

Sensing Base Modifications in Non-Canonically Folded DNA with an Optimized Host:Guest Sensing Array

Junyi Chen,² Briana L. Hickey,¹ Ziting Gao,¹ Alexie Andrea P. Raz,¹ Richard J. Hooley^{1*} and Wenwan Zhong^{1, 2*}

¹Department of Chemistry; ²Environmental Toxicology Graduate Program; University of California-Riverside, Riverside, CA 92521, U.S.A. * E-mail: richard.hooley@ucr.edu; wenwan.zhong@ucr.edu

ABSTRACT: An arrayed host:guest fluorescence sensor system can discriminate DNA G-quadruplex structures that differ only in the presence of single oxidation or methylation modification in the guanine base. These small modifications make subtle changes to G4 folding that are often not detectable by CD, but induce differential fluorescence responses in the array. The sensing is functional in diluted serum, and is capable of distinguishing individual modifications in DNA mixtures, providing a powerful method of detecting folding changes caused by DNA damage.

KEYWORDS: *Molecular recognition, Biosensors, DNA Structures, DNA Damage, Modified Nucleobases*

Structural modifications in DNA bases are important disease markers,¹ and can be caused by a number of agents, including toxic chemicals, oxidative damage and ionizing radiation.² These modifications can greatly affect the structure of non-canonically folded DNA: for example, guanine is prone to oxidation and alkylation, and G-rich sequences tend to form non-canonical G-quadruplex (G4) structures.³ But as non-Watson-Crick folding motifs are less prevalent and less stable than duplex DNA, the effects of base modification can be more variable and more challenging to detect.⁴ Guanine methylation and oxidation modifications alter the size, shape and H-bonding properties of individual bases, but the changes they induce are often small: they often only destabilize folding motifs, rather than effecting complete unfolding or a discrete topology switch.

Comprehensive analysis of structural changes induced by nucleotide modification is possible using X-Ray crystallography and/or multidimensional NMR spectroscopy,⁵ but is expensive, time-consuming and requires large amounts of material, and other methods such as base-resolution sequencing or LC-MS/MS are destructive.⁶ As such, a rapid optical method of detecting the presence of base modifications in DNA would be highly desirable. Common optical techniques such as circular dichroism (CD) can provide some information,⁷ but fall short when distinguishing small changes in structure. Differential sensing⁸ is a powerful tool for detecting small changes in structure for different biological targets,⁹ including oligonucleotides.¹⁰ We have recently shown that multicomponent cavitand:dye arrays can distinguish different types of G-quadruplex (G4) topology,¹¹ as well as other folding patterns.^{11b} As these arrays showed selectivity for small structural changes of non-canonically folded DNA (such as G4s with vacancies or bulges), they

seemed ideal for detecting even smaller differences in DNA structure, namely modifications of individual guanine bases in G4 strands. Here we show that cavitand:dye arrays, optimized under the guidance of machine learning, can differentiate G4s containing only a single modified G base in the strand.

One challenge in analyzing non-canonically folded DNA is that small changes in environment (such as salt concentration and pH) can cause defects and changes in the folding topology from sample to sample. Therefore, we chose modified DNA targets that were already known in the literature and displayed characterized folding topologies upon modification. We focused on two types of modified guanine base (Figure 1): either methylation of the oxygen at the 6 position (6-mG), or oxidation at carbon 8 (8-oxoG). While these modifications involve small differences in base size (replacing two hydrogens with a single oxygen atom, for example), they alter the H-bonding properties of the bases, which can have highly variable effects on G4 topology. As such, we targeted a pool of 9 DNA G4s (3 strand types) of different lengths and sequences, each with a single modification of either 6-mG or 8-oxoG at varying positions, either in one of the G-quartets, or in the loop (Figure 1d). The three strand types were based on 22-nt strand *AG22* (with 6-mG at the 2 and 3 positions, both on the G-quartet),¹² 26-nt strand *PCNA* (modified at the 3 and 7 positions with 8-oxoG)¹³ and 32-nt strand *32R* (modified at the 6 and 11 positions with 8-oxoG).¹⁴ Two other control G4 strands were also tested, *HT-T5 original* (hybrid) and *MYOG-3332* (antiparallel in the experimental conditions used, see Supporting Information for full sequences).

