGGAA	Parallel	22	1, 3, 4
c-kit1	AGGGAGGGCGCTGGGAGGAGGG	Parallel	22	3, 5
pu22	TGAGGGTGGGTAGGGTGGGTAA	Parallel	22	1, 6
PS5.M	GTGGGTCATTGTGGGTGGGTGTGG	Parallel	24	7
AG22	AGGGTTAGGGTTAGGGTTAGGG	Na^+ Antiparallel K^+ Hybrid	22	1, 3
bcl-2 2345	GGGCGCGGGAGGAATTGGGCGGG	Hybrid	23	3, 8
wtTel23	TAGGGTTAGGGTTAGGGTTAGGG	Hybrid	23	9
TP3	TGGGGGCCGAGGCGGGGCTTGGG	Hybrid	23	10
wtTel24	TAGGGTTAGGGTTAGGGTTAGGGT	Hybrid	24	9
Telo24	TTAGGGTTAGGGTTAGGGTTAGGG	Hybrid	24	11
H24	TTGGGTTAGGGTTAGGGTTAGGGA	Hybrid	24	3
wtTel26	TTAGGGTTAGGGTTAGGGTTAGGGTT	Hybrid	26	9
Tel26	AAAGGGTTAGGGTTAGGGTTAGGGAA	Hybrid	26	12
TBA	GGTTGGTGTGGTTGG	Antiparallel	15	1, 3
Bom19	TAGGTTAGGTTAGGTTAGG	Antiparallel	19	14
2KF8	GGGTTAGGGTTAGGGTTAGGGT	Antiparallel	22	15
442A3	GGGTTT TGGGTTTTGGGTTGGG	Antiparallel	22	16
TTT-L13	AGGGTT TGGGTTAGGGTTTGGGT	Antiparallel	23	13
2KKA	AGGGTTAGGGTTAGGGTTAGGGT	Antiparallel	23	15
TA2	GGGGTTGGGGTGTGGGGTTGGGG	Antiparallel	23	2
6FTU	GGGGGAGGGGTACAGGGGTACAGGGG	Antiparallel	26	17

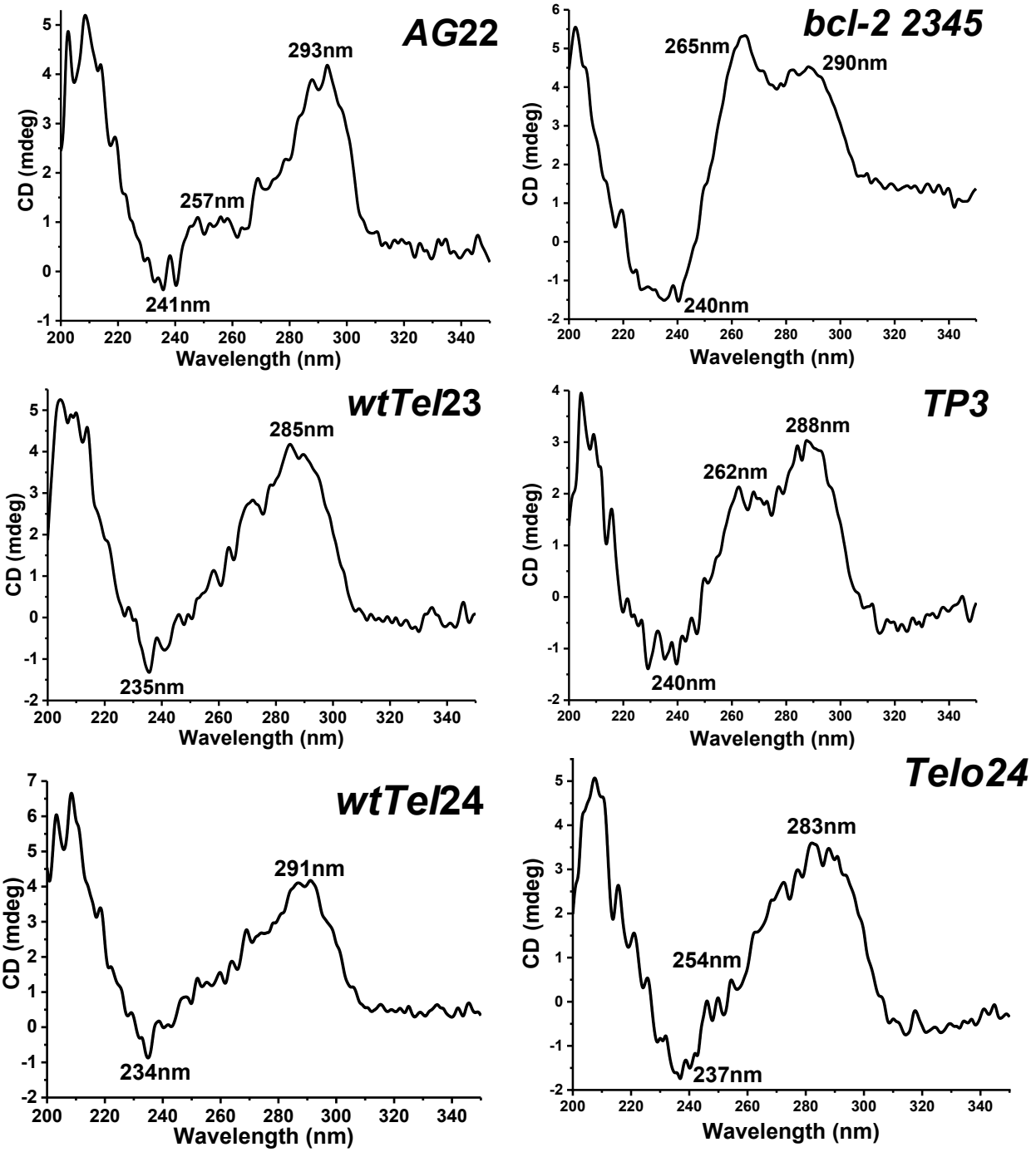
1.2 Circular Dichroism (CD) of G4 Topology

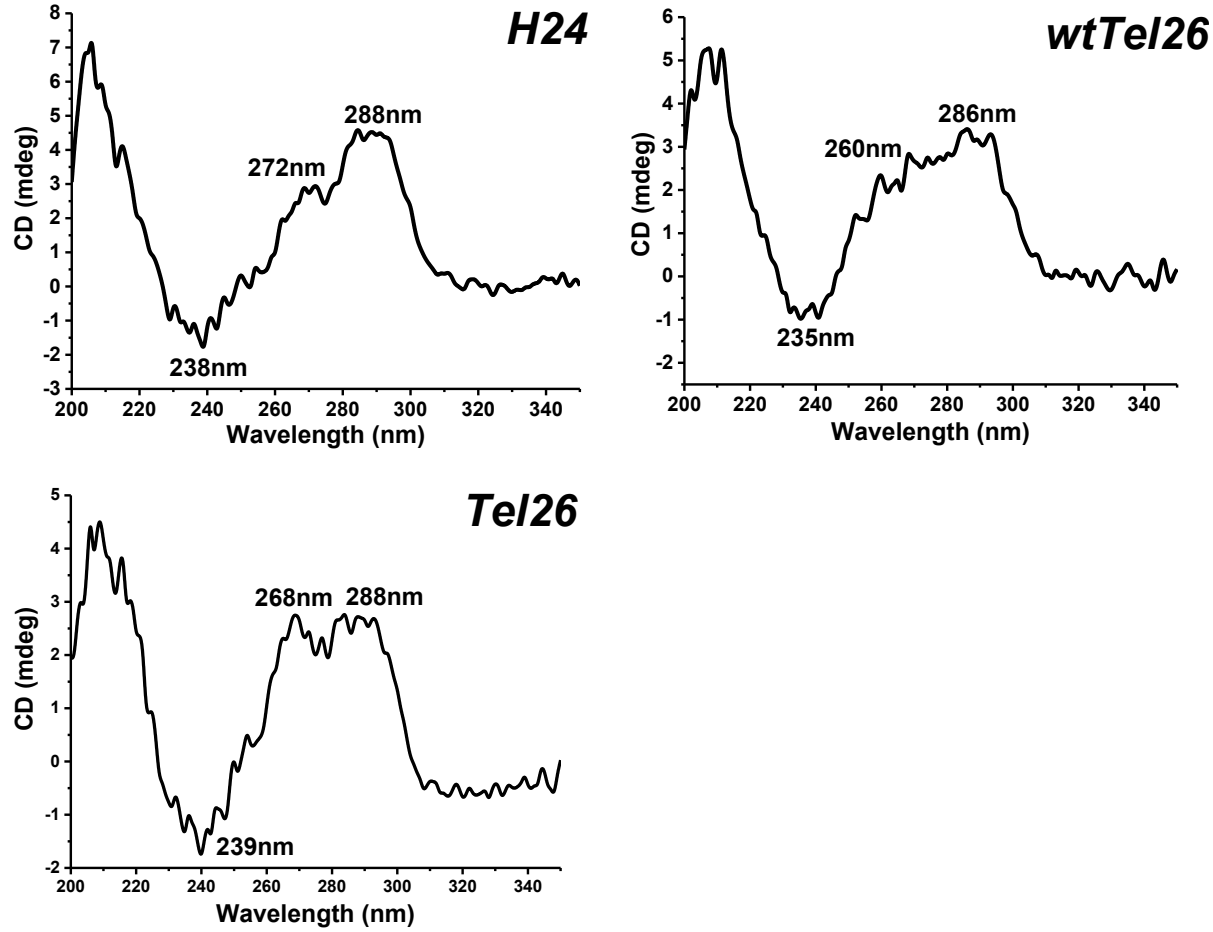
Parallel G4 sequences



Supplementary Fig. 1. CD spectra with baseline correction of 10 μM parallel G4 DNA in K^+ buffer (10mM K_2HPO_4/KH_2PO_4, 1mM EDTA, pH 7.4). DNA was denatured at 95°C for 5 min, then re-annealed at 4°C overnight before experiment.

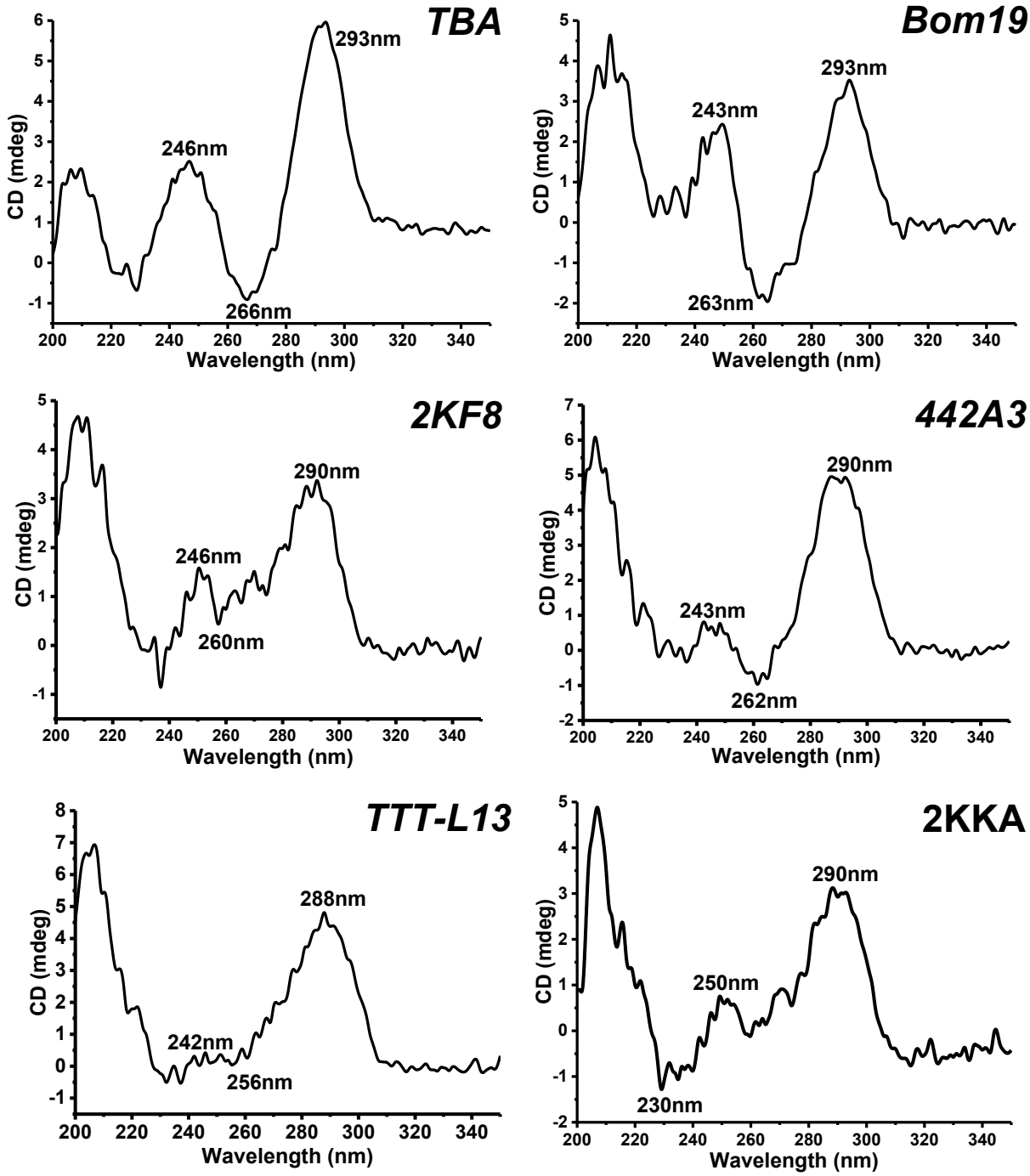
Hybrid G4 sequences

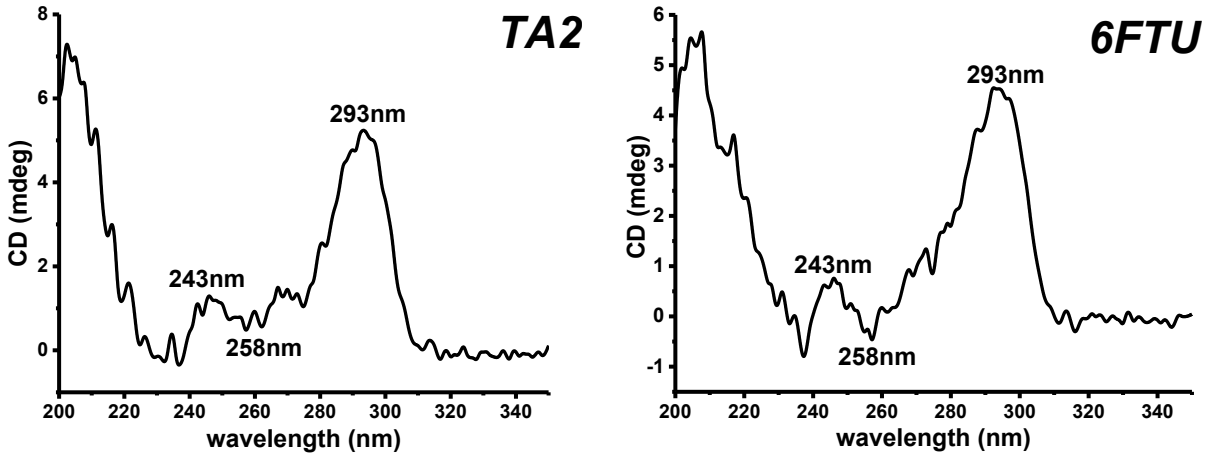




Supplementary Fig. 2. CD spectra with baseline correction of 10 μM hybrid G4 DNA in K^+ buffer (10mM K_2HPO_4/KH_2PO_4, 1mM EDTA, pH 7.4). DNA was denatured at 95°C for 5 min, then re-annealed at 4°C overnight before experiment.

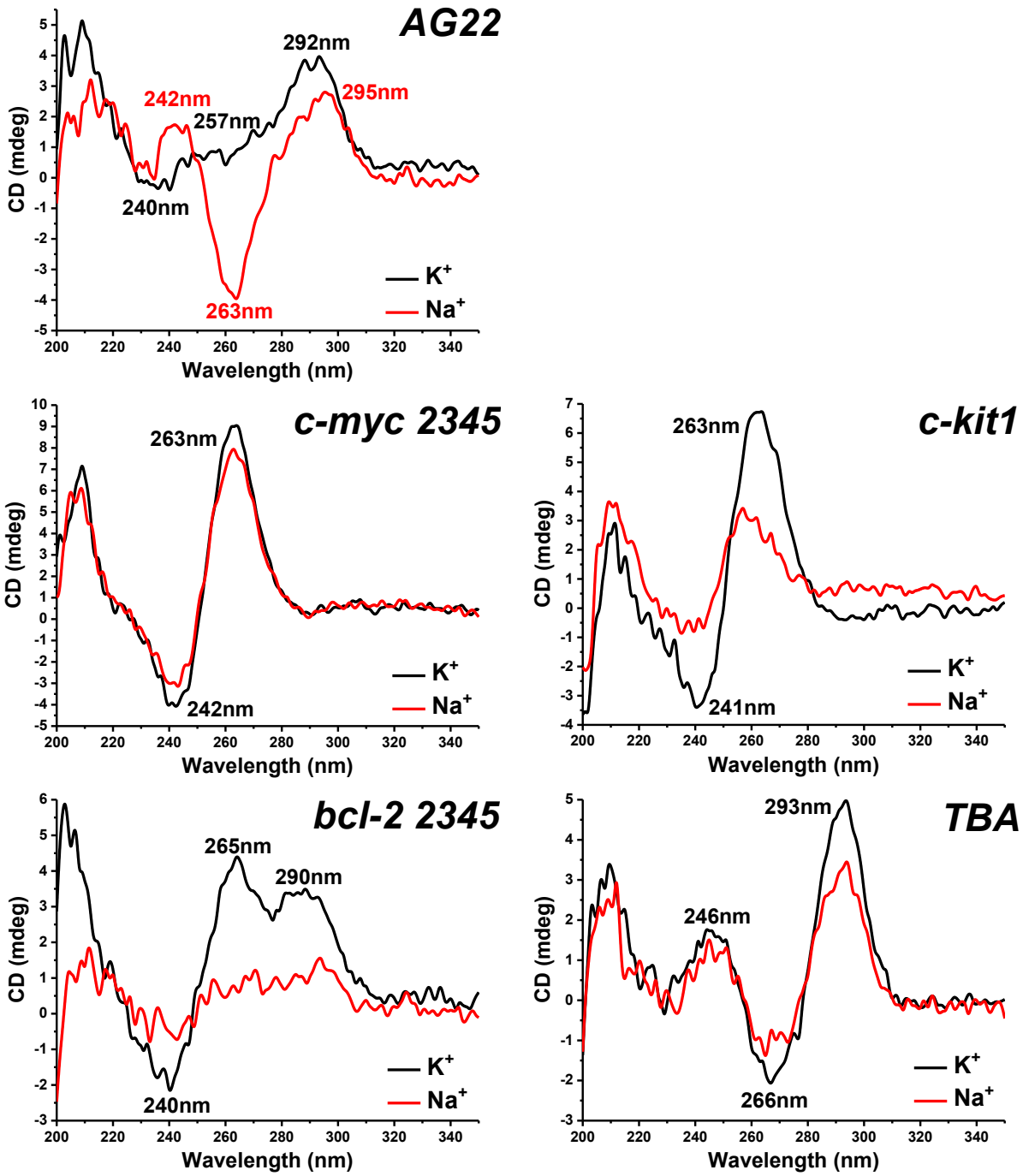
Antiparallel G4 sequences





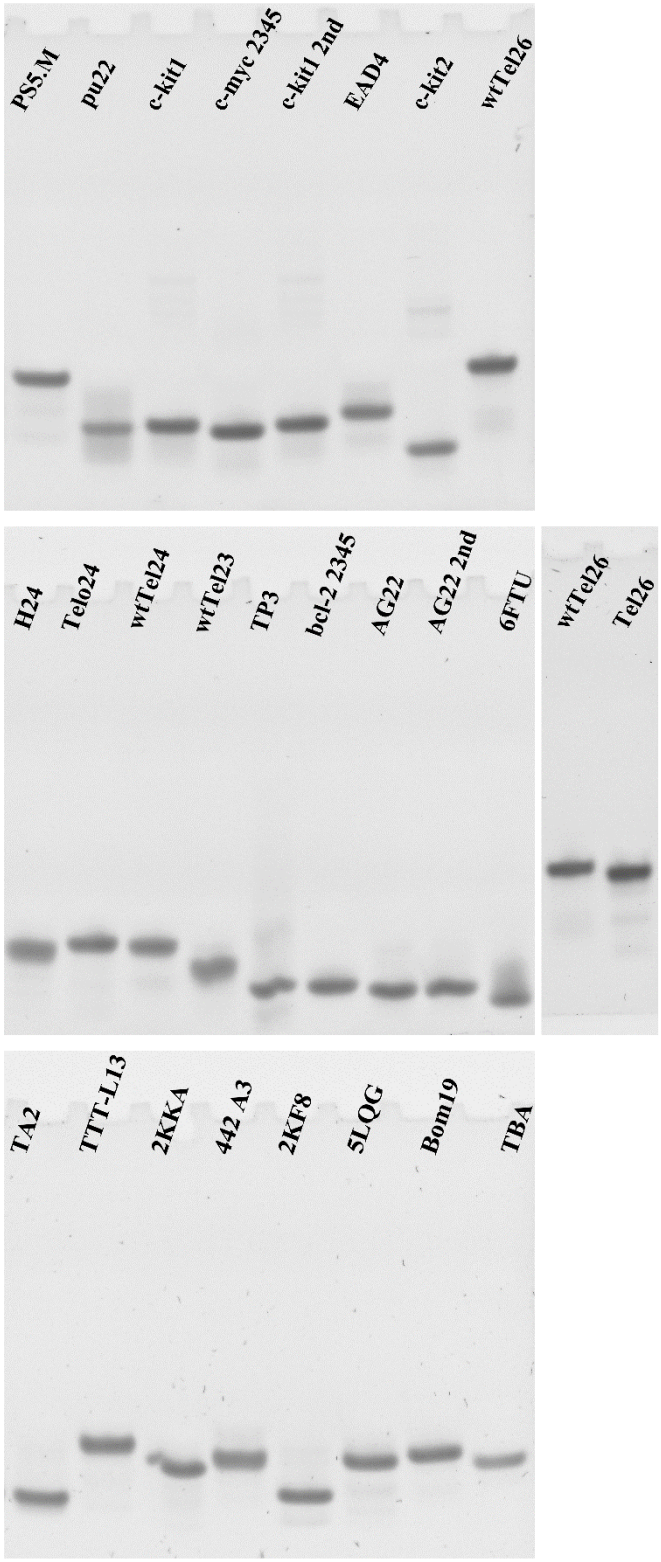
Supplementary Fig. 3. CD spectra with baseline correction of 10 μM antiparallel G4 DNA in K^+ buffer (10mM K_2HPO_4/KH_2PO_4, 1mM EDTA, pH 7.4). DNA was denatured at 95°C for 5 min, then re-annealed at 4°C overnight before experiment.