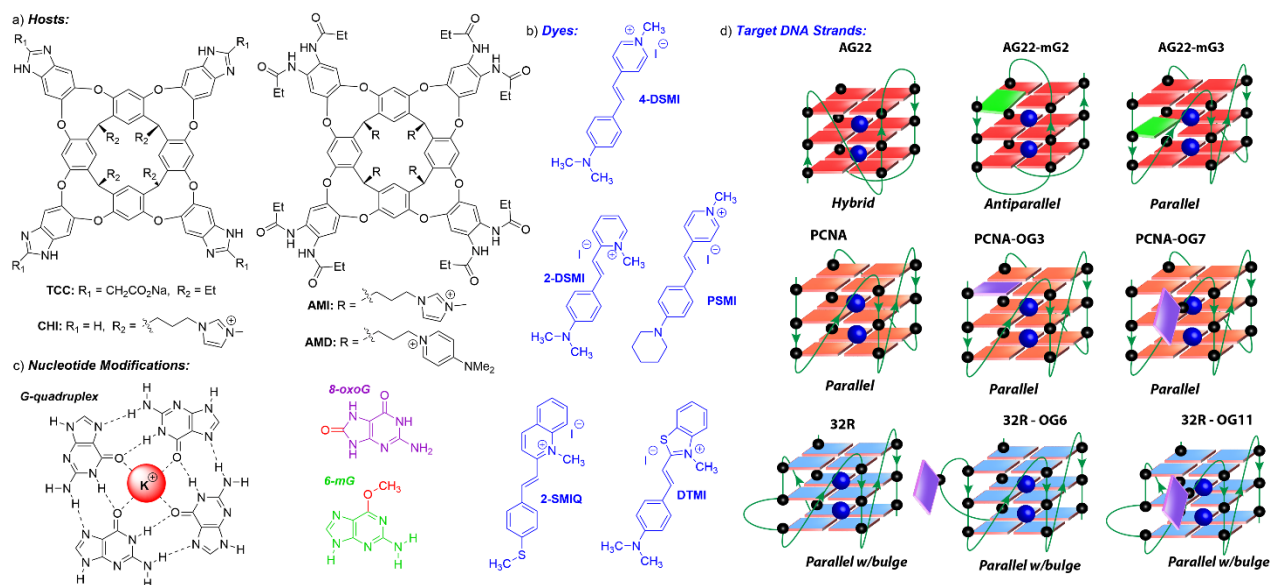


Figure 1. Host:Guest fluorescence sensing array. a) Host structures; b) Dye structures; c) Structures of an unmodified G-quartet and guanine modifications. d) Series of modified DNAs (and parent structures) used, including their topologies.

Our DNA-targeting sensor arrays¹¹ consist of a suite of water-soluble deep cavitands and styrylpyridinium dyes (Figure 1a,b). The flat, cationic dyes bind to both the cavitand hosts and the G4s, forming a competitive recognition equilibrium that includes ternary complexes of **host•dye•G4**.^{11b} The fluorescence emission profiles of each **host•dye** combination provide a variable fingerprint that can be processed by differential analysis techniques.^{8b} We expected that distinguishing base modifications would be challenging, so performed the initial tests with a large, unbiased 25-component array of 4 cavitands and 5 dyes (Figure 1 a,b). Some of these dyes and hosts have been used previously,¹¹ but others (notably **AMD** cavitand and **2-DSMI**, **DTMI** and **2-SMIQ** dyes) were synthesized for this task (see Supporting Information), to enhance the selectivity of the array.