1.3 CD Analysis of AG22 Topology in Na^+/K^+ Buffer



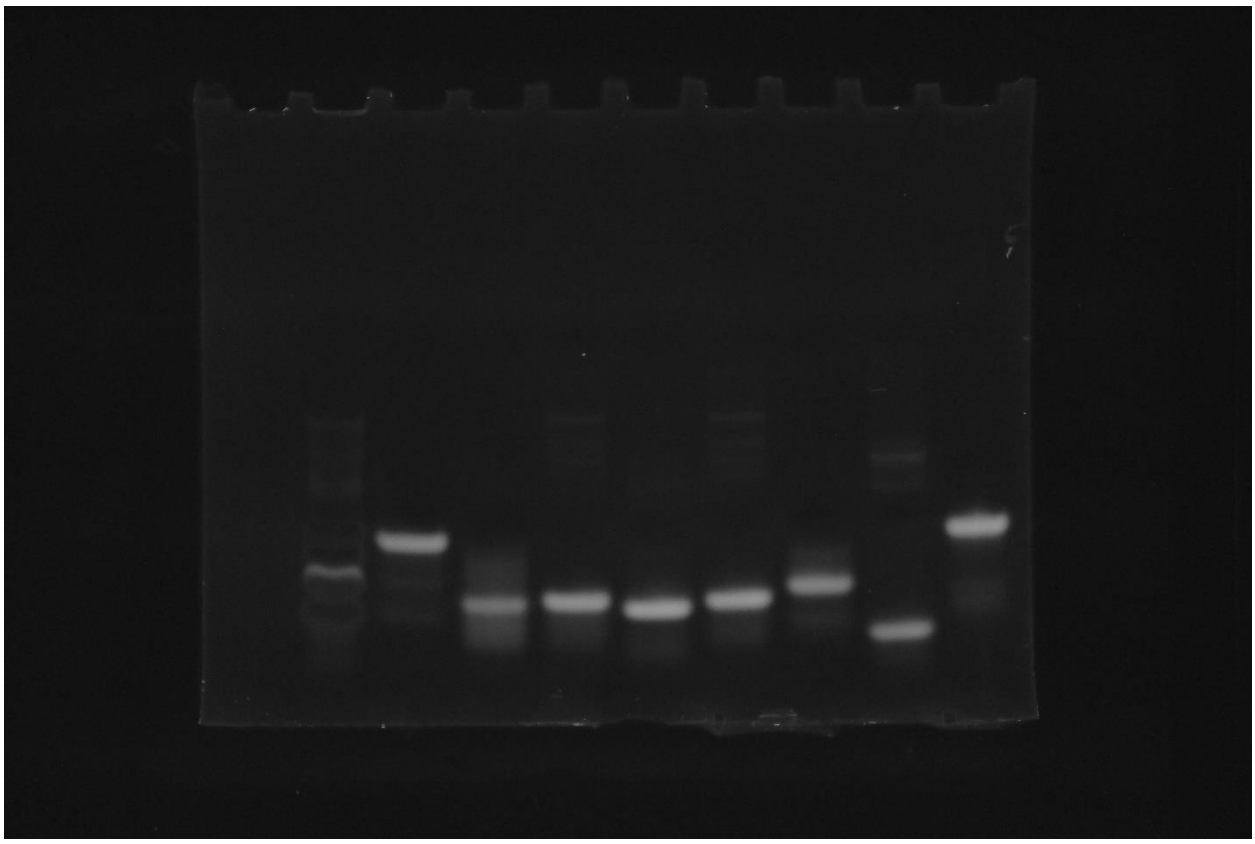
Supplementary Fig. 4. CD spectra with baseline correction of 10 μM *c-myc 2345*, *c-kit1*, *bcl-2 2345*, *TBA* and *AG22* in K^+ (10 mM K_2HPO_4/KH_2PO_4) (black line) or Na^+ buffer (10 mM Na_2HPO_4/NaH_2PO_4), both with 1 mM EDTA, pH 7.4. DNA was denatured at 95°C for 5 min, then re-annealed at 4°C overnight before experiment.

1.4 Gel Electrophoresis of G4 Structure

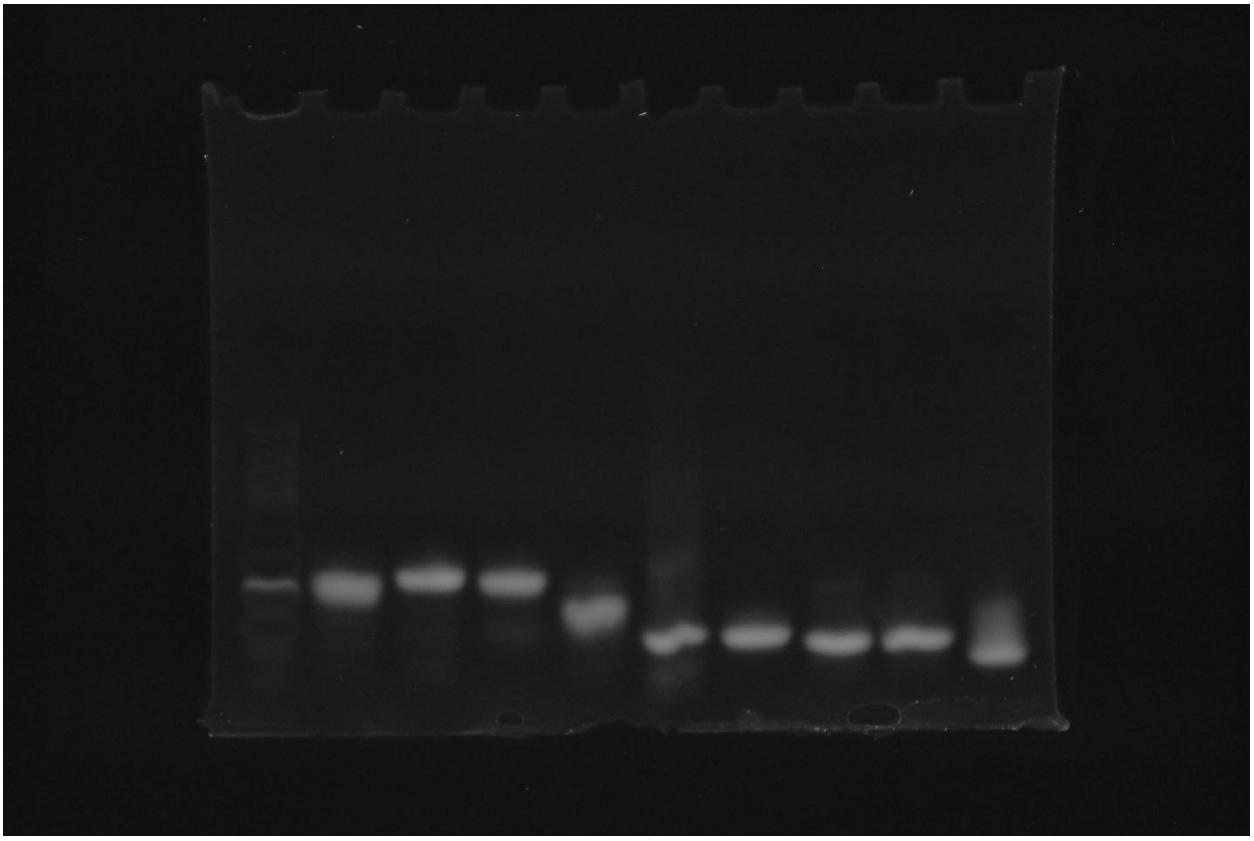


Supplementary Fig. 5. The gradient native polyacrylamide gel electrophoresis (PAGE) gel (4%-20%) results of DNA G-quadruplex sequences.

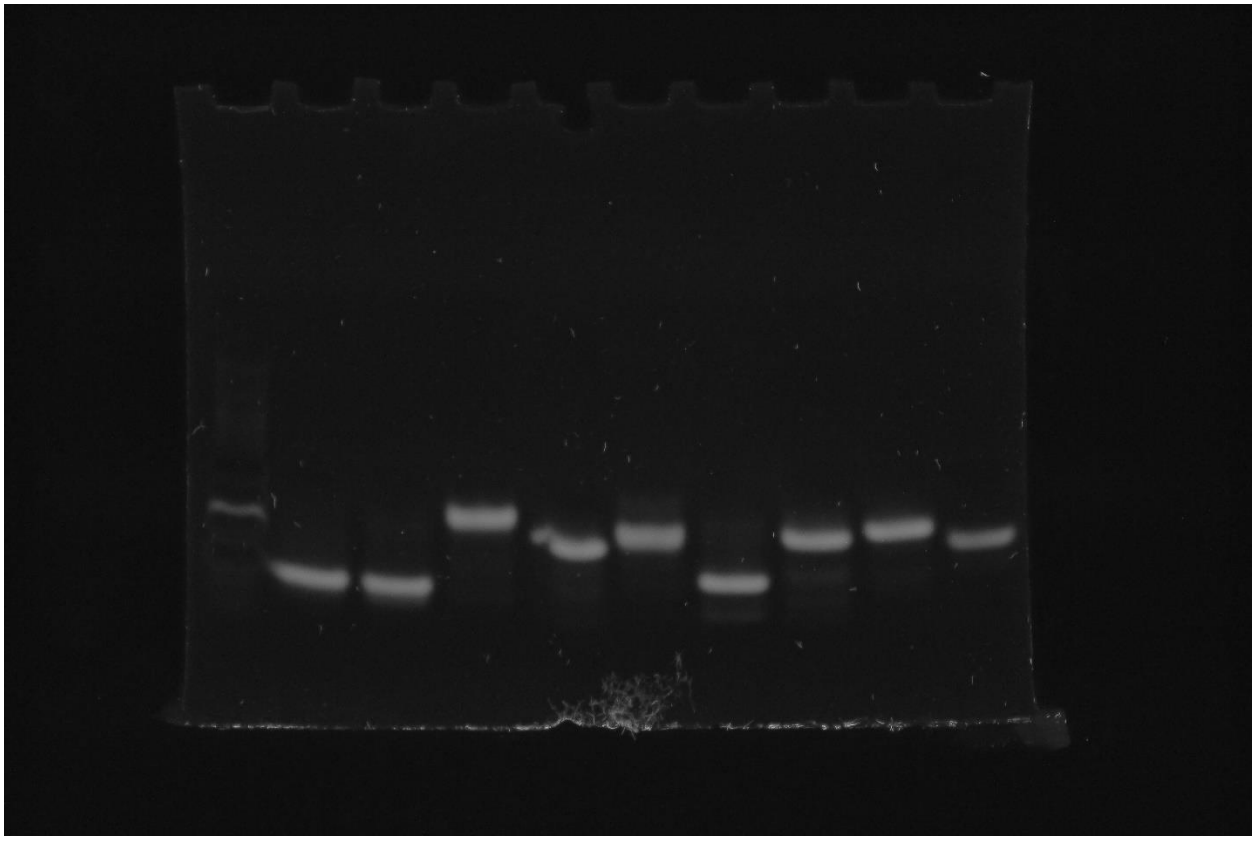
a



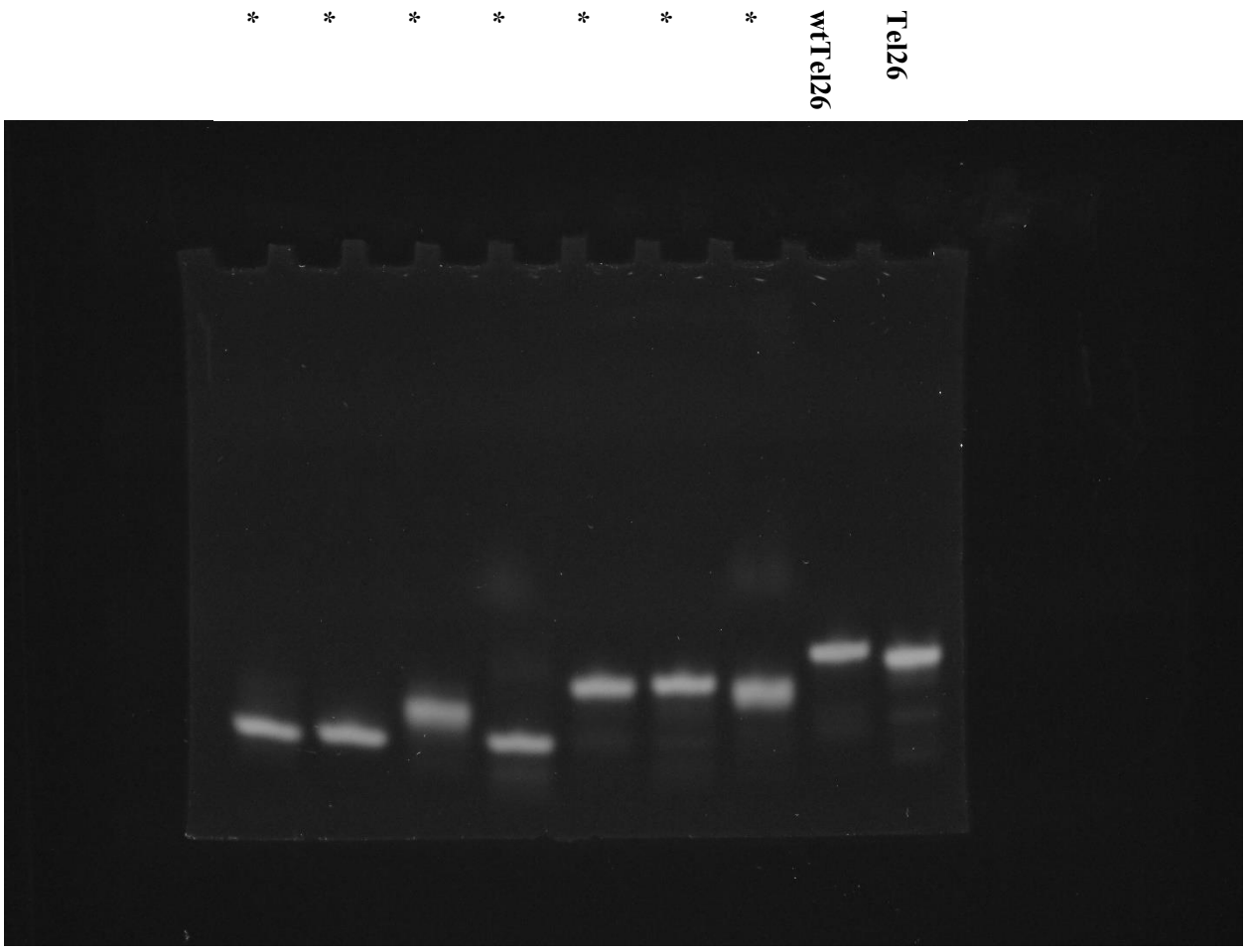
b



c



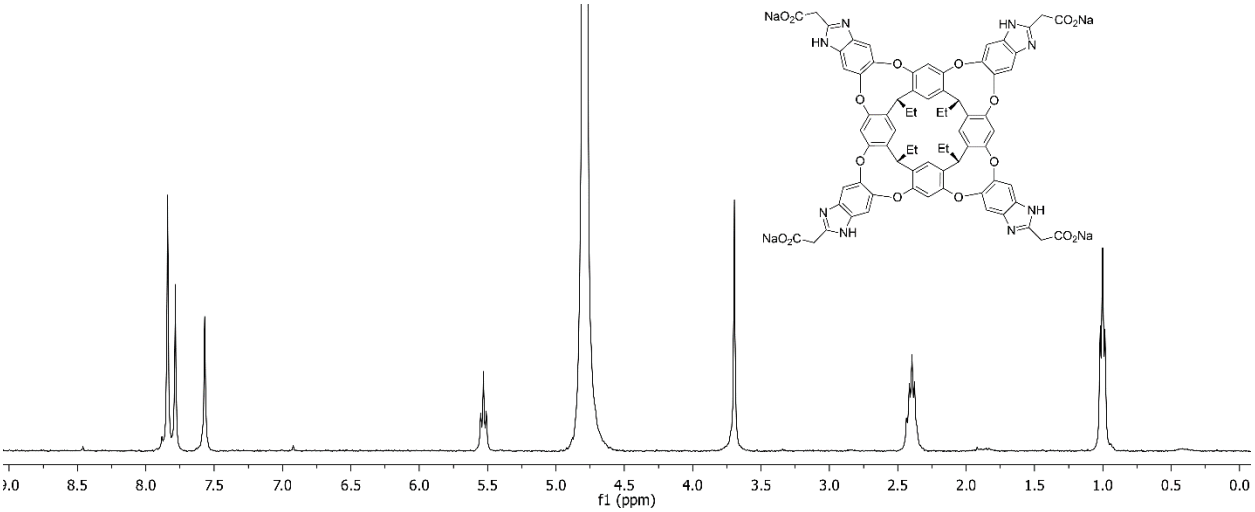
d



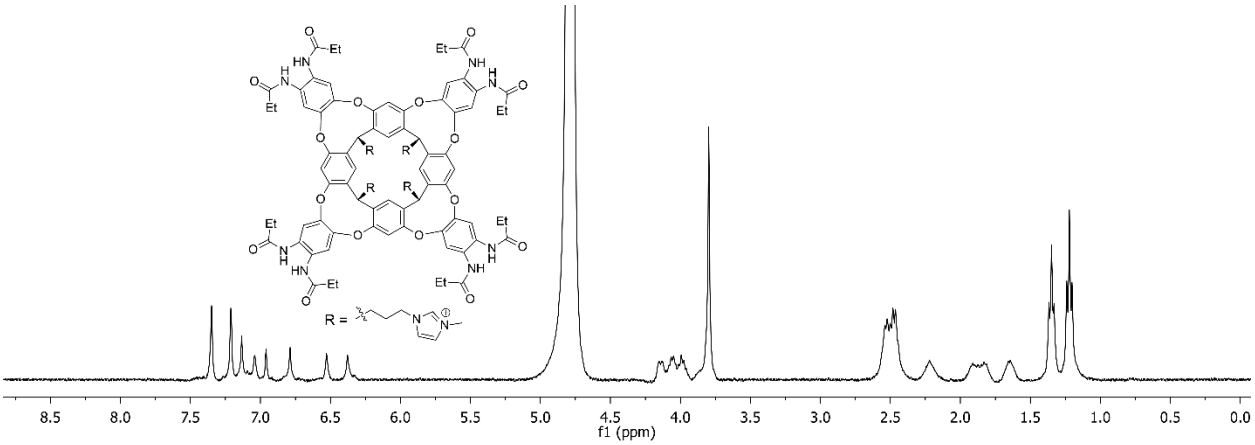
Supplementary Fig. 6. The unprocessed gel: the gradient native polyacrylamide gel electrophoresis (PAGE) gel (4%-20%) results of DNA G-quadruplex sequences. The DNA marked with a * are not used in this work.