The 11 target DNA strands ([DNA] = 0.1 μ M) were added to the full array ([host]=[dye] = 0.625 μ M) in 10 mM potassium phosphate buffer, 1 mM EDTA, pH 7.4, and the fluorescence recorded in each well (five repeats) at the optimal Ex/Em wavelengths for each dye (F and F₀ are defined as the fluorescence of the Host:Guest element, with and without addition of target DNA; see Supporting Information for full fluorescence data). The fluorescence responses were analyzed by Principal Component Analysis (PCA): the scores plot is shown in Figure 2a. The results were quite unsatisfactory, especially considering the size of the sensing pool: multiple overlaps in the scores plot can be seen, even between ostensibly different G4 topologies such as **PCNA** and modified **AG22**. This illustrates the weakness of a “brute-force” differential sensing approach: more array elements do not necessarily lead to better sensing and discrimination. Evidently certain cavitand:dye combinations had a deleterious effect on the overall sensing outcome,¹⁵ so a judicious narrowing of the sensing pool was in order. The loading vectors resulting from PCA were analyzed, and showed that the sensor elements involving **2-SMIQ** and **TCC** had lower contributions than other dye/host sensor elements to PC1, which summarized 67.2% variance of the dataset (Figure S-22, 23). In addition, the loading vectors for

the **TCC**-based sensors were similar to the those without cavitand, and the vectors for the **2-SMIQ** elements were quite comparable to each other. As a result, these elements (thioquinoline dye **2-SMIQ** and the anionic cavitand **TCC**) were deemed “failures”, so were removed from the pool, and the data reanalyzed.

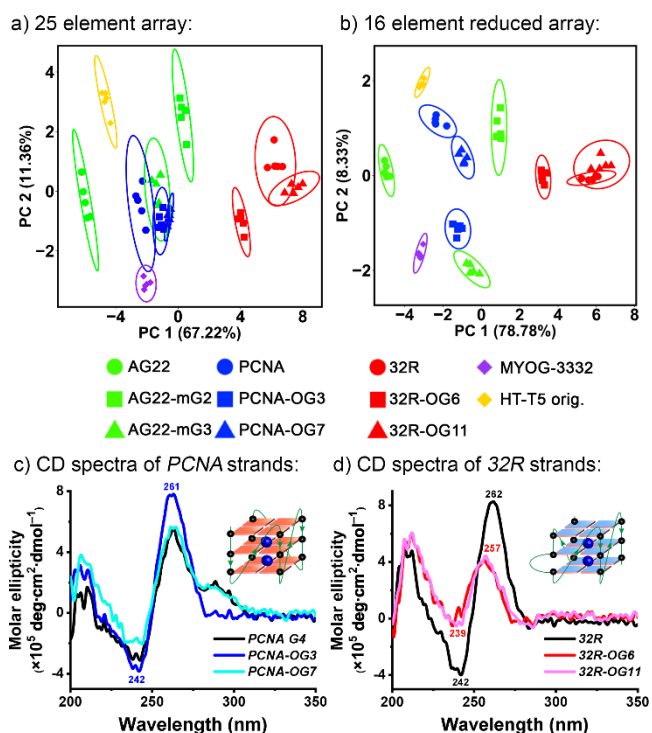


Figure 2. PCA scores plots with a) a 25-element array; b) manually optimized 16-element array. Conditions: [DNA] = 0.1 μ M, 10mM potassium phosphate buffer, 1 mM EDTA, pH 7.4, [host] = [dye] = 0.625 μ M, ellipses at 95% confidence interval. CD spectra of c) parent and modified **PCNA** strands; d) parent and modified **32R** strands.

This rudimentary optimization was almost successful, and as can be seen in Figure 2b, complete discrimination is possible for 9 of the 11 DNAs tested, using 95% confidence

ellipses. For the unmodified strands (*AG22*, *PCNA*, *32R*, *HT-T5 original*, and *MYOG-3332*) the array can still differentiate their G4 topology, even distinguishing the perfect G4s and that with a vacancy (*MYOG-3332*). The two strands that could not be discriminated were *32R* and *32R-OG11*, which is unsurprising when the structural similarities of the DNA targets are considered. We performed CD analysis on the strands to confirm the literature topology assignments, and also conducted CD in the presence of cavitaand and dye, to confirm that these additives did not alter DNA folding (see Figure 2c,d and Supporting Information). The structural changes of the modified DNAs are, in most cases, vanishingly small. The 22nt *AG22* strands show the largest structural change upon modification (in this case, with 6-mG), as guanine methylation at the O⁶ position weakens its H-bonding and K⁺ coordination.^{12,16} CD confirms that the parent *AG22* is a hybrid G4, but *AG22-mG2* switches to an antiparallel topology and *AG22-mG3* to a parallel G4. In contrast, introducing 8-oxoG modifications to *PCNA* have variable, and in some cases quite subtle effects on the structure. *PCNA-OG3* contains the modification in a G-quartet, but *PCNA-OG7* has the modification in an external loop, and in each case, the folding G4 topology remains parallel. As can be seen in Figure 2c, the changes in CD spectrum between the three strands are almost imperceptible, with only a slightly increased intensity seen for *PCNA-OG3* at 261 nm. The base modifications on *32R* are both on external loops, which should not induce significant topology or stability changes. As such, *32R-OG6* and *32R-OG11* have identical topologies, cannot be distinguished in any way by CD (Figure 2d), and are only slightly different than the parent *32R*.