2. NMR Spectra of Components Used

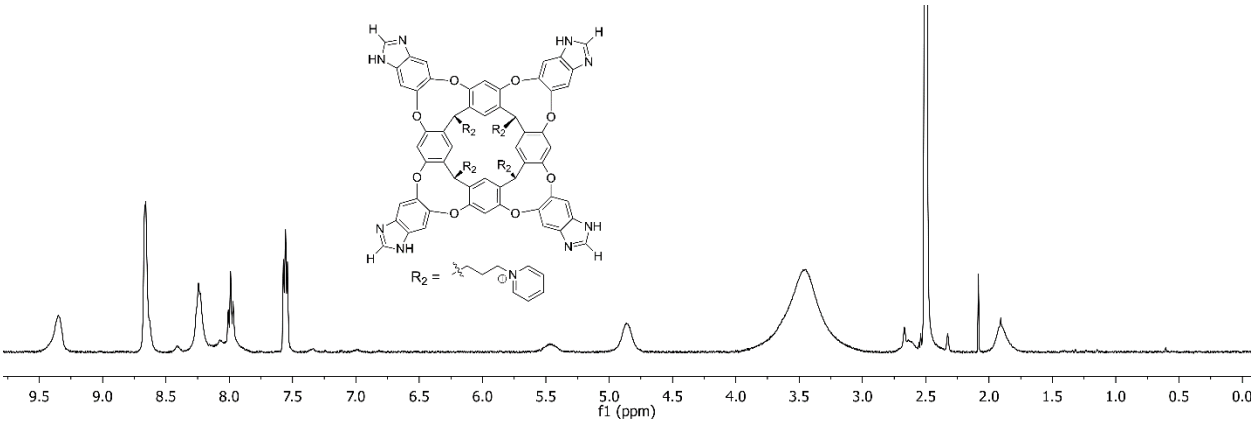
All hosts and guests are known and their spectra have been previously published, see methods section for the references. ^{1}H NMR spectra of the specific samples used are published here to illustrate their purity and ensure reproducibility. Some spectra were taken in DMSO-d_6 to maximize signal:noise ratio.



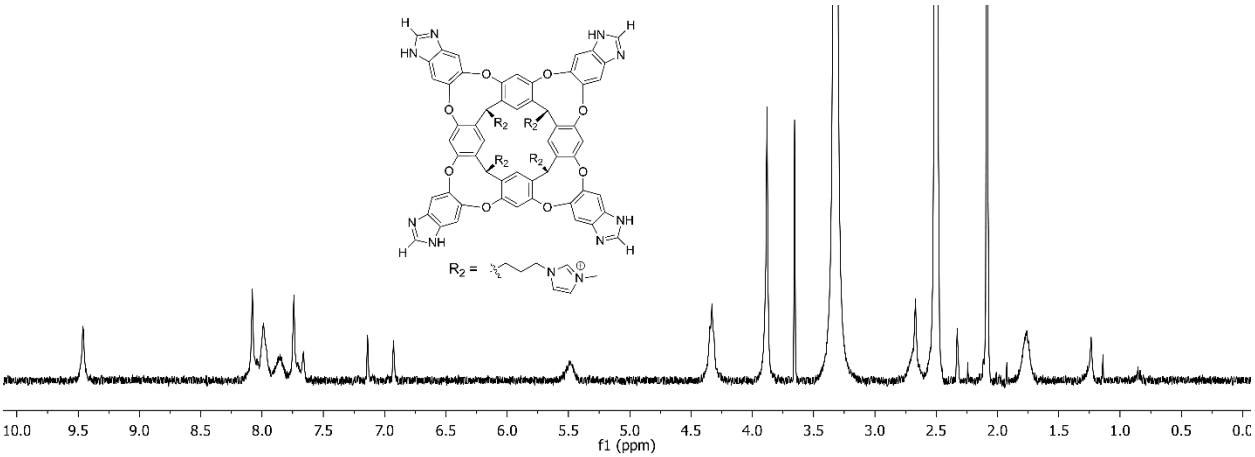
Supplementary Fig. 7. ^{1}H NMR spectrum of cavitand **1** (D_2O, 400 MHz, 298K).



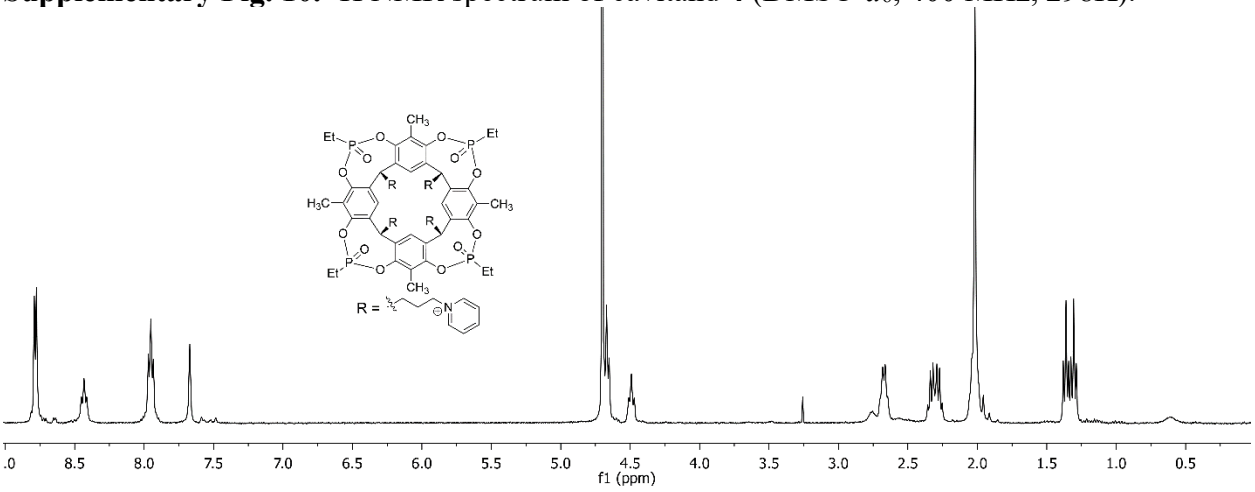
Supplementary Fig. 8. ^{1}H NMR spectrum of cavitand **2** (D_2O, 400 MHz, 298K).



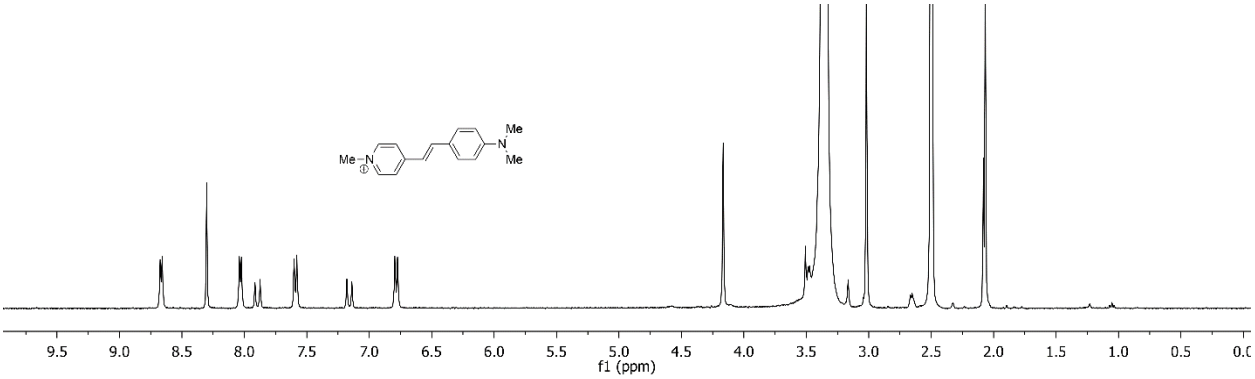
Supplementary Fig. 9. ^{1}H NMR spectrum of cavitand **3** (DMSO-d_6, 400 MHz, 298K).



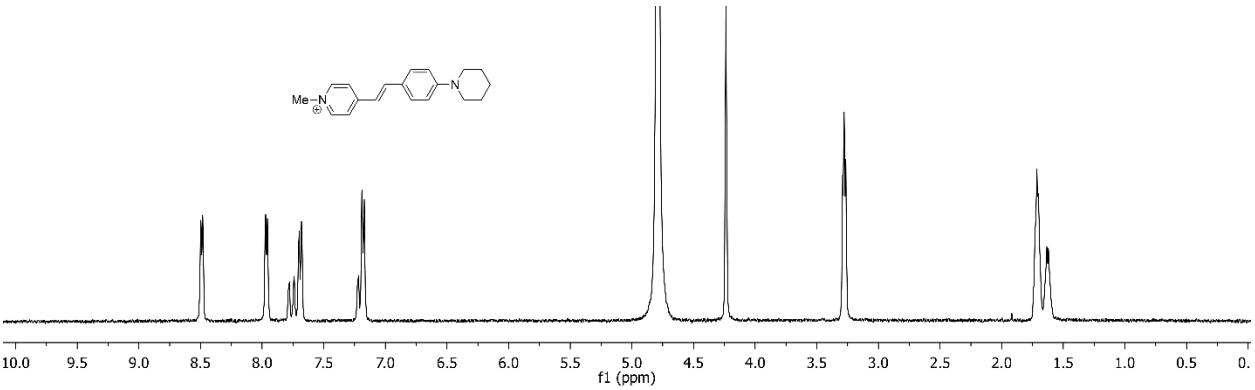
Supplementary Fig. 10. ^{1}H NMR spectrum of cavitand **4** (DMSO-d_6, 400 MHz, 298K).



Supplementary Fig. 11. ^{1}H NMR spectrum of cavitand **5** (D_2O, 400 MHz, 298K).



Supplementary Fig. 12. ^{1}H NMR spectrum of **DSMI** dye (DMSO-d_6, 400 MHz, 298K).

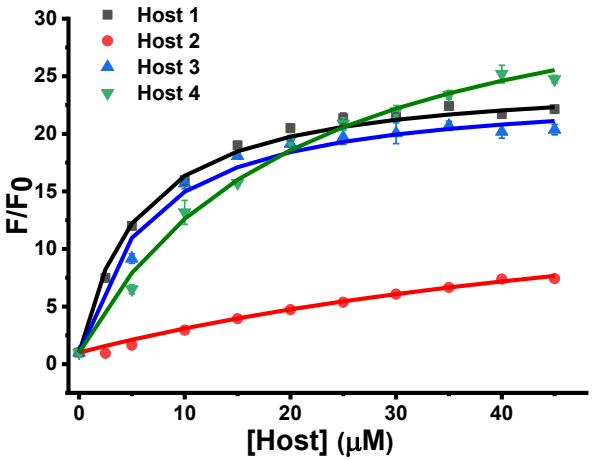


Supplementary Fig. 13. ^{1}H NMR spectrum of **PSMI** dye (D_2O, 400 MHz, 298K).

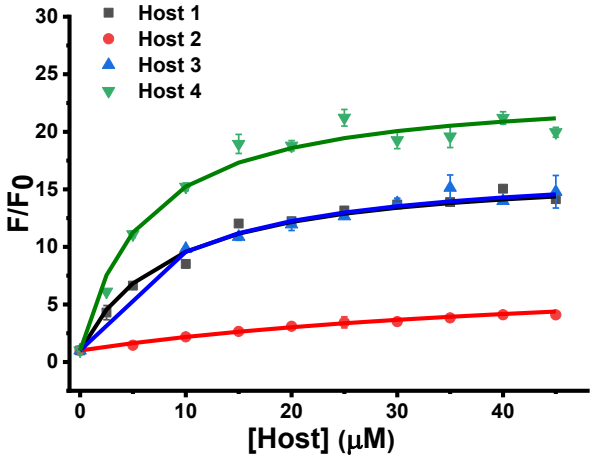
3. Binding Affinity Measurements

3.1 Fluorescence Titrations of Dyes and Hosts 1-5

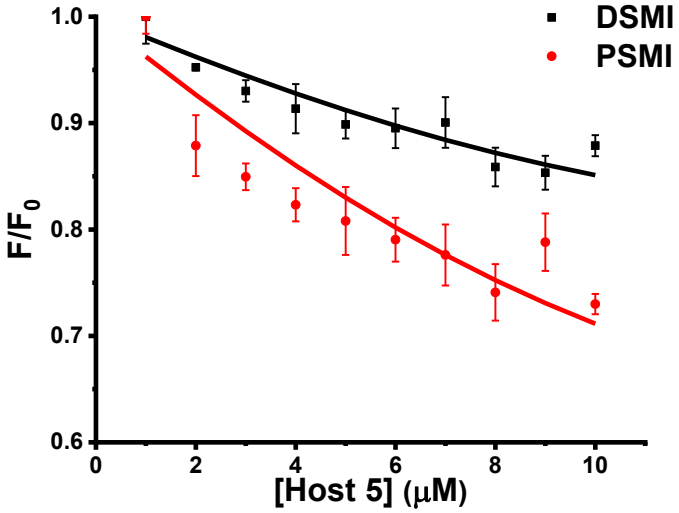
a) **DSMI + Host 1-4**



b) **PSMI + Host 1-4**

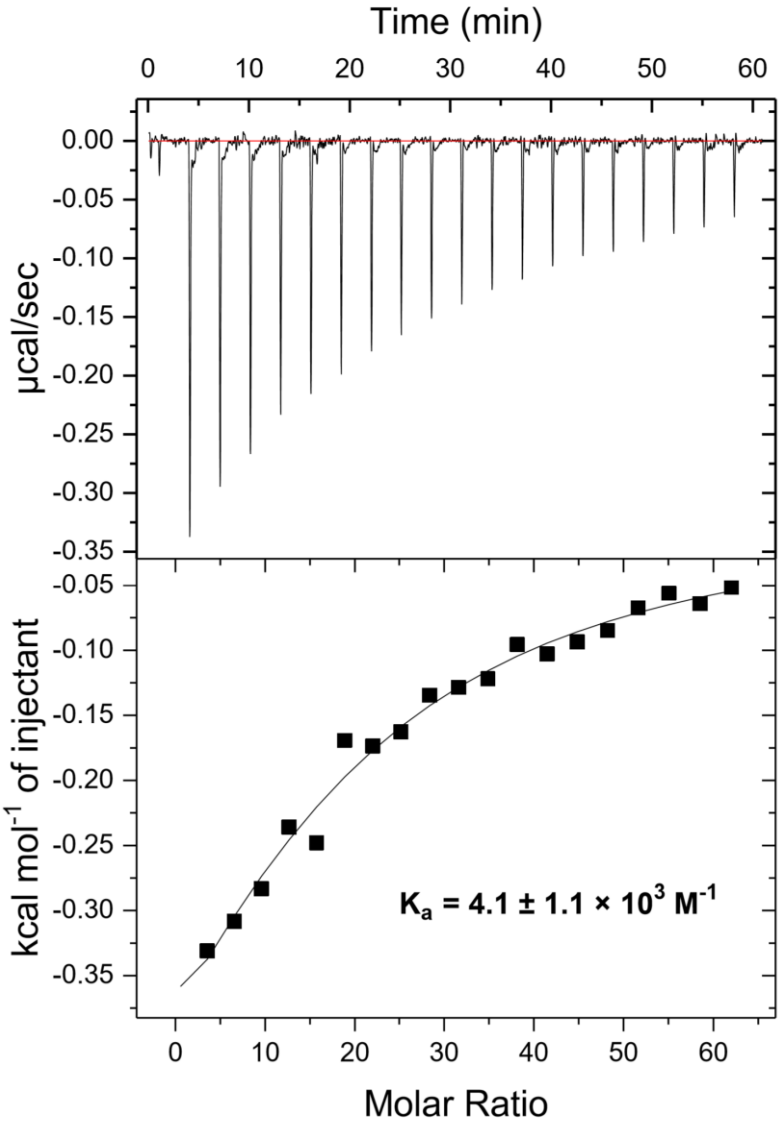


c) **DSMI/PSMI + Host 5**

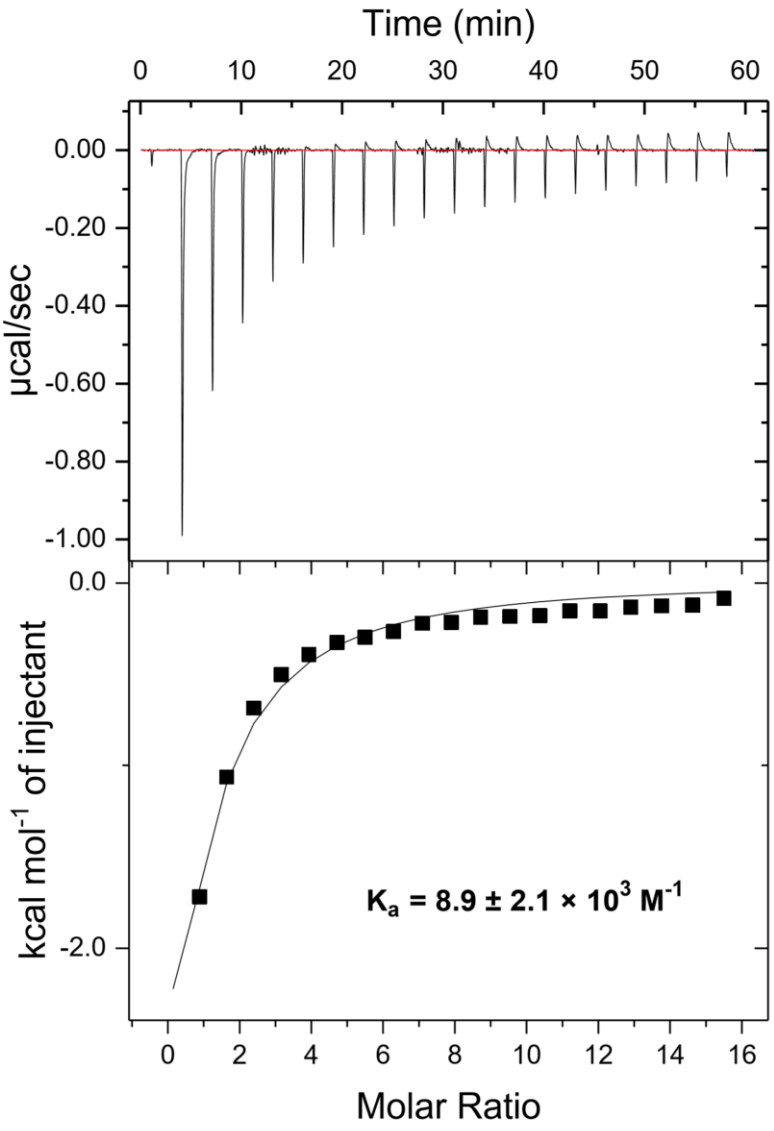


Supplementary Fig. 14. Affinity measurement of Dye: **DSMI**/**PSMI** with Host **1-5** via fluorescence. a) **DSMI** + Host **1-4**; b) **PSMI** + Host **1-4**; c) **DSMI**/**PSMI** + Host **5**. For Host **1-4**, [Dye] = 0.625μM; for Host **5**, [Dye] = 10μM. Buffer: 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm.

3.2 Isothermal Calorimetry Analysis of Dye•*c-myc 2345* Binding

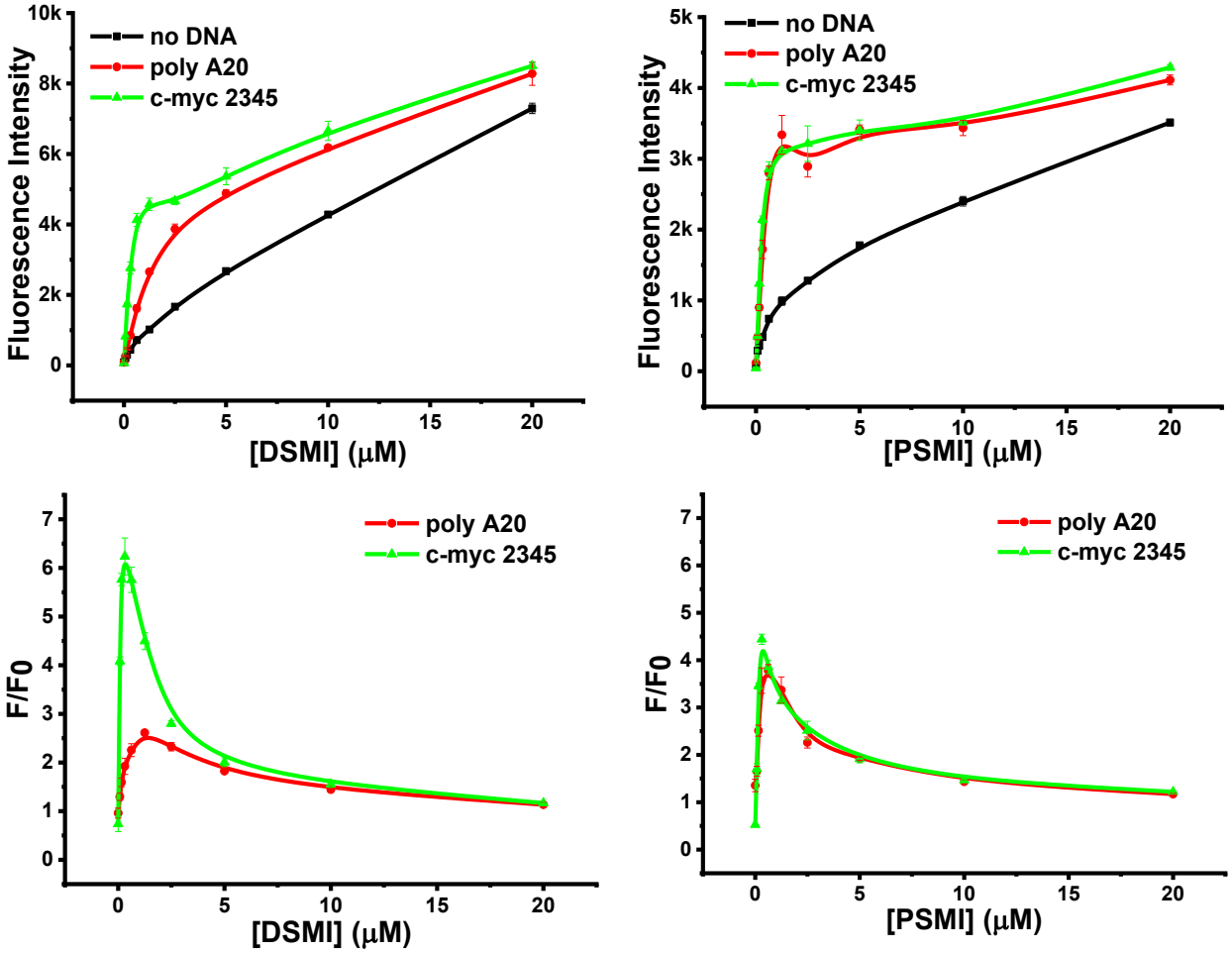


Supplementary Fig. 15. ITC titrations of increasing amounts of **DSMI** with 10µM *c-myc 2345*, measured at 25 °C. The 10µM *c-myc 2345* was placed in the cell and the 3mM **DSMI** in the syringe. Both solutions were diluted with buffer 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4. Top trace: raw data for the ITC titration. Bottom trace: binding isotherm of the integrated calorimetric titration data. The heat of dilution, measured by the injection of titrant into the buffer solution, was subtracted for each titration to obtain the net reaction heat value. *c-myc 2345* was denatured at 95°C 5 min, then re-annealing at 4°C overnight before experiment.



Supplementary Fig. 16. ITC titrations of increasing amounts of **PSMI** with 40µM *c-myc 2345*, measured at 25 °C. The 40µM *c-myc 2345* was placed in the cell and the 3mM **PSMI** in the syringe. Both solutions were diluted with 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4. Top trace: raw data for the ITC titration. Bottom trace: binding isotherm of the integrated calorimetric titration data. The heat of dilution, measured by the injection of titrant into the buffer solution, was subtracted for each titration to obtain the net reaction heat value. *c-myc 2345* was denatured at 95°C 5 min, then re-annealing at 4°C overnight before experiment.

3.3 Fluorescence titration of Dye-DNA



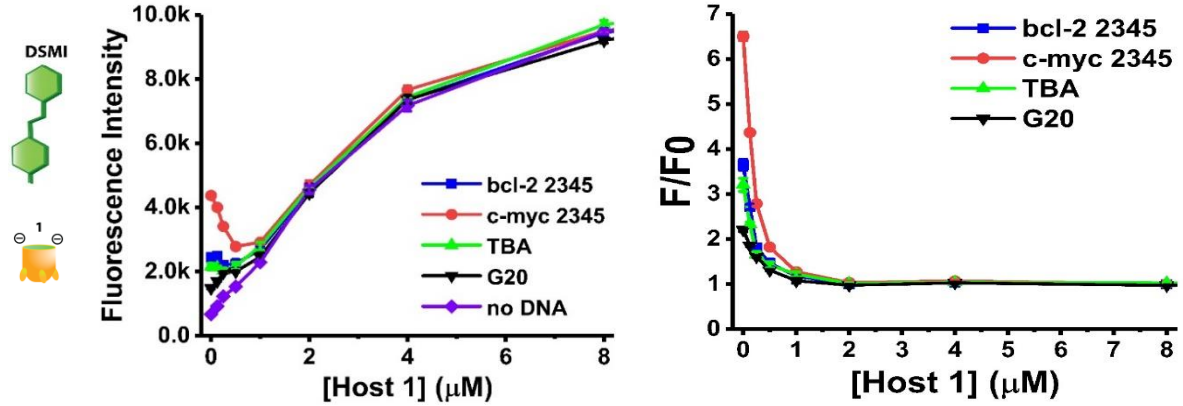
Supplementary Fig. 17. Fluorescence response curves of poly A20 or *c-myc 2345* with increasing concentration (0-20 μM) of **DSMI**/**PSMI**. Top: plots using the raw fluorescence counts; Bottom: plots using the fluorescence normalized against that of the dye (F_0 being the dye fluorescence in the absence of DNA). [**DSMI**/**PSMI**] = 0-20 μM, [DNA] = 0.1 μM, 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. DNA was denatured at 95°C 5 min, then re-annealing at 4°C overnight before experiment.

4. Fluorescence Response Curves

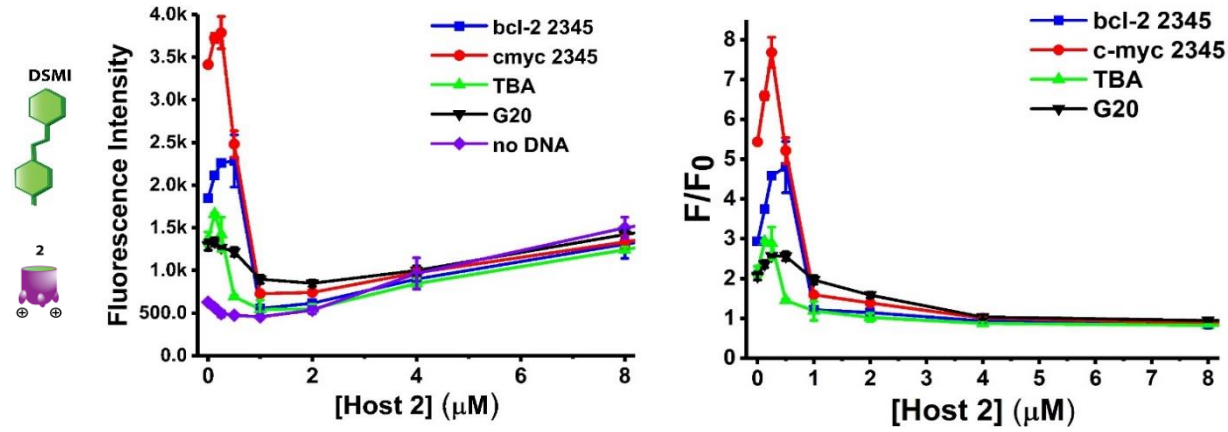
4.1 Host Addition to Dye•DNA Complexes

4.1.1 Host Addition to DSMI•DNA Complexes

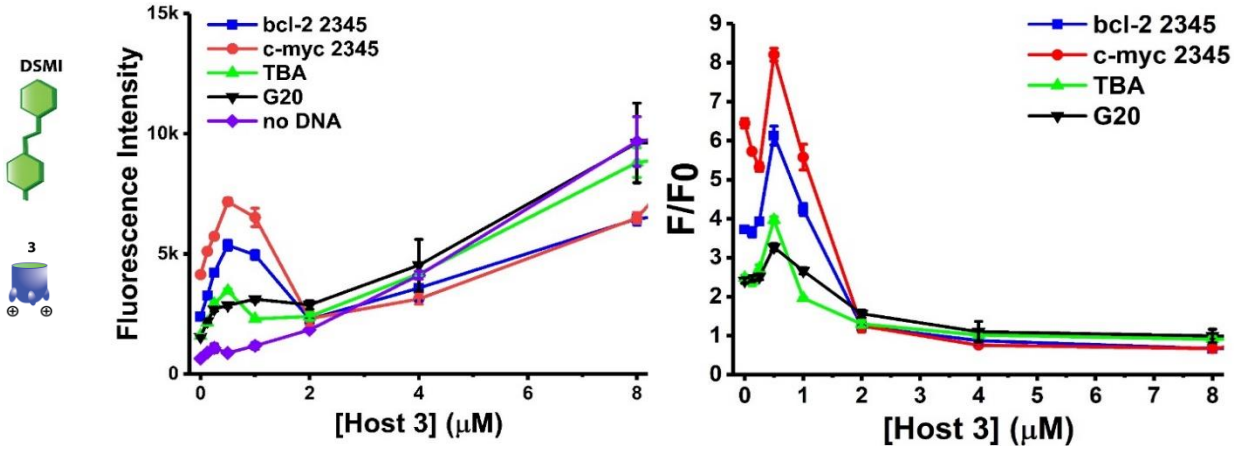
a)



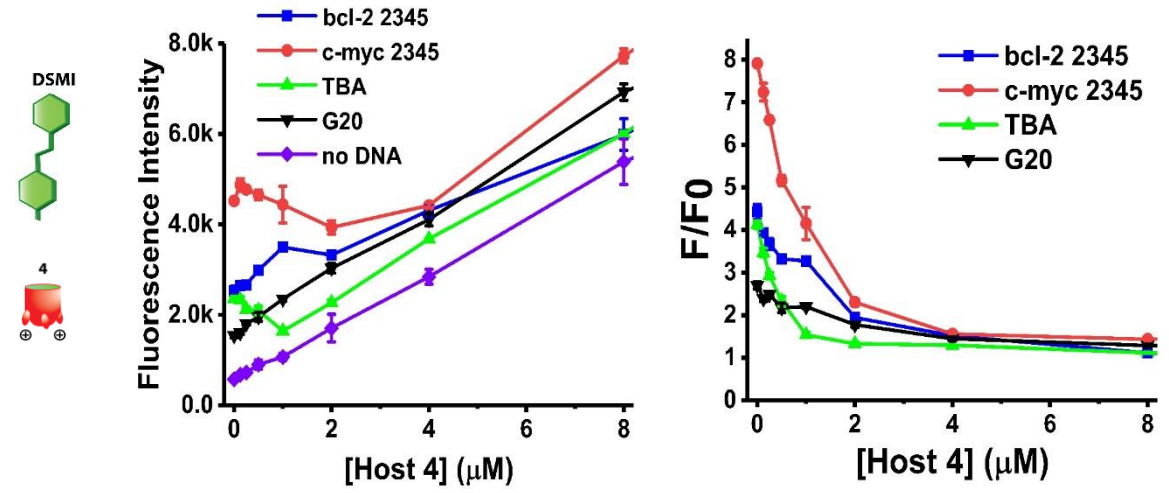
b)



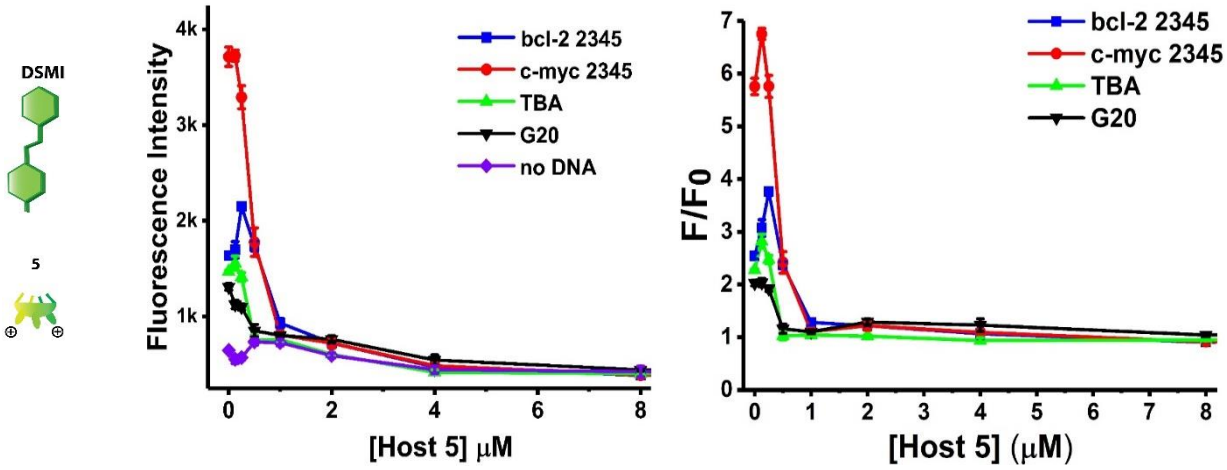
c)



d)



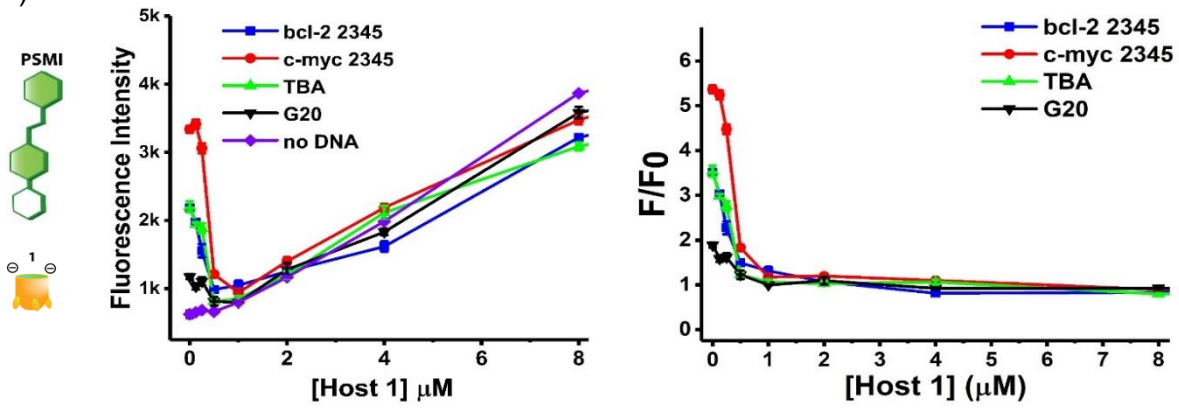
e)



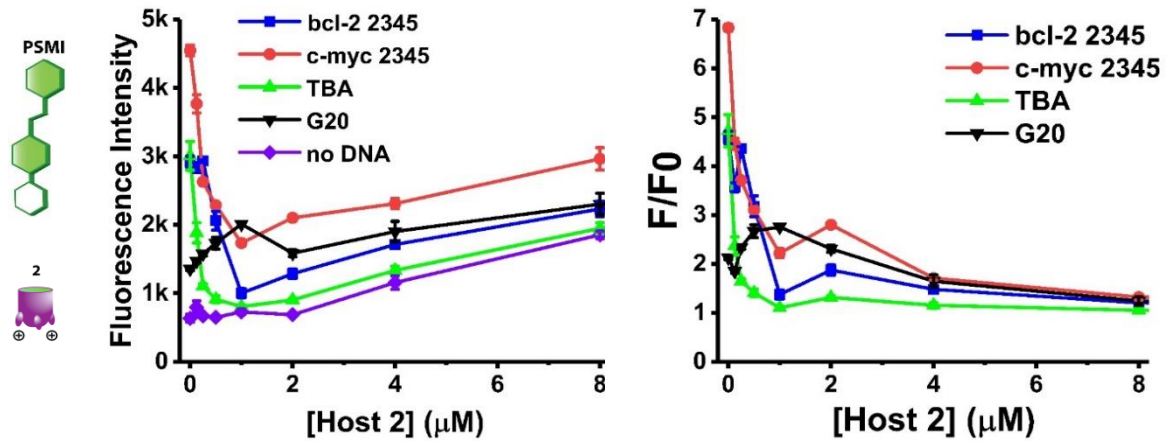
Supplementary Fig. 18. Fluorescence response curves of **DSMI**•DNA complexes upon titration of hosts 1-5. ***Left:*** the raw fluorescence counts (**DSMI** + Host **1-5**); ***Right:*** plots normalized to the response of cavitand-**DSMI** in the absence of DNA (F_0). [**DSMI**] = 0.625 μM, [DNA] = 0.1 μM, [Host] = 0-8 μM, 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. DNA was denatured at 95°C 5 min, then re-annealed at 4°C overnight before experiment.

4.1.2 Host addition to PSMI•DNA Complexes

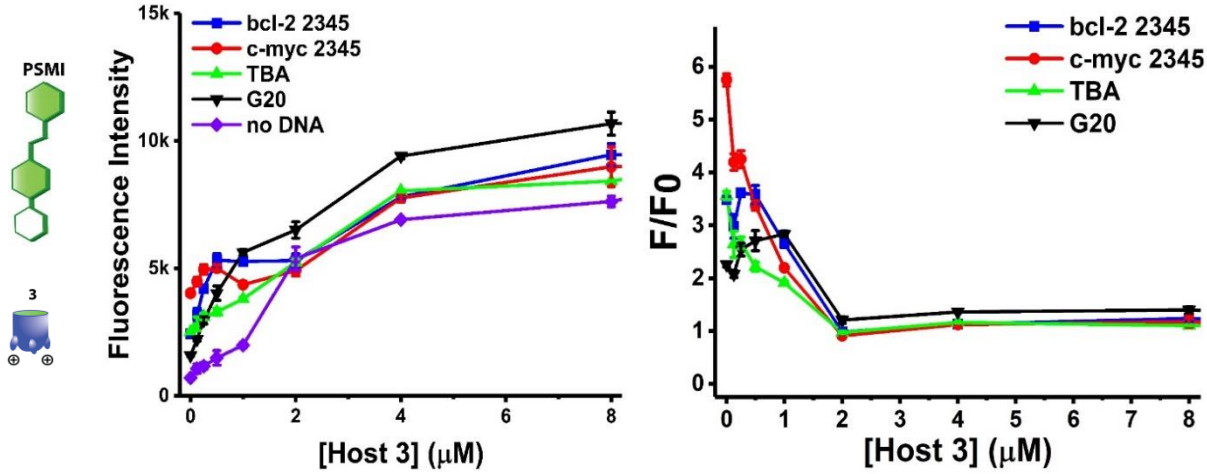
a)

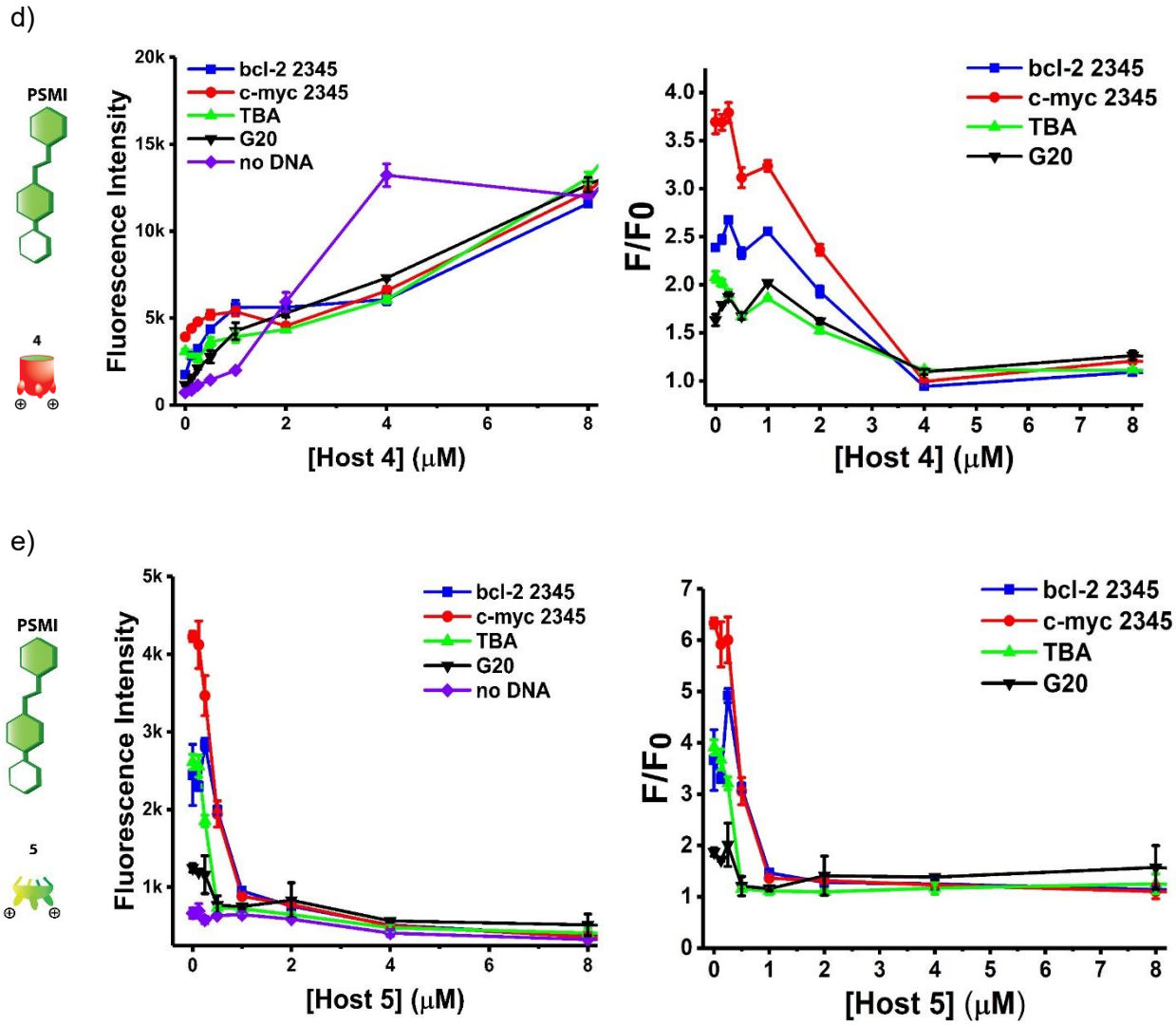


b)



c)



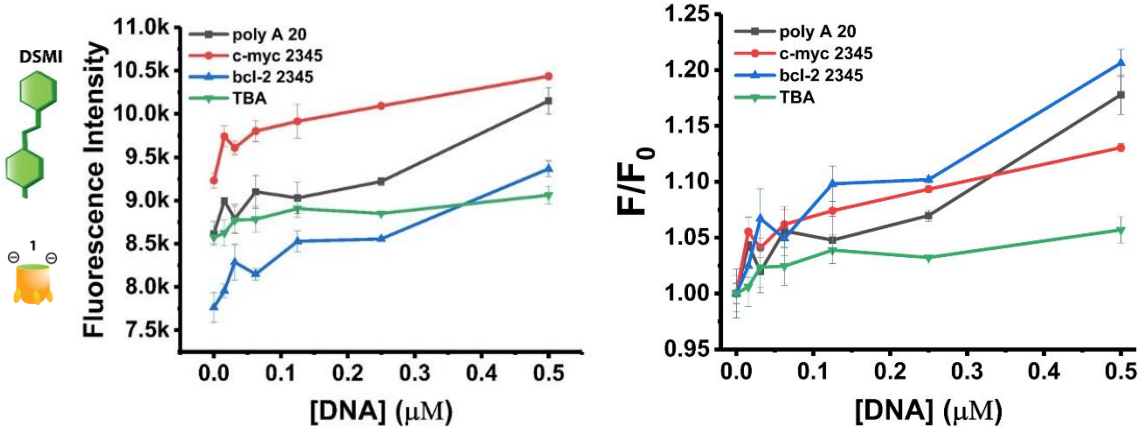


Supplementary Fig. 19. Fluorescence response curves of **PSMI**•DNA complexes upon titration of Hosts **1-5**. ***Left:*** the raw fluorescence counts (**PSMI** + Host **1-5**); ***right:*** plots normalized to the response of cavitand-**PSMI** (in the absence of DNA). [**PSMI**] = 0.625 μM, [DNA] = 0.1 μM, [Host] = 0-8 μM, 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. DNA was denatured at 95°C 5 min, then re-annealed at 4°C overnight before experiment.