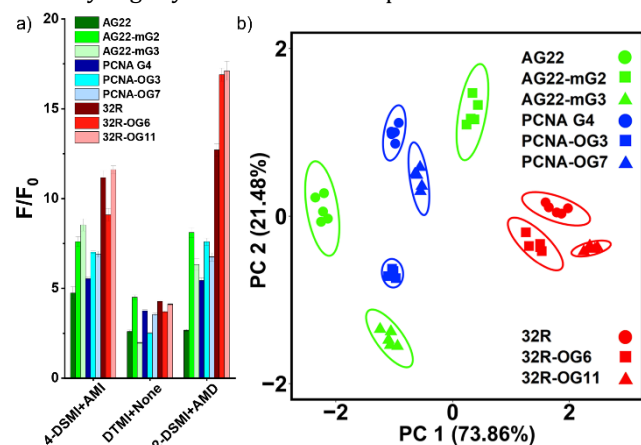


Figure 3. Optimized sensing. a) F/F_0 bar plot of 9 G4s measured with the optimized 3-element array: **4-DSMI+AMI**, **DTMI+None**, and **2-DSMI+AMD**. b) PCA scores plot obtained from the data shown in a), with ellipses at 95% confidence interval. Conditions identical to Fig. 2.

Despite these structural similarities, the minimally modified 16-element screen is quite (if not completely) effective, but also needs further optimization. The challenge is that identifying the optimal sensing elements can be difficult in a large pool. Manual elimination of elements showing low or repetitive contributions after analysis of the loading vectors from PCA was attempted, but. This process was time-consuming and gave unsatisfactory results (Figure S27-33). For more accurate optimization, we turned to machine learning (see Supporting Information).^{11b} To achieve this, the array data of 9 DNA (for the three sets of modified DNAs,

with the two “control” strands removed) was treated with SVM-RFECV (support vector machine, recursive feature elimination cross validation).¹⁷ SVM-RFECV adapts the linear SVM-RFE with a stratified-N-fold cross-validation strategy to find the minimum features set on which the SVM can provide the highest accuracy.^{17c} As such, it is more effective at feature optimization, showing greater robustness and accuracy when compared to manual selection from PCA vectors. After standardization of the F/F_0 data, SVM-RFECV selected the optimal number of features that can fully discriminate each DNA, followed by cross validation. Only 3 features are necessary to differentiate these 9 DNAs: **4-DSMI + AMI**, **DTMI alone**, and **2-DSMI + AMD** (Figure 3a). The PCA plot using these three elements (Figure 3b) shows that all 9 DNA can be separated with no overlap between any of the 95% confidence ellipses, even those that were indistinguishable by CD (i.e. *32R-OG6* and *32R-OG11*, or *PCNA* and *PCNA-OG3*), or those that overlapped significantly in the previous PCA plots (i.e. *32R* and *32R-OG11*). Notably, this simple, 3-element sensor array allows discrimination between 32-nt DNA G4 strands that differ by only 2 atoms (i.e. CH vs. C=O), outside the G4 quartet!

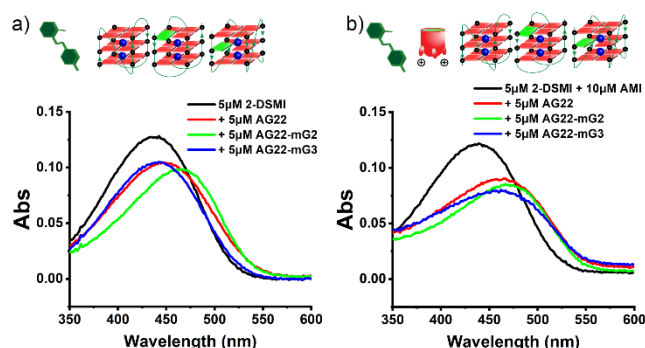


Figure 4. UV/Vis absorption spectra of a) **2-DSMI + AG22** series DNAs, and b) **2-DSMI+AMI** complex + AG22 series DNA. [2-DSMI] = 5 μM, [AMI] = 10 μM, [DNA] = 5 μM, potassium phosphate buffer.

The sensing mechanism appears consistent with that previously shown,¹¹ involving multiple binding states between the hosts, dyes and DNA. The dyes bind to the DNA, and the response is modulated by both competitive displacement by the cavitaands and the formation of **host•dye•DNA** heteroternary complexes, which provide an orthogonal sensing mechanism to the dyes alone. This is illustrated by correlation analysis: the non-cavitaand element **DTMI alone** has a low correlation coefficient (≤ 0.5) with the other two elements, while **2-DSMI + AMD** and **4-DSMI + AMI** are much more closely correlated (0.85). To better understand how this minimal array functions, we examined the individual binding profiles of the dyes with their competitive targets, namely cavitaands and the G4 DNA, (focusing on the *AG22* series, see Supporting Information). At dye concentrations < 5 μM, the **2-DSMI** and **DTMI** exhibit a 1:1 binding ratio with the G4 DNA (in the absence of cavitaand), as confirmed by Job plot analysis. The other dye:DNA affinities are quite similar, on the order of 2 μM. More detailed analysis of **DTMI** binding to the *AG22* series shows that the introduction of a 6-mG modification does not change the binding affinity appreciably - K_d (**DTMI**) for *AG22*, *AG22-mG2* and *AG22-mG3* were all between 1.9 and 2.5 μM. The most notable change was the fluorescence response F/F_0 , which was ~1.6-fold higher for

AG22-*mG2* than AG22 (Figure S-16). This modification occurs in the upper G-quartet, and effects a topology shift from hybrid to antiparallel suggesting that the reorganization of the G4 upon guanine methylation could provide a more hydrophobic environment for the dye, increasing emission. The affinities of the **2-DSMI** and **DTMI** dyes for the three cationic hosts (**CHI**, **AMI**, and **AMD**) are generally similar, with K_d in the range of 6–20 μM (see Supporting Information), which overall is slightly weaker than the dye:DNA affinities. The overall trend is that the different cavitations all display slightly different affinities to the dyes within a set range, which allows competition with the dye-DNA binding and introduces variables for differential analysis.

To further interrogate the effect of cavitand on the response, UV-Vis absorbance spectra were acquired, focusing on the **2-DSMI/AMI** interaction with the AG22 strands (Figure 4), and the 32R strands (see Supporting Information). A significant red shift ($\Delta\lambda = 25\text{ nm}$) of the peak absorbance occurred when the dye was mixed with AG22-*mG2*, but was not observed with the other two AG22 strands. Interestingly, in the presence of **AMI**, similar levels of red shift ($\Delta\lambda = 23\text{--}28\text{ nm}$) were observed with all AG22 strands, indicating that this cavitand host may strengthen the dye-DNA interaction.