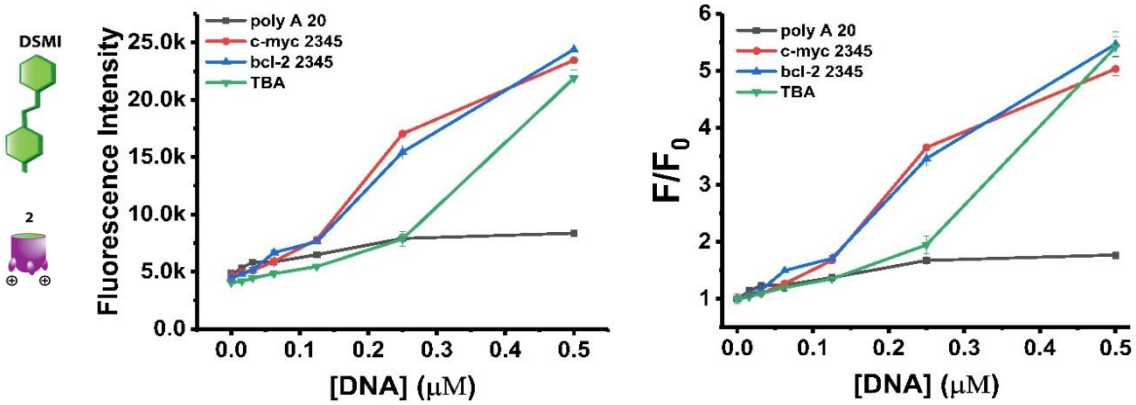
4.2 DNA Addition to Dye•Host Complexes

4.2.1 DNA Addition to DSMI•Host Complexes

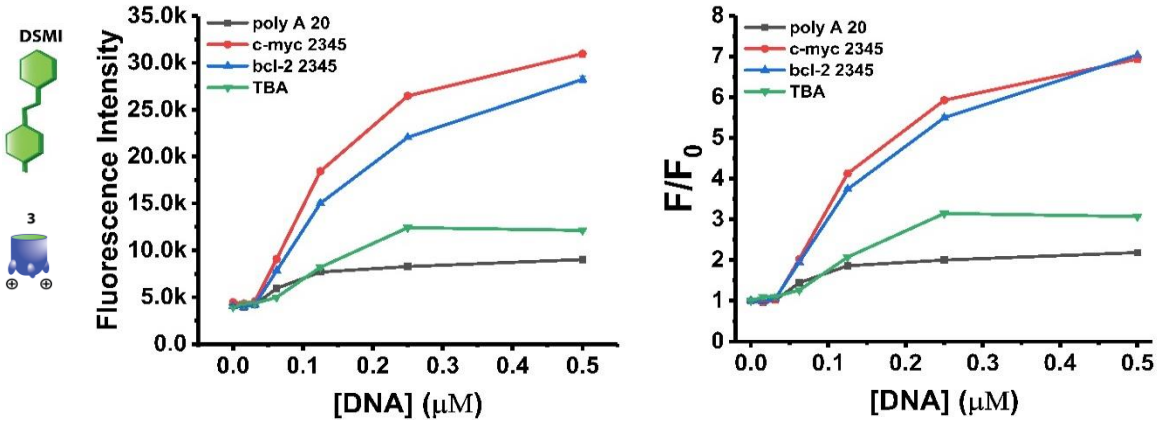
a)



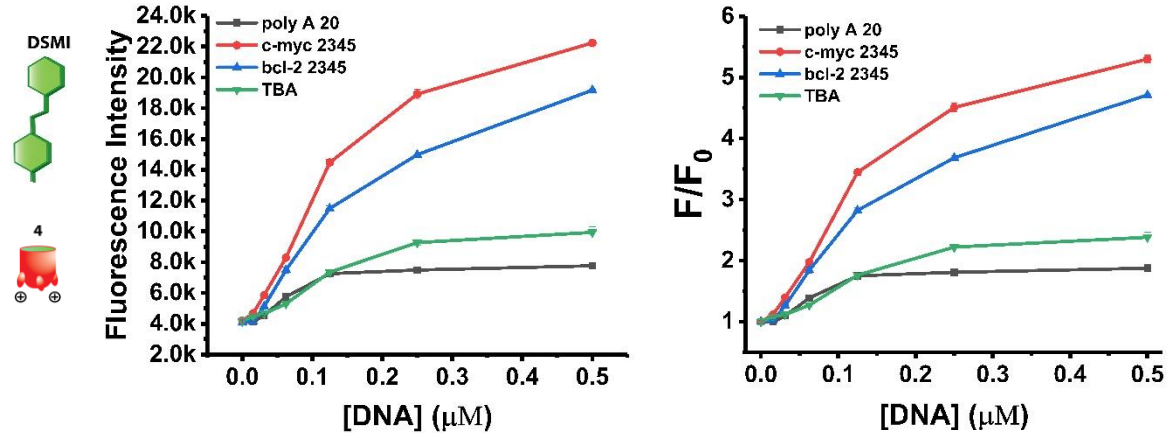
b)



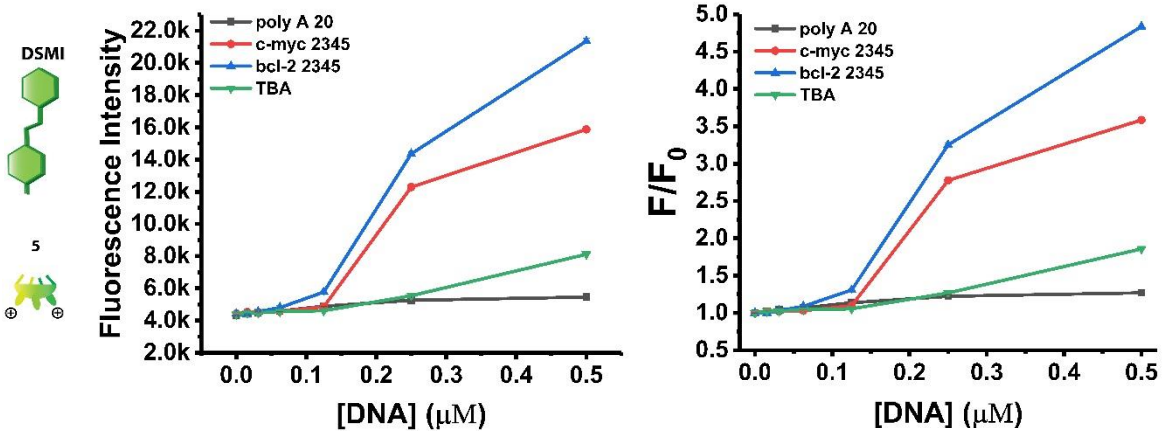
c)



d)



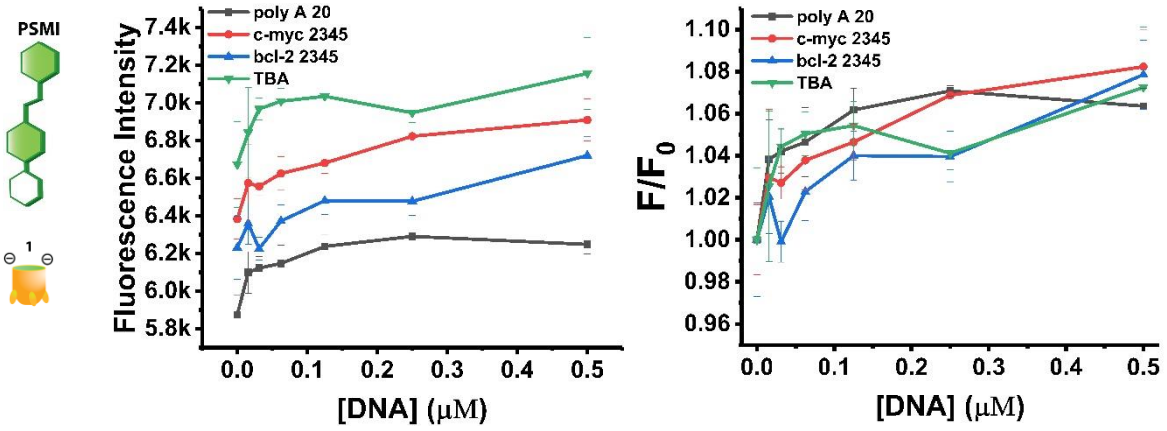
e)



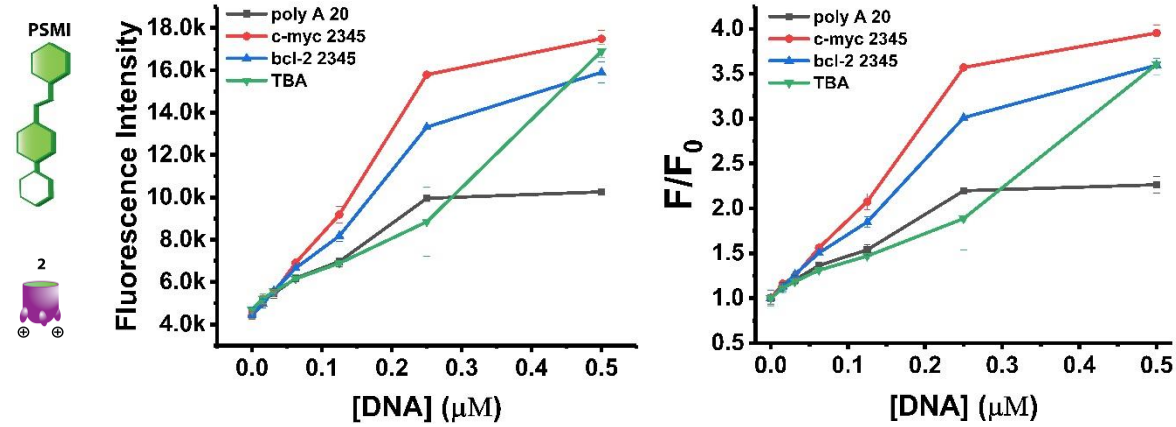
Supplementary Fig. 20. Fluorescence response curves of **DSMI•Host** complexes upon titration of DNA. ***Left:*** the raw fluorescence counts (**DSMI** + Host **1-5**); ***Right:*** plots normalized to the response of cavitand-**DSMI** in the absence of DNA (F_0). [**DSMI**] = 0.625 μM, [Host] = 1 μM, [DNA] = 0 - 0.5 μM, 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4. DNA was denatured at 95°C 5 min, then re-annealed at 4°C overnight before experiment. Note: This set of data were acquired in BioTek Synergy HT microplate reader with the filter set of Ex485/Em600nm.

4.2.2 DNA Addition to PSMI•Host Complexes

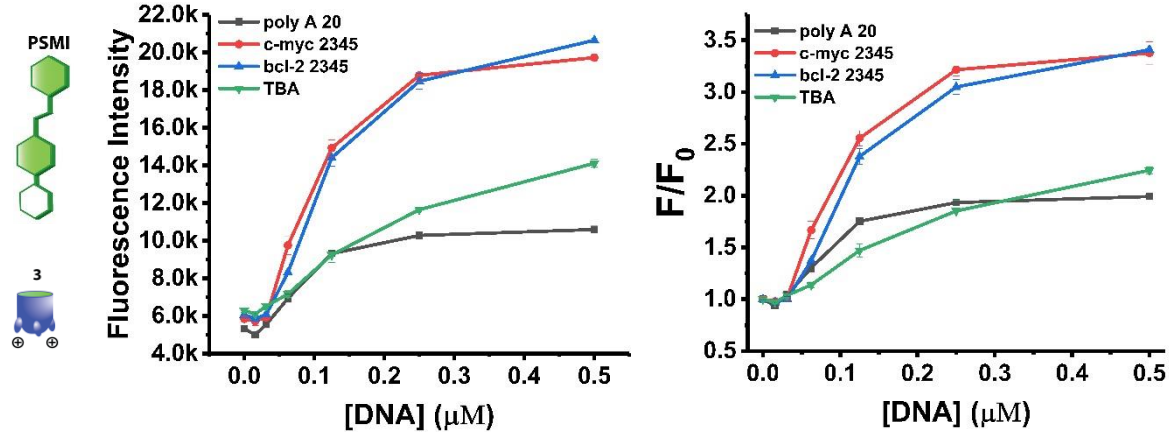
a)



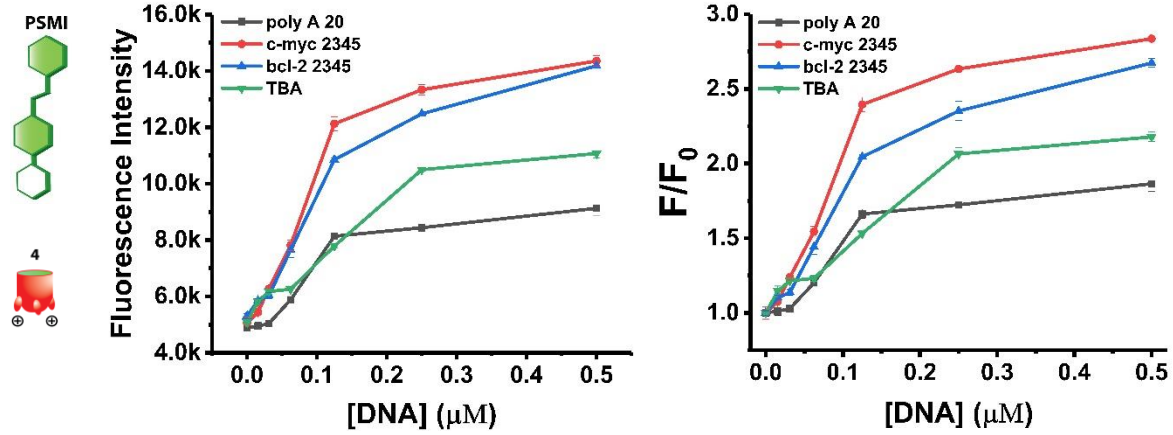
b)



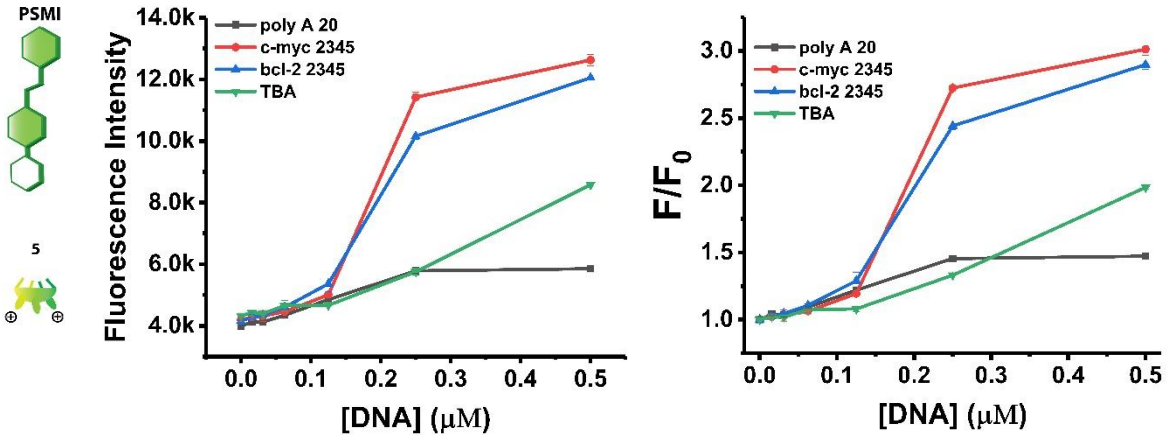
c)



d)



e)



Supplementary Fig. 21. Fluorescence response curves (a-e) of **PSMI•Host** complexes upon titration of DNA. ***Left:*** the raw fluorescence counts (**PSMI** + Host **1-5**); ***Right:*** plots normalized to the response of cavitand-**PSMI** in the absence of DNA (F_0). [**PSMI**] = 0.625 μM, [Host] = 1 μM, [DNA] = 0 - 0.5 μM, 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4. DNA was denatured at 95°C 5 min, then re-annealed at 4°C overnight before experiment. Note: This set of data were acquired in BioTek Synergy HT microplate reader with the filter set of Ex485/Em600nm.

5. Limit of Detection (LOD) for Each Host•Dye Complex

5.1 DNA Detection with DSMI

Supplementary Table 2 LOD calculated from the titration curves shown in Section 5.2.1 using the 3σ method.

LOD/DSMI	A20	*c-myc 2345*
Host 2	0.00779±0.00720μM	0.00616±0.00555μM
Host 3	0.02480±0.00070μM	0.02985±0.00186μM
Host 4	0.01925±0.00026μM	0.00983±0.00020μM
Host 5	0.01173±0.00005μM	0.00476±0.00005μM
No Host	0.01368±0.00042μM	0.00357±0.00002μM

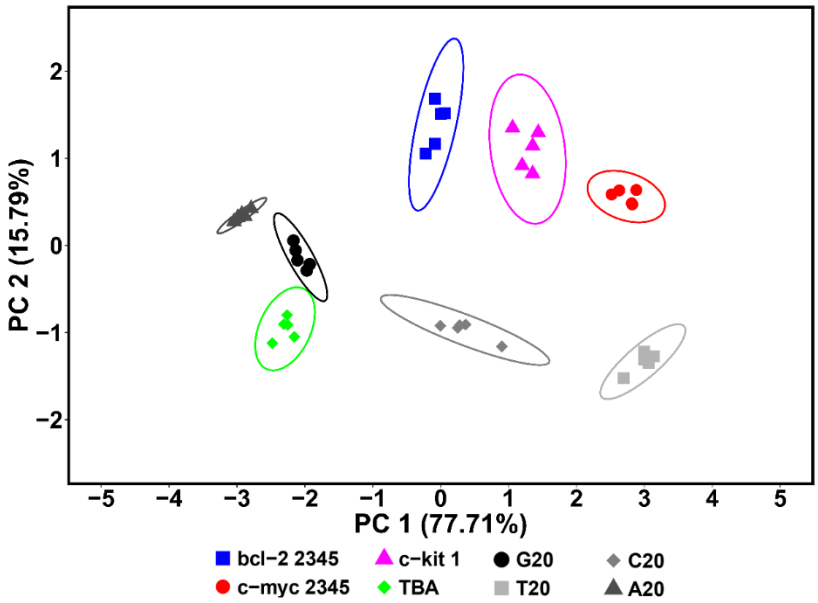
5.2 DNA Detection with PSMI

Supplementary Table 3 LOD calculated from the titration curves shown in Section 5.2.2 using the 3σ method.