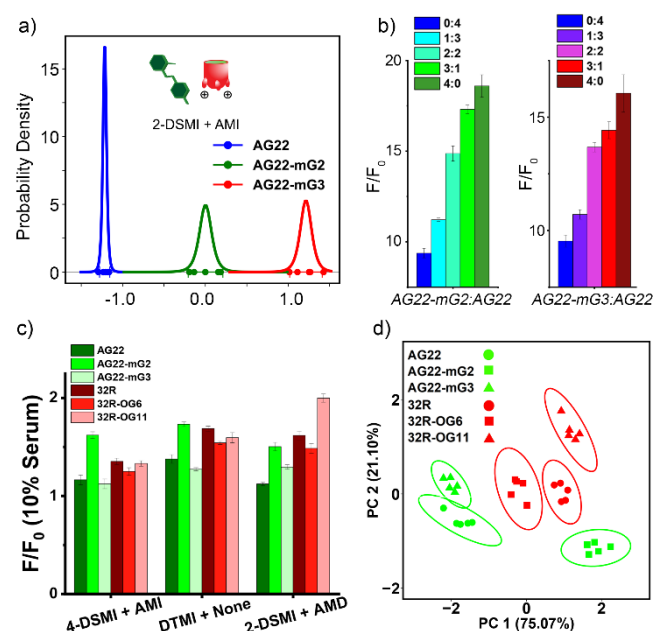


Figure 5. More Complex Sensing. a) t-distribution curve of the AG22 DNAs using the **2-DSMI+AMI** element, [Dye] = [Host] = 0.625 μM , [DNA] = 0.1 μM , potassium phosphate buffer, vertical markers = 95% confidence intervals. b) F/F_0 plots of two mixtures of the AG22 DNAs (total [DNA] = 0.5 μM), using the **2-DSMI+AMI** element. c) F/F_0 plot and d) PCA scores plot for discrimination of AG22 and 32R series in 10% serum, using **2-DSMI+AMD**, **DTMI+None**, and **4-DSMI+AMI**. [Dye] = [Host] = 0.625 μM , [DNA] = 0.1 μM , 10% serum in potassium phosphate buffer. Ellipses at 95% confidence interval.

To illustrate the potential of this system, it was tested in more stringent conditions, such as detecting base modifications in binary mixtures of DNA strands, and sensing in bio-media, namely diluted human serum (Figure 5). Detecting changes in modification in DNA mixtures is simplest (and most attractive) when a single sensor component is used, so we again used SVM-RFECV to identify the optimal sensor for

this purpose. The **2-DSMI+AMI** element was identified as optimal in all performance metrics for the AG22 series (Table S-9). Figure 5a shows the power of the **2-DSMI+AMI** element, whereby it completely discriminates the three AG22 strands (*via* t-distribution test). In addition, upon mixing one of the modified AG22 strands with parent AG22, (keeping total [DNA] = 0.5 μM , Figure 5b), the fluorescence gradually increased with increasing modified AG22-*mG2/3* strand concentration.

The most challenging test is to see whether differentiation of modified DNA strands is possible in diluted serum, and, as can be seen in Figure 5c/d, the sensor array performs quite well. Using the minimized 3-element array (**4-DSMI+AMI**, **DTMI + None**, **2-DSMI + AMD**), each of the 6 DNA strands in the AG22 and 32R series can be fully differentiated in 10% serum. We were unable to effect complete separation of the PCNA series, and the presence of serum causes quite high background fluorescence likely due to cell-free DNA or serum proteins. Nevertheless, the performance is impressive, illustrating the selectivity of the sensor for small differences in DNA structure.

In conclusion, we have shown that machine-learning-optimized arrayed cavitand:dye complexes can distinguish DNA G4 structures containing either oxidation or methylation modifications on individual guanine bases in the strand. The modifications can make subtle changes to the G4 folding that are undetectable by CD, but induce differential fluorescence responses in the array. The array is functional in diluted serum, and can recognize the presence of the modified guanines in DNA mixtures using only a single array element. Machine learning tools are a powerful method of optimizing sensor arrays, and in this case, they illustrate the necessity of careful array design: combining variably structured hosts and dyes can provide orthogonal sensing mechanisms, and judicious element choice can reduce the array size, but inclusion of non-effective sensors can harm the differentiation effect.

ASSOCIATED CONTENT

Supporting Information

DNA sequences and characterization, characterization of new molecules, fluorescence spectra, UV-Vis absorption spectra and full array data for analysis of the various DNA pools. This material is available free of charge via the Internet at <http://pubs.acs.org>.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (CHE-1707347 to W.Z., R.J.H.) for funding, and Adam D. Gill for initial dye synthesis.

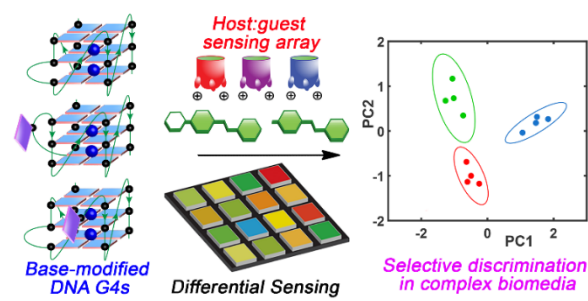
REFERENCES

- (1) (a) Chen, K.; Zhao, B. S.; He, C. Nucleic acid modifications in regulation of gene expression. *Cell Chem. Biol.* **2016**, *23*, 74–85. (b) Chen, Y.; Hong, T.; Wang, S.; Mo, J.; Tian, T.; Zhou, X. Epigenetic modification of nucleic acids: from basic studies to medical applications. *Chem. Soc. Rev.* **2017**, *46*, 2844–2872.
- (2) Luo, C.; Hajkova, P.; Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science* **2018**, *361*, 1336–1340. (b) Portela, A.; Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **2010**, *28*, 1057–1068. (c) Carell, T.; Kurz, M.

- Q.; Muller, M.; Rossa, M.; Spada, F. Non-canonical bases in the genome: The regulatory information layer in DNA. *Angew. Chem., Int. Ed.* **2018**, *57*, 4296–4312. (d) Wang, G.; Vasquez, K. M. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair* **2014**, *19*, 143–151.
- (3) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- (4) (a) Steenken, S.; Jovanovic, S.V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* **1997**, *119*, 617–618. (b) Szalai, V. A.; Singer, M. J.; Thorp, H. H. Site-specific probing of oxidative reactivity and telomerase function using 7,8-dihydro-8-oxoguanine in telomeric DNA. *J. Am. Chem. Soc.* **2002**, *124*, 1625–1631. (c) Fleming, A. M.; Burrows, C. J. G-quadruplex folds of the human telomere sequence alter the site reactivity and reaction pathway of guanine oxidation compared to duplex DNA. *Chem. Res. Toxicol.* **2013**, *26*, 593–607. (d) Fleming, A. M.; Ding, Y.; Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 2604–2609. (e) Lee, H.-T.; Bose, A.; Lee, C.-Y.; Opreko, P. L.; Myong, S. Molecular mechanisms by which oxidative DNA damage promotes telomerase activity. *Nucleic Acids Res.* **2017**, *45*, 11752–11765.
- (5) (a) Winnerdy, F. R.; Bakalar, B.; Maity, A.; Vandana, J. J.; Mechulam, Y.; Schmitt, E.; Phan, A. T. NMR solution and X-ray crystal structures of a DNA containing both right- and left-handed parallel-stranded G-quadruplexes *Nucleic Acids Res.* **2019**, *47*, 8272–8281. (b) Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **2015**, *6*, 3314–3320.
- (6) (a) Hofer, A.; Liu, Z. J.; Balasubramanian, S. Detection, structure and function of modified DNA bases. *J. Am. Chem. Soc.* **2019**, *141*, 6420–6429. (b) Heck, C.; Michaeli, Y.; Bald, I.; Ebenstein, Y. Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. *Curr. Opin. Biotechnol.* **2019**, *55*, 151–158. (c) Yuan, B.-F. Assessment of DNA epigenetic modifications. *Chem. Res. Toxicol.* **2020**, *33*, 695–708.
- (7) del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex secondary structure obtained from Circular Dichroism spectroscopy. *Angew. Chem. Int. Ed.* **2018**, *57*, 7171–7175.
- (8) (a) You, L.; Zha, D.; Anslyn, E.V. Recent advances in supramolecular analytical chemistry using optical sensing. *Chem. Rev.* **2015**, *115*, 7840–7892. (b) Stewart, S.; Ivy, M. A.; Anslyn, E. V. The use of principal component analysis and discriminant analysis in differential sensing routines. *Chem. Soc. Rev.* **2014**, *43*, 70–84.