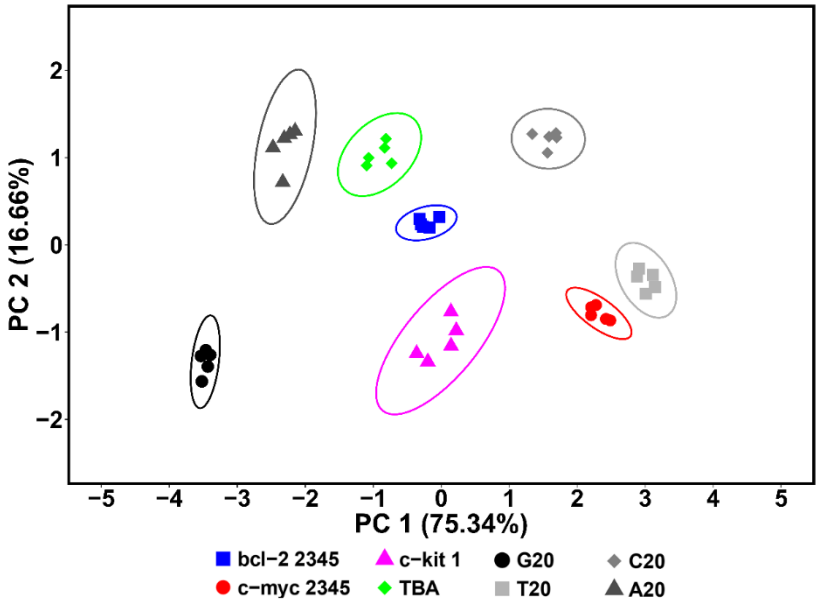
LOD/PSMI	A20	*c-myc 2345*
Host 2	0.03125±0.00206μM	0.02000±0.00084μM
Host 3	0.02878±0.00038μM	0.03348±0.00061μM
Host 4	0.03151±0.00031μM	0.01986±0.00082μM
Host 5	0.03469±0.00074μM	0.03585±0.00053μM
No Host	0.00307±0.00002μM	0.02120±0.00095μM

6. Array Analysis for differentiation of 8 DNAs

6.1 PCA Results with the Type 1 or Type 2 array

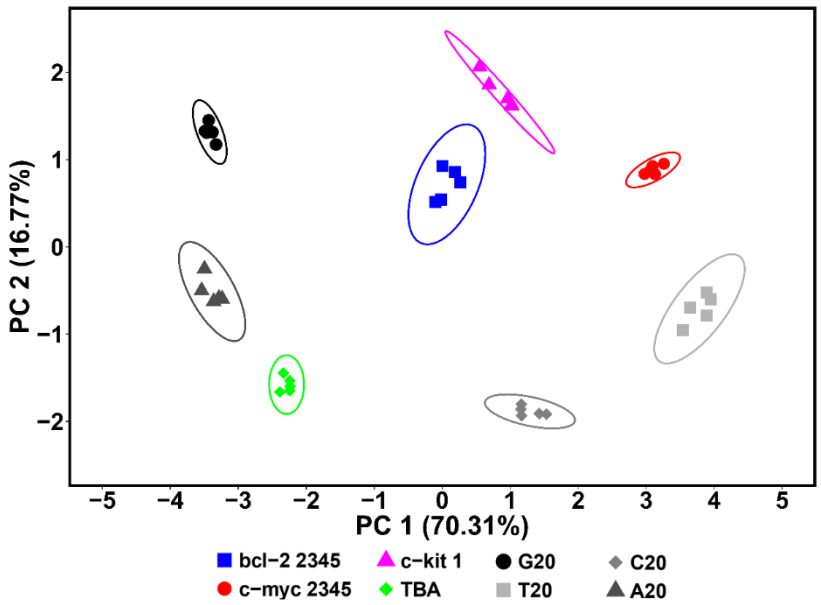


Supplementary Fig. 22. PCA scores plot for selective sensing of 8 DNA structures using **Type 1 Array** (hosts **1-5** or no host, **DSMI** dye)**:** [**DSMI**] = 0.625 μM, [**1**, **4**, or **5**] = 0.25 μM, [**2** or **3**] = 0.50 μM. [DNA] = 0.1 μM, buffer: 10 mM KH_2PO_4/K_2HPO_4, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm. Ellipses indicate 95% confidence.



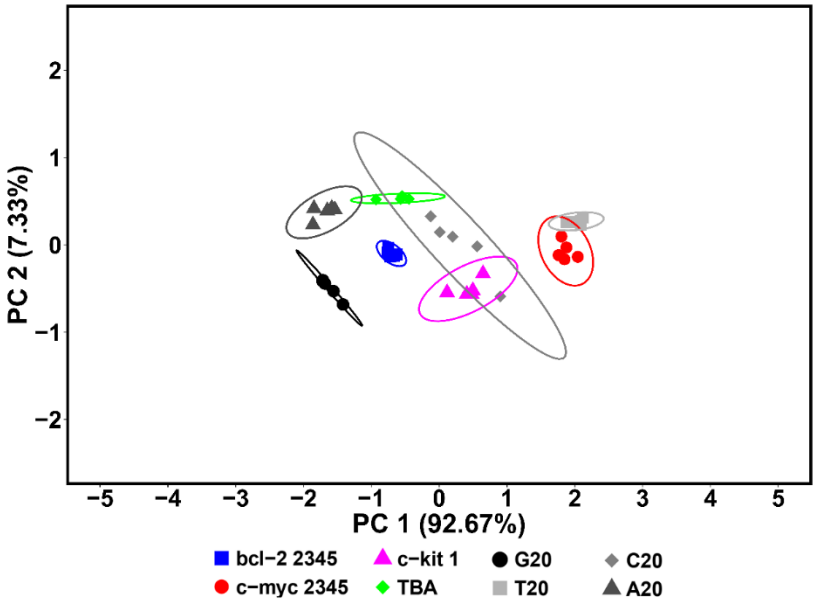
Supplementary Fig. 23. PCA scores plot for selective sensing of 8 DNA structures using **Type 2 Array** (hosts **1-5** or no host, **PSMI** dye)**:** [**PSMI**] = 0.625 μM, [**1**, **3**, **or 5**] = 0.25 μM, [**2**] = 1.0 μM, [**4**] = 0.50 μM. Other sensor conditions identical to those described in Supplementary Fig. 22. Ellipses indicate 95% confidence.

6.2 PCA Results with the 10 Host-Dye Array Elements



Supplementary Fig. 24. PCA scores plot for selective sensing of 8 DNA structures using the **10-component array** (the combination of **Type 1** and **Type 2** arrays, but not including "no host"). Other sensor conditions identical to those described in Supplementary Fig. 22. Ellipses indicate 95% confidence.

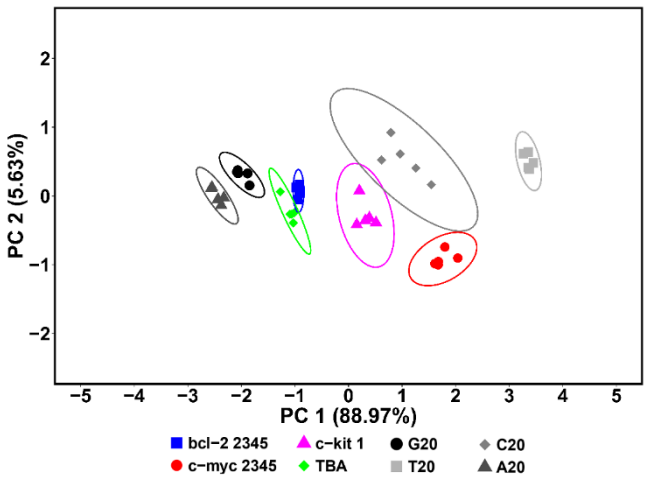
6.3 PCA results using only 2 dyes as array elements



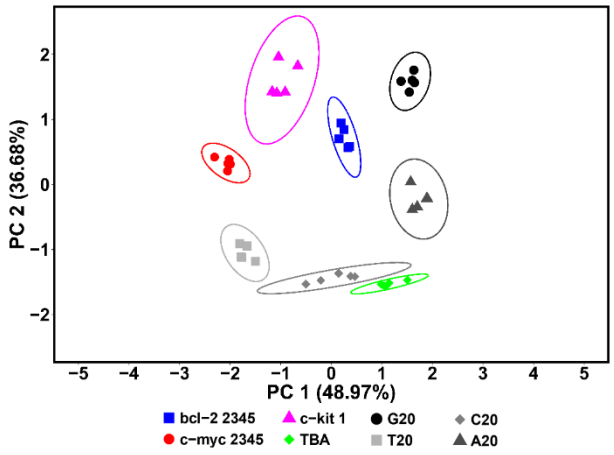
Supplementary Fig. 25. PCA scores plot for selective sensing of 8 DNA structures using the **2-component array** (only **DSMI** or **PSMI** dye)**:** [**DSMI/PSMI**] = 0.625 μM. Other sensor conditions identical to those described in Supplementary Fig. 22. Ellipses indicate 95% confidence.

6.4 PCA results using 4 array elements: dyes only plus Host 1, 2, or 4

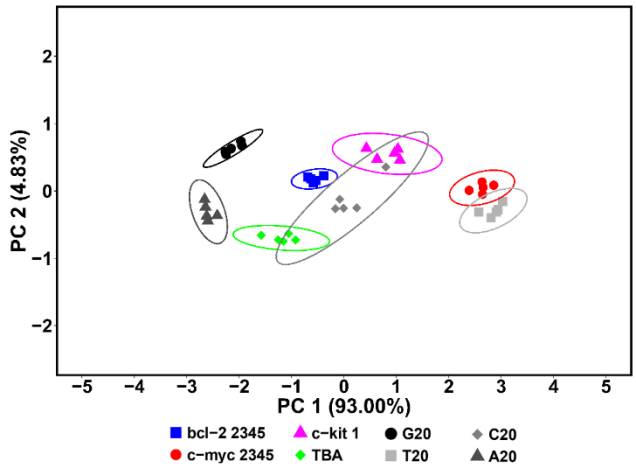
a)



b)



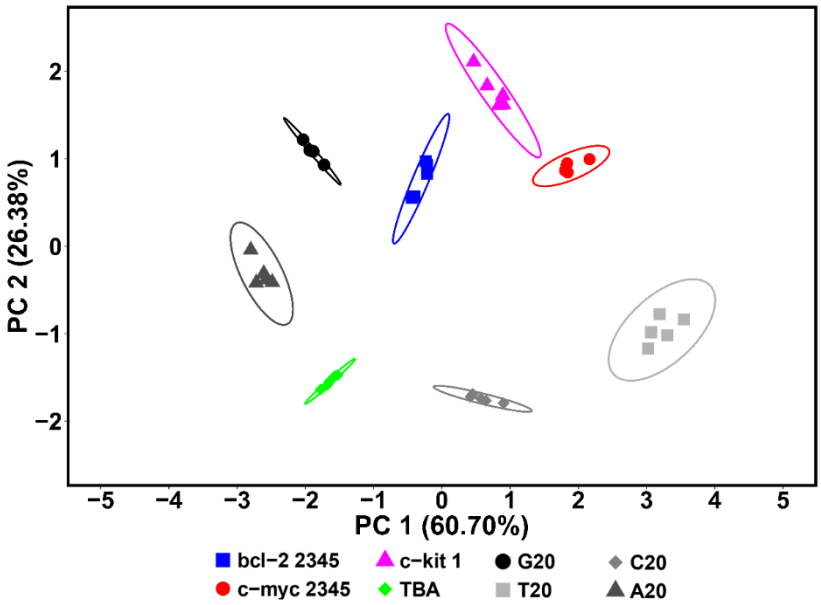
c)



Supplementary Fig. 26. PCA scores plots for selective sensing of 8 DNA structures using the **4-component array:** a) **DSMI/PSMI** + host **1** or no host from **Type 1** and **Type 2** arrays; b) **DSMI/PSMI** + host **2** or no host from **Type 1** and **Type 2** arrays; c) **DSMI/PSMI** + host **4** or no

host from **Type 1** and **Type 2** arrays. Other sensor conditions identical to those described in Supplementary Fig. 22. Ellipses indicate 95% confidence.

6.5 PCA results using 6 array elements: Dye·Host 1, 2, and 4

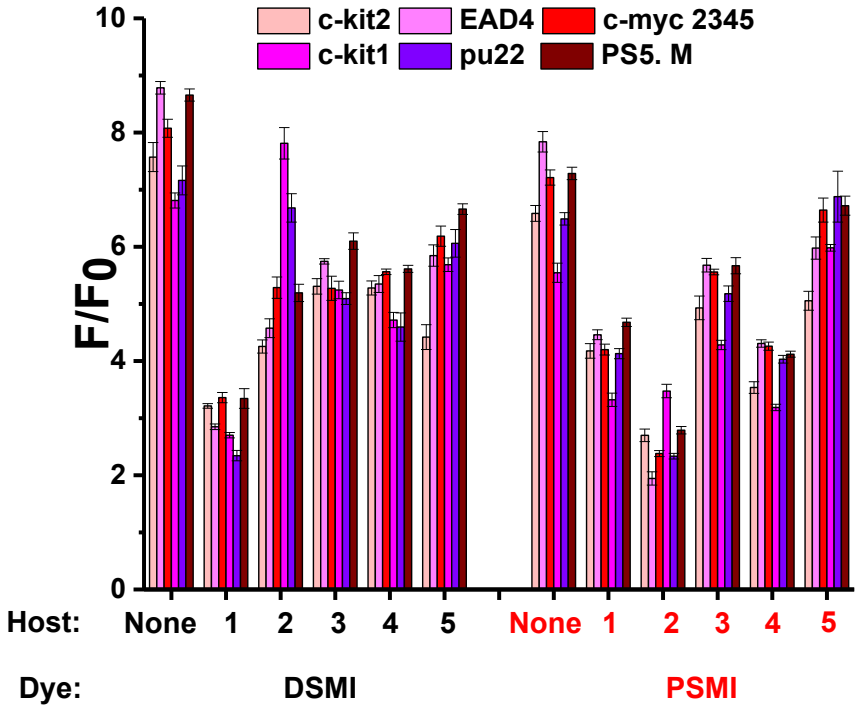


Supplementary Fig. 27. PCA scores plot for selective sensing of 8 DNA structures using the **6-component array** (**DSMI**/**PSMI** + hosts **1**, **2**, or **4** from **Type 1** and **Type 2** arrays). Other sensor conditions identical to those described in Supplementary Fig. 22. Ellipses indicate 95% confidence.

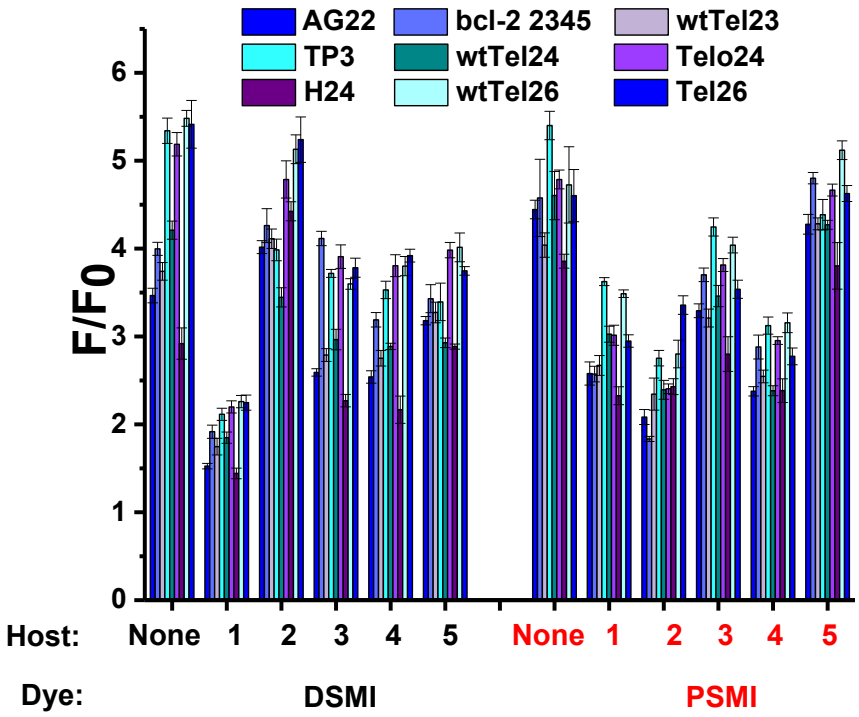
7. Array Analysis for Differentiation of 23 G4 DNA

7.1 Bar Plots for Array Signals from G4 with Different Topologies

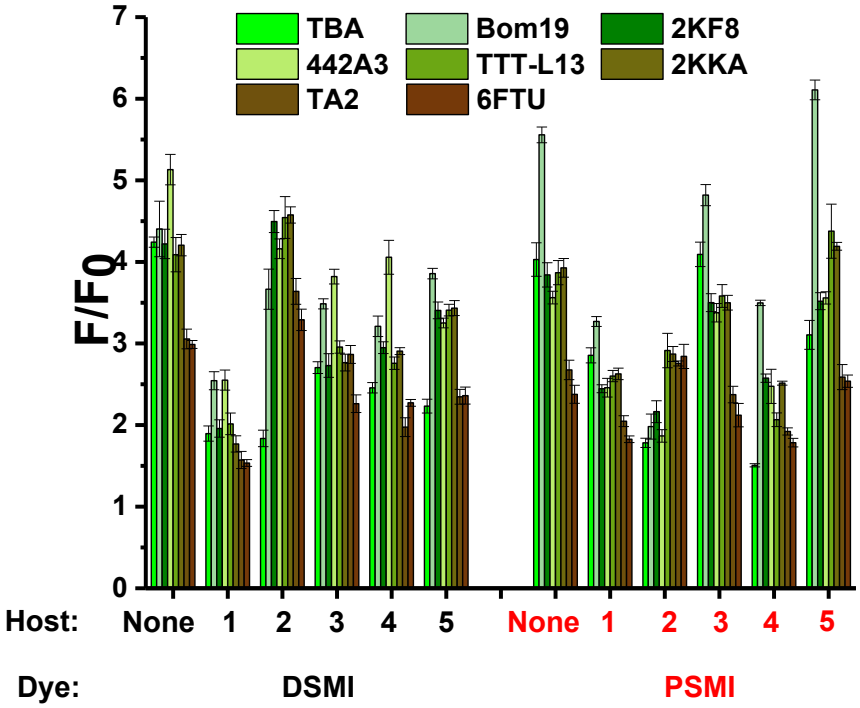
a) **Parallel G4**



b) **Hybrid G4**



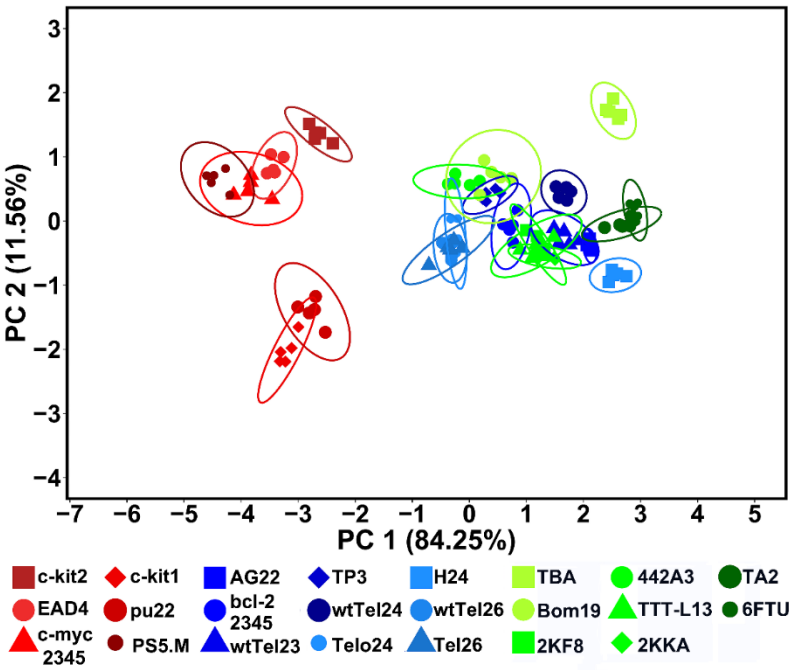
c) **Antiparallel G4**



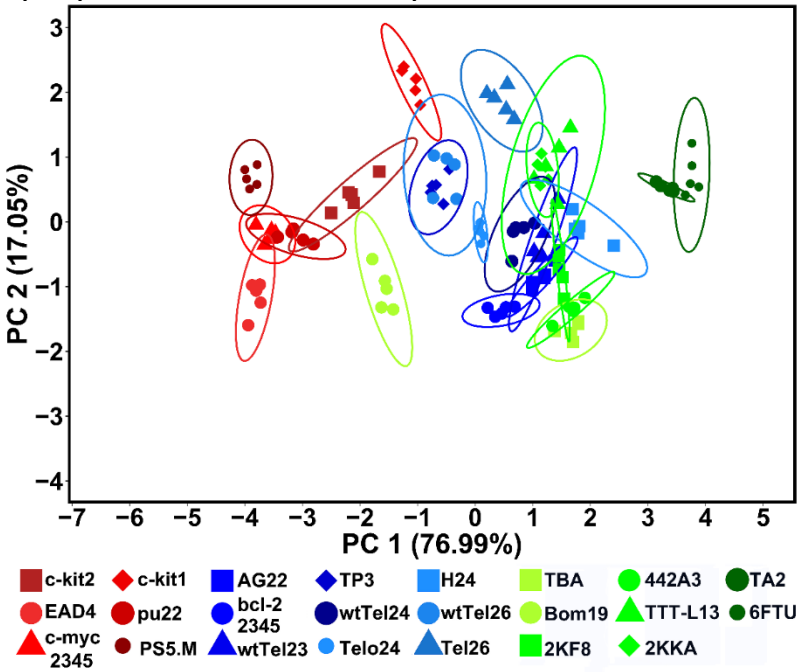
Supplementary Fig. 28. Full fluorescence response plots of a) parallel G4 sequences, b) hybrid G4 sequences, c) antiparallel G4 sequences, obtained with the **12-component array** (the combination of **Type 1** and **Type 2** arrays). Other sensor conditions identical to those described in Supplementary Fig. 22.

7.2 PCA Score Plots for Differentiation of all 23 G4

7.2.1 PCA Results with the Type 1, or Type 2 Array



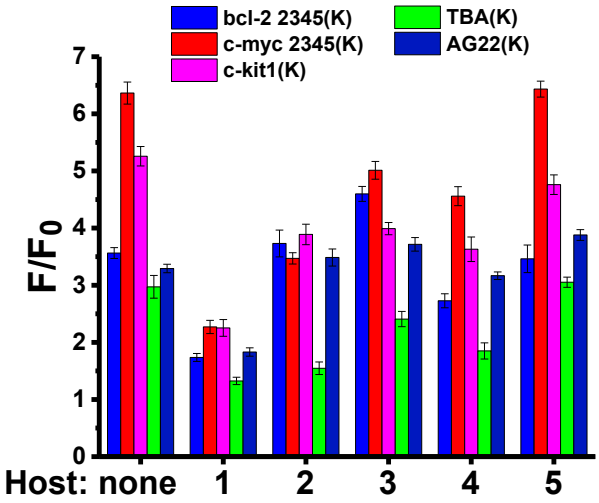
Supplementary Fig. 29. G-quadruplex classification with only the **Type 1 Array**. PCA scores plot generated from the corresponding data in Supplementary Fig. 28, by treating each repeat as one individual sample, processed in RStudio. Ellipses indicate 95% confidence.



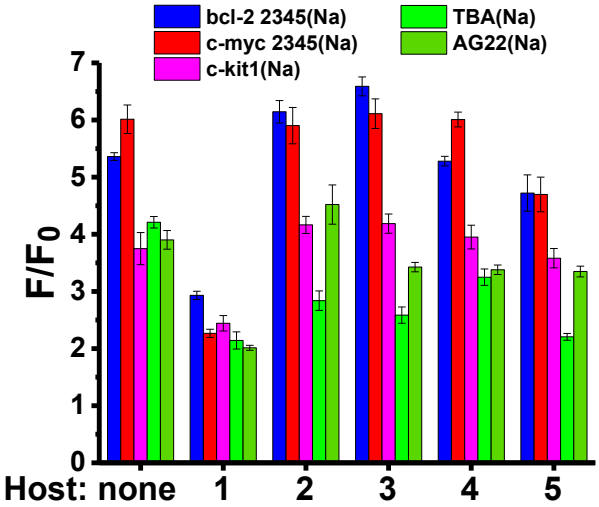
Supplementary Fig. 30. G-quadruplex classification with only the **Type 2 Array**. PCA scores plot generated from the corresponding data in Supplementary Fig. 28, by treating each repeat as one individual sample, processed in RStudio. Ellipses indicate 95% confidence.

8. Array Analysis for Topology Switch in K^+ and Na^+ Buffer

a)



b)



Supplementary Fig. 31. Fluorescence response plots for structural Switching of ***AG22*** detected by the host-guest sensor array in a) 10 mM KH_2PO_4/K_2HPO_4 buffer; or b) 10 mM NaH_2PO_4/Na_2HPO_4 buffer, using the **Type 1** sensing array. [DNA] = 0.1 μM, [buffer] = 10 mM, 1 mM EDTA, pH 7.4, Ex/Em = 485nm/605nm.

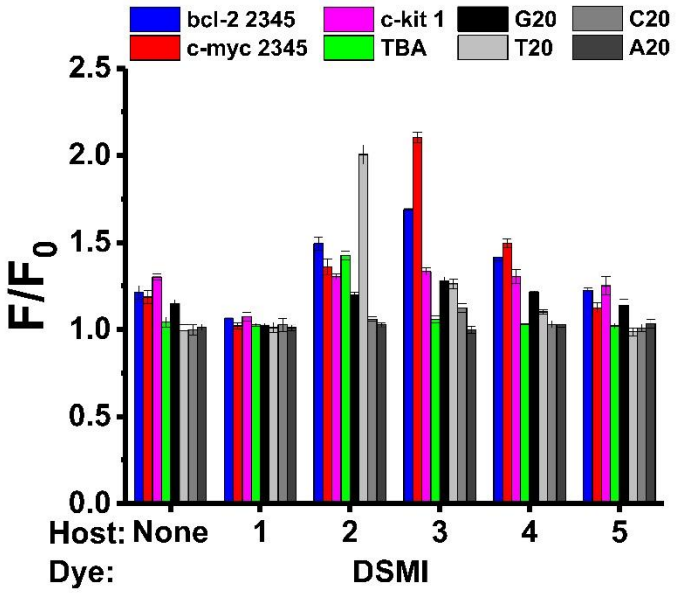
9. Array Performance in Complex Media.

Note: The following data were acquired in BioTek Synergy HT microplate reader with the filter set of Ex485/Em600nm.

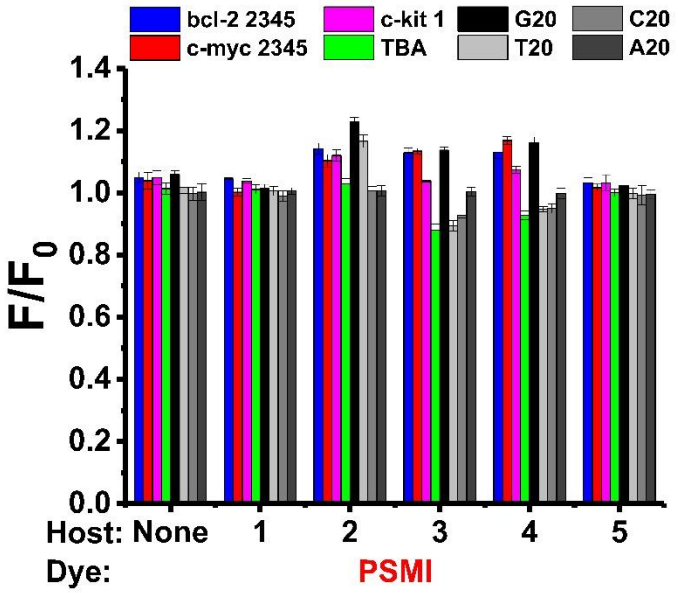
9.1 In Solutions Containing Fetal Bovine Serum (FBS)

9.1.1 Bar Plots for DNA Sensing Array in Solutions Containing 5% FBS

a)



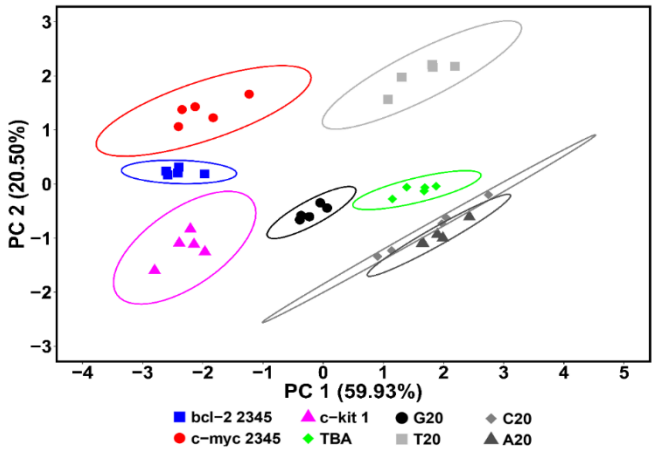
b)



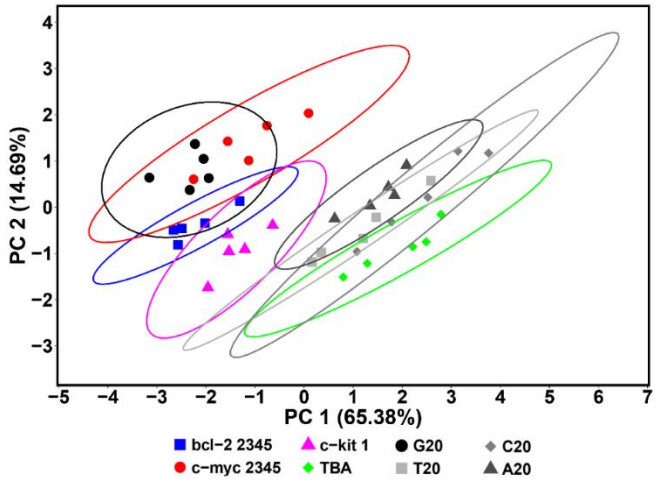
Supplementary Fig. 32. Fluorescence responses upon addition of the eight DNA strands to the **host•dye** sensor components in solutions containing 5% FBS. a) **Type 1** array; b) **Type 2** array. [DNA] = 0.1 μM, 5% FBS, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

9.1.2 PCA Score Plots for DNA Sensing Array in Solutions Containing 5% FBS

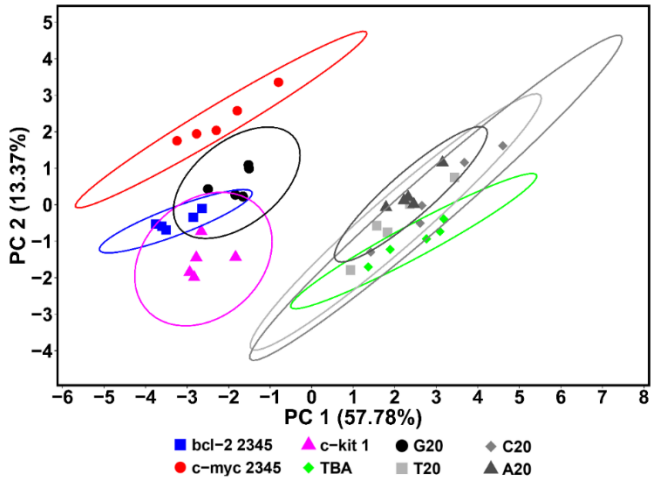
a)



b)



c)

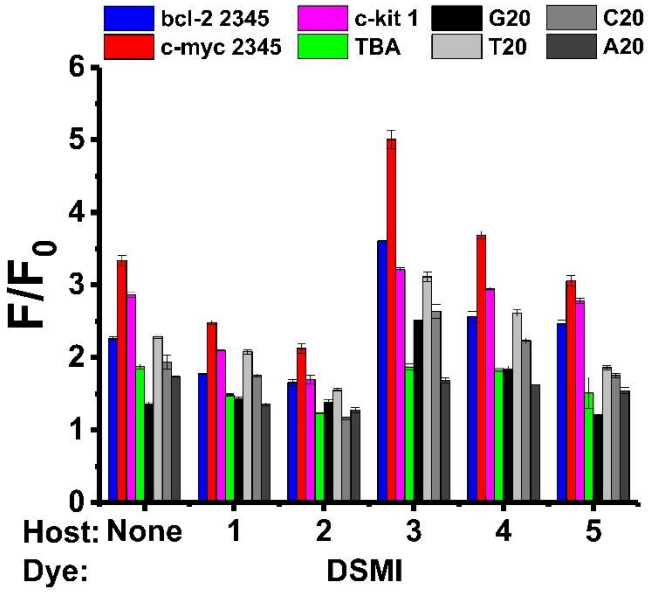


Supplementary Fig. 33. PCA scores plot generated from the data for DNA sensing array in solutions containing 5% FBS in Supplementary Fig. 34. a) **Type 1** array; b) **Type 2** array; c) the combination of **Type 1** and **Type 2** arrays. [DNA] = 0.1 μM, 5% FBS, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4. Ellipses indicate 95% confidence.

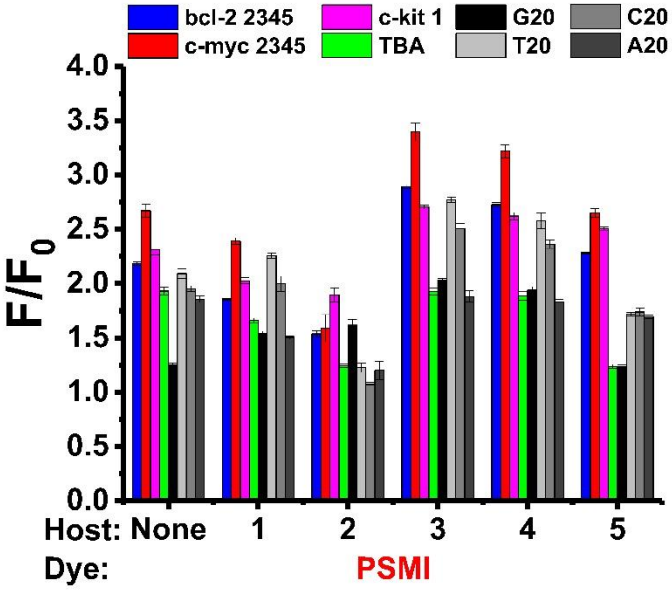
9.2 In the Presence of lysozyme

9.2.1 Bar Plots for Sensing Array in the Presence of 0.1 μM Lysozyme

a)



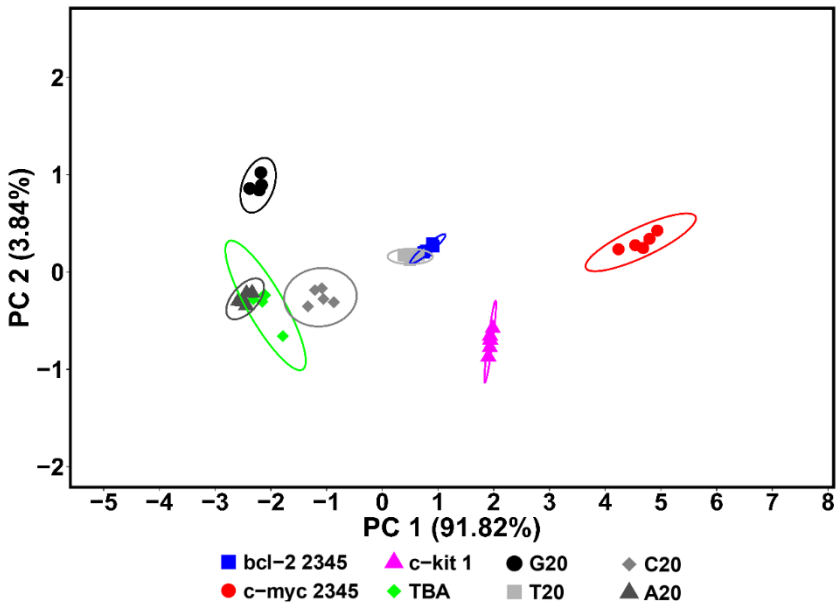
b)



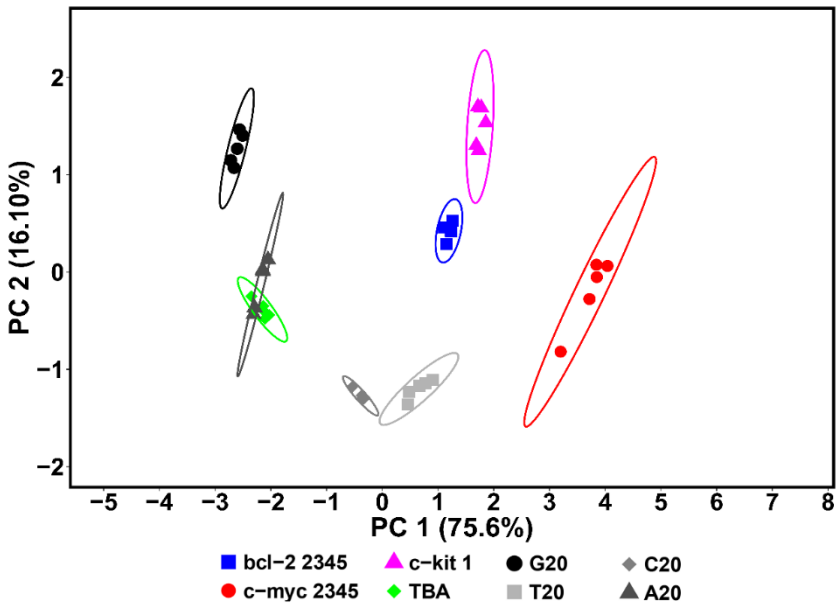
Supplementary Fig. 34. Fluorescence responses upon addition of the eight DNA strands to the **host•dye** sensor components with the presence of 0.1 μM lysozyme. a) **Type 1** array; b) **Type 2** array. [DNA] = 0.1 μM, 0.1 μM lysozyme, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

9.2.2 PCA Score Plots for DNA Sensing Array in the Presence of 0.1 µM Lysozyme

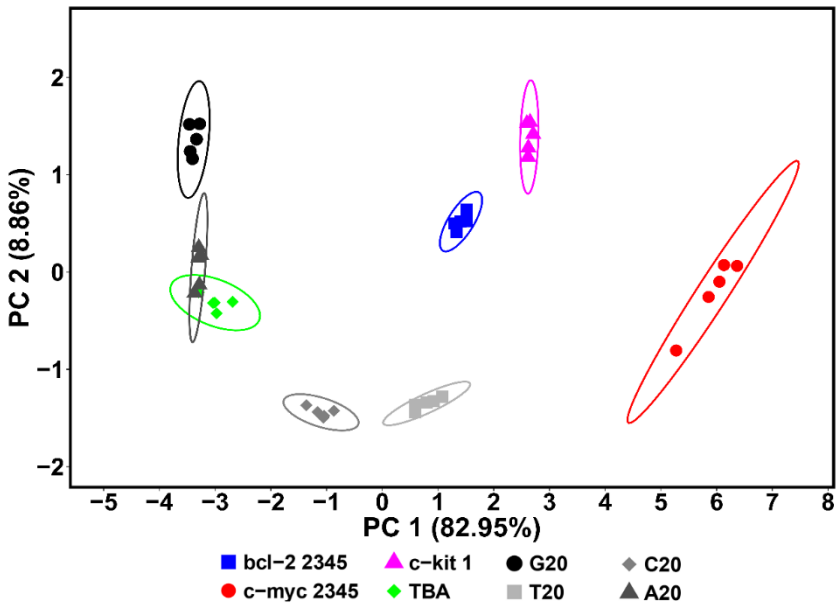
a)



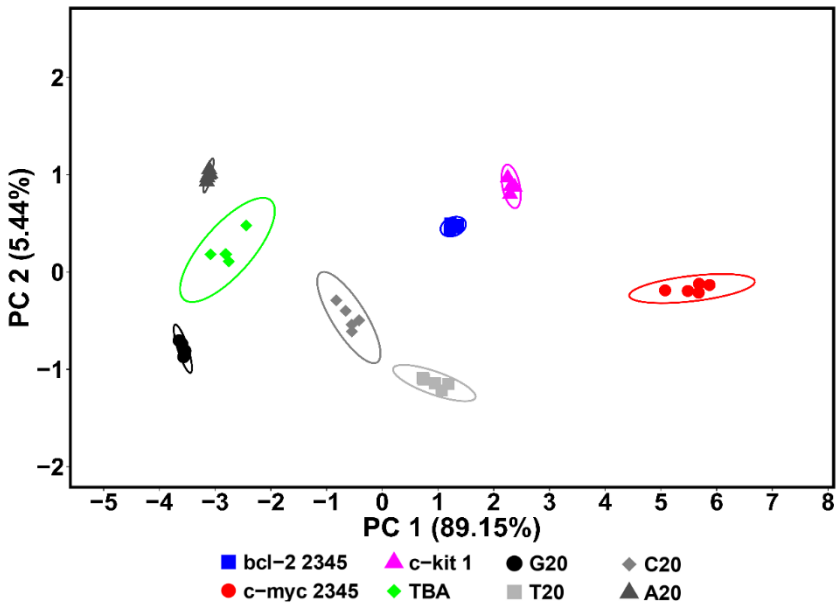
b)



c)



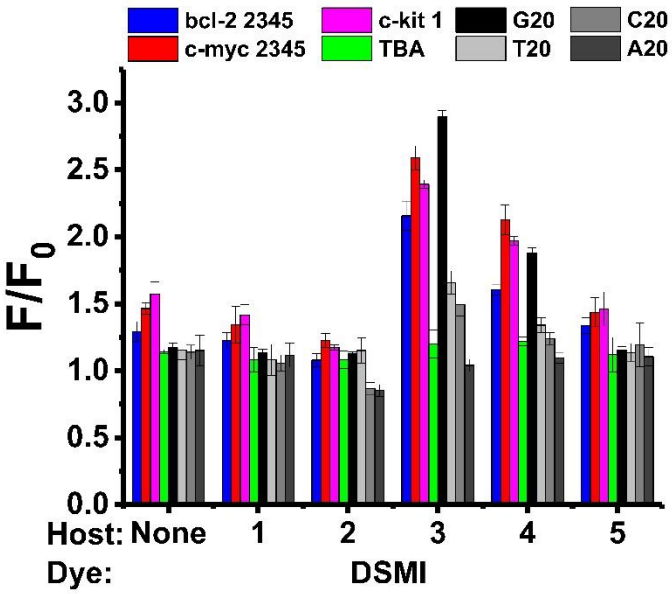
d)



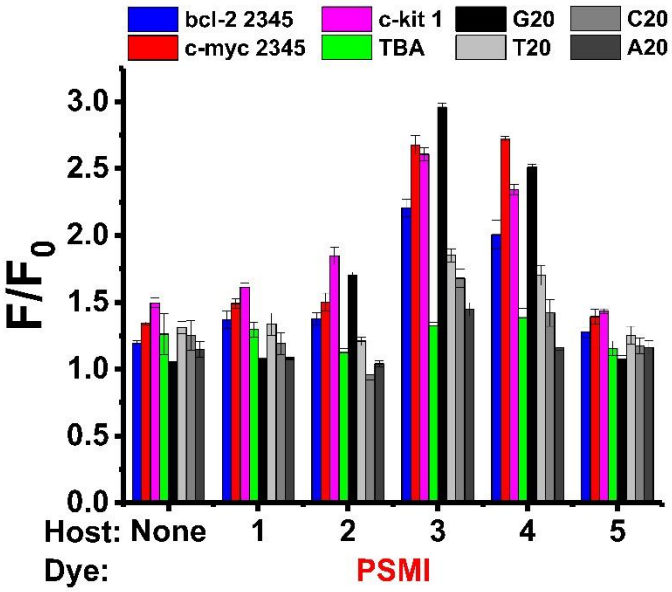
Supplementary Fig. 35. PCA scores plot generated from the data for DNA sensing array with the presence of 0.1 μM lysozyme in Supplementary Fig. 34. a) **Type 1** array; b) **Type 2** array; c) the combination of **Type 1** and **Type 2** arrays; d) the 10-component array (not including "host **2**"). [DNA] = 0.1 μM, 0.1 μM lysozyme, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

9.2.3 Bar Plots for Sensing Array in the Presence of 1 μM Lysozyme

a)

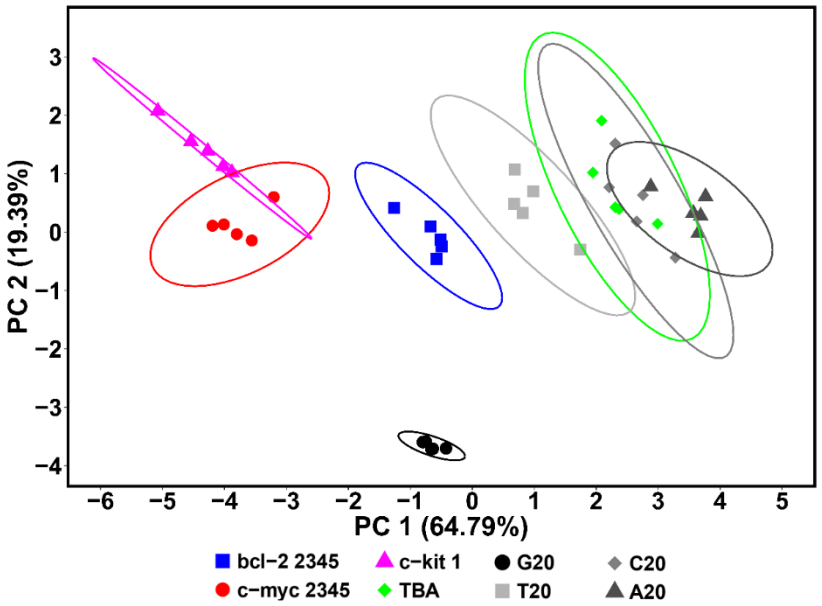


b)



Supplementary Fig. 36. Fluorescence responses upon addition of the eight DNA strands to the **host•dye** sensor components with the presence of 1 μM lysozyme. a) **Type 1** array; b) **Type 2** array. [DNA] = 0.1 μM, 1 μM lysozyme, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

9.2.4 PCA Score Plots for Sensing Array in the Presence of 1 μM Lysozyme

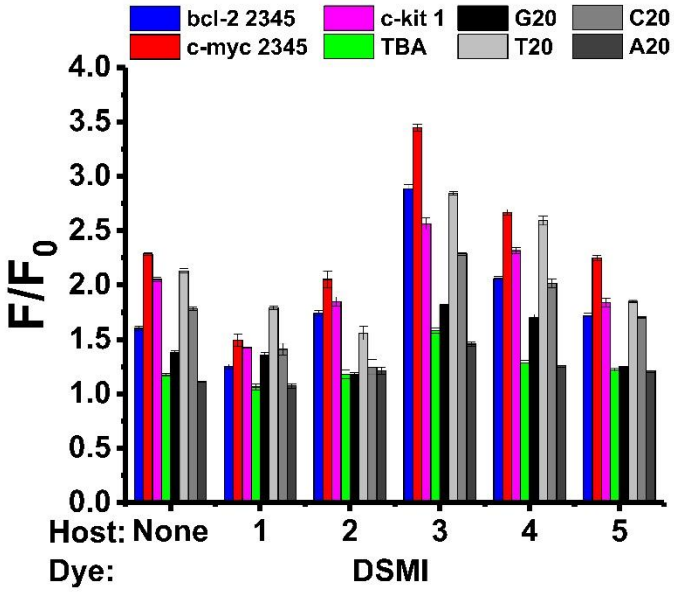


Supplementary Fig. 37. PCA scores plot generated from the data for DNA sensing array with the presence of 1 μM lysozyme in Supplementary Fig. 36 using the 12-component array (the combination of **Type 1** and **Type 2** arrays. [DNA] = 0.1 μM, 1 μM lysozyme, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

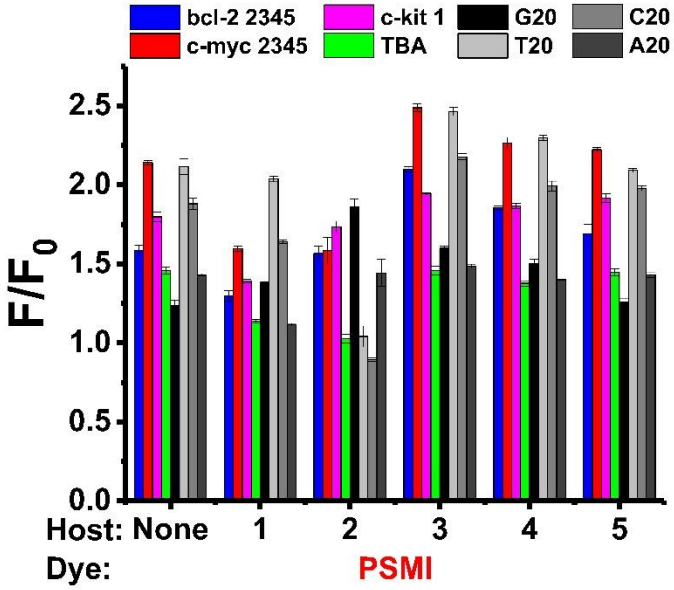
9.3 In the presence of Lactose

9.3.1 Bar Plots for Sensing Array in the Presence of 20 μM Lactose

a)



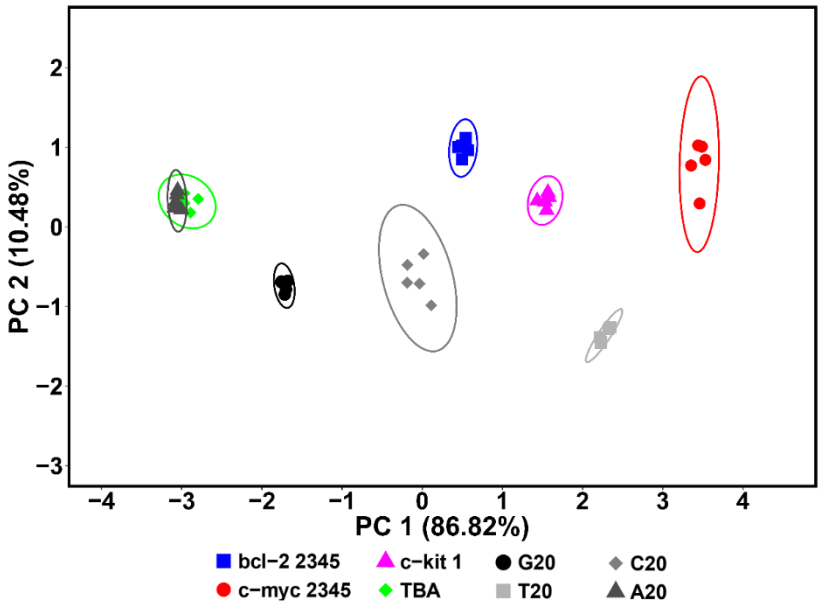
b)



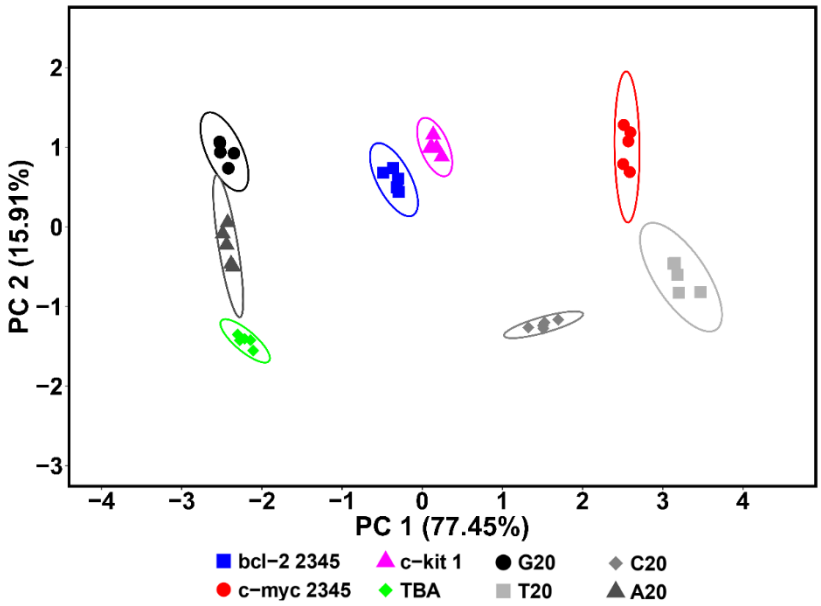
Supplementary Fig. 38. Fluorescence responses upon addition of the eight DNA strands to the **host•dye** sensor components with the presence of 20 μM lactose. a) **Type 1** array; b) **Type 2** array. [DNA] = 0.1 μM, 20 μM lactose, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

9.3.2 PCA Score Plots for Sensing Array in the Presence of 20 µM Lactose

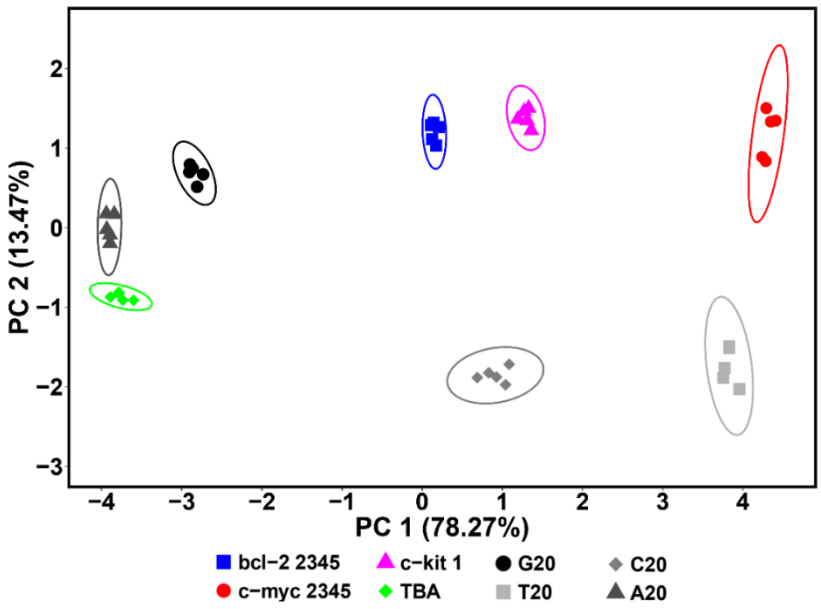
a)



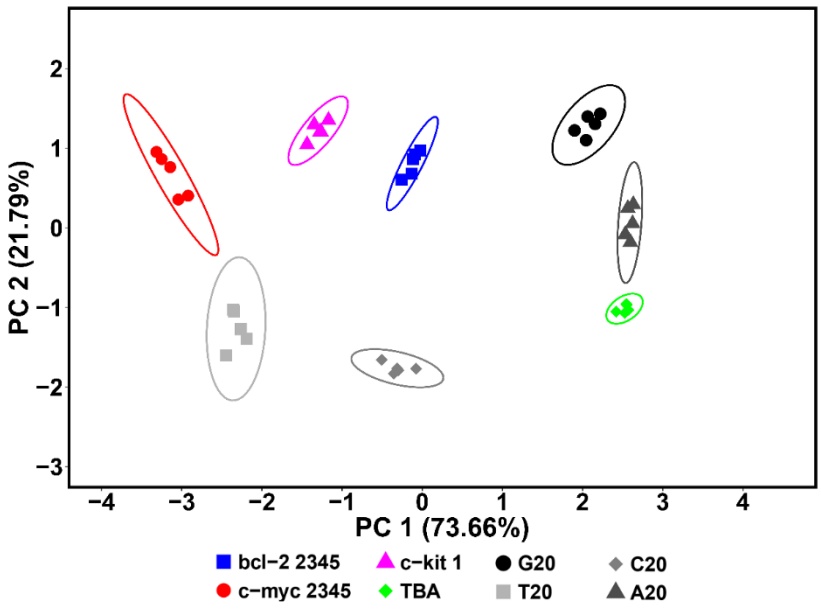
b)



c)



d)



Supplementary Fig. 39. PCA scores plot generated from the data for DNA sensing array with the presence of 20 μM lactose in Supplementary Fig. 38. a) **Type 1** array; b) **Type 2** array; c) the combination of **Type 1** and **Type 2** arrays; d) the 6-component array: **DSMI**/**PSMI** + hosts 2, 4 or no host from **Type 1** and **Type 2** arrays. [DNA] = 0.1 μM, 20 μM lactose, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

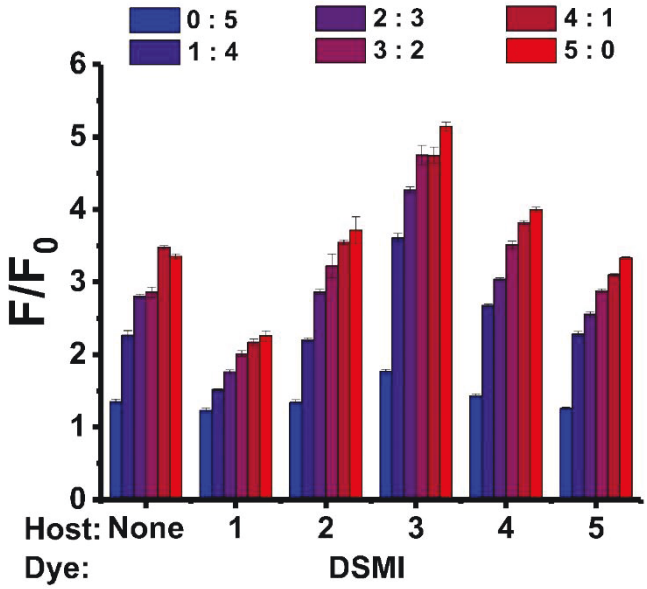
10. Array-based Recognition of the Mixture of Non-structural DNA and G4.

Note: The following data were acquired in BioTek Synergy HT microplate reader with the filter set of Ex485/Em600nm.

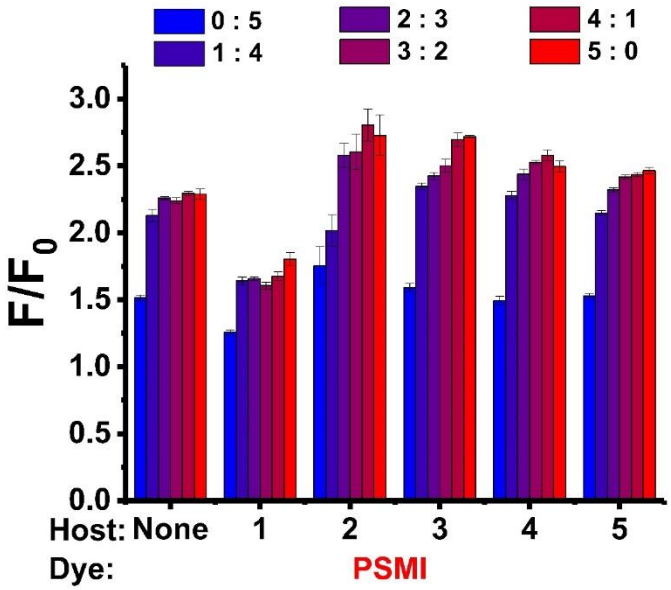
a)

[DNA]/uM	①	②	③	④	⑤	⑥	blank
poly A 20	0.5	0.4	0.3	0.2	0.1	0	0
c-myc 2345	0	0.1	0.2	0.3	0.4	0.5	0
Mole ratio of c-myc 2345 to poly A20	0: 5	1 : 4	2 : 3	3 : 2	4 : 1	5 : 0	-

b)



c)



Supplementary Fig. 40. Array-based analysis of the mixtures of *A20* and *c-myc 2345*. a) The component concentrations in each test solution. Fluorescence responses for b) the **DSMI**-based array elements and c) the **PSMI**-based array elements to the solutions. Total [DNA] = 0.5 μM, 10 mM KH_2PO_4/K_2HPO_4 buffer, 1 mM EDTA, pH 7.4.

11. References

1. Jin, M. *et al.* Thiazole Orange-Modified Carbon Dots for Ratiometric Fluorescence Detection of G-Quadruplex and Double-Stranded DNA. *ACS Appl. Mater. Interfaces* **10**, 25166-25173 (2018).
2. Cheng, X., Liu, X., Bing, T., Cao, Z. & Shangguan, D. General Peroxidase Activity of G-Quadruplex−Hemin Complexes and Its Application in Ligand Screening. *Biochemistry* **48**, 7817-7823 (2010).
3. Yang, Q. *et al.* Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: II. The binding characterization with specific intramolecular G-quadruplex and the recognizing mechanism. *Nucleic Acids Res.* **38**, 1022-1033 (2009).
4. Lin, D. *et al.* A benzindole substituted carbazole cyanine dye: a novel targeting fluorescent probe for parallel c-myc G-quadruplexes. *Analyst* **140**, 5772-5780 (2015).
5. Ferreira, J. *et al.* Naphthalene amine support for G-quadruplex isolation. *Analyst* **142**, 2982-2994 (2017).
6. 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+: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.* **41**, 10584-10592 (2013).
7. Peng, H. *et al.* DNAzyme-mediated assays for amplified detection of nucleic acids and proteins. *Anal. Chem.* **90**, 190-207 (2017).
8. Gu, Y. *et al.* A light-up probe targeting for Bcl-2 2345 G-quadruplex DNA with carbazole TO. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **191**, 180-188 (2018).
9. Lin, C. *et al.* Molecular Recognition of the Hybrid-2 Human Telomeric G-Quadruplex by Epiberberine: Insights into Conversion of Telomeric G-Quadruplex Structures. *Angew. Chem.* **130**, 11054-11059 (2018).
10. Sengar, A. *et al.* Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. *Nucleic Acids Res.* **47**, 1564-1572 (2019).
11. Ma, Y. *et al.* Development of G-quadruplex ligands for selective induction of a parallel-type topology. *Org. Biomol. Chem.* **16**, 7375-7382 (2018).
12. Wu, G., Chen, L., Liu, W. & Yang, D. Molecular Recognition of the Hybrid-Type G-Quadruplexes in Human Telomeres. *Molecules* **24**, 1578 (2019).
13. Hao, F., Ma, Y. & Guan, Y. Effects of Central Loop Length and Metal Ions on the Thermal Stability of G-Quadruplexes. *Molecules* **24**, 1863 (2019).
14. Amrane, S. *et al.* A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence. *Nucleic Acids Res.* **37**, 931-938 (2009).
15. Del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **57**, 7171-7175 (2018).
16. Cheng, M. *et al.* Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.* **46**, 9264-9275 (2018).
17. Guédin, A. *et al.* Quadruplexes in 'Dicty': crystal structure of a four-quartet G-quadruplex formed by G-rich motif found in the Dictyostelium discoideum genome. *Nucleic Acids Res.* **46**, 5297-5307 (2018).