- (9) (a) Zhang, T.; Edwards, N. Y.; Bonizzoni, M.; Anslyn, E. V. The Use of Differential Receptors to Pattern Peptide Phosphorylation. *J. Am. Chem. Soc.* **2009**, *131*, 11976–11984. (b) 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.* **2017**, *139*, 409–416. (c) Eubanks, C. S.; Zhao, B.; Patwardhan, N. N.; Thompson, R. D.; Zhang, Q.; Hargrove, A. E. Visualizing RNA conformational changes via pattern recognition of RNA by small molecules. *J. Am. Chem. Soc.* **2019**, *141*, 5692–5698. (d) Beatty, M. A.; Selinger, A. J.; Li, Y.; Hof, F. Parallel synthesis and screening of supramolecular chemosensors that achieve fluorescent turn-on detection of drugs in saliva. *J. Am. Chem. Soc.* **2019**, *141*, 16763–16771. (e) 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.* **2010**, *8*, 1033–1039. (f) Peacor, B. C.; Ramsay, C. M.; Waters, M. L. Fluorogenic sensor platform for the histone code using receptors from dynamic combinatorial libraries. *Chem. Sci.* **2017**, *8*, 1422–1428.
- (10) (a) 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.* **2018**, *46*, e41. (b) Zuffo, M.; Xie, X.; Granzhan, A. Strength in numbers: development of a fluorescence sensor array for secondary structures of DNA. *Chem.-Eur. J.* **2019**, *25*, 1821–1818.
- (11) (a) Chen, J.; Hickey, B. L.; Wang, L.; Lee, J.; Gill, A. D.; Favero, A.; Pinalli, R.; Dalcanele, E.; Hooley, R. J.; Zhong, W. Selective Discrimination and Classification of G-Quadruplex Structures with a Host:Guest Sensing Array. *Nat. Chem.* **2021**, *13*, 488–495. (b) Chen, J.; Gill, A.D.; Hickey, B. L.; Gao, Z.; Cui, X.; Hooley, R. J.; Zhong, W. Machine Learning Aids Classification and Discrimination of Non-canonical DNA Folding Motifs by an Arrayed Host:guest Sensing System. *J. Am. Chem. Soc.* **2021**, *143*, 12791–12799.
- (12) Mekmaysy, C. S.; Petraccone, L.; Garbett, N. C.; Ragazzon, P. A.; Gray, R.; Trent, J. O.; Chaires, J. B. Effect of O6-Methylguanine on the Stability of G-Quadruplex DNA. *J. Am. Chem. Soc.* **2008**, *130*, 6710–6711.
- (13) Redstone, S. C. J.; Fleming, A. M.; Burrows, C. J. Oxidative Modification of the Potential G-Quadruplex Sequence in the PCNA Gene Promoter Can Turn on Transcription. *Chem. Res. Toxicol.* **2019**, *32*, 437–446.
- (14) Cogoi, S.; Ferino, A.; Miglietta, G.; Pedersen, E. B.; Xodo, L. E. The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription. *Nucleic Acids Res.* **2017**, *46*, 661–676.
- (15) Hickey, B. L.; Chen, J.; Zou, Y.; Gill, A. D.; Zhong, W.; Millar, J. G.; Hooley, R. J. Enantioselective Sensing of Insect Pheromones in Water. *Chem. Commun.* **2021**, *57*, 13341–13344.
- (16) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **1989**, *59*, 871–880.
- (17) (a) Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297. (b) Ivanciuc, O. Applications of Support Vector Machines in chemistry. *Rev. Comput. Chem.* **2007**, 291–400. (c) Wang, C.; Xiao, Z.; Wu, J. Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data. *Phys. Med.* **2019**, *65*, 99–105.

For
Only:

TOC



Sensing Base Modifications in Non-Canonically Folded DNA with an Optimized Host:Guest Sensing Array

Junyi Chen,^a Briana L. Hickey,^b Ziting Gao,^b Alexie Andrea P. Raz,^b Richard J. Hooley^{b} and Wenwan Zhong^{a,b*}*

*^aEnvironmental Toxicology Graduate Program; ^bDepartment of Chemistry,
University of California - Riverside, Riverside, CA 92521, U.S.A.*

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

Supporting Information

Table of Contents

1. General Information	S-3
2. DNA Sequences and Characterization.....	S-6
2.1 DNA Sequences.....	S-6
2.2 Circular Dichroism (CD) Spectra for DNA Folding Confirmation.....	S-7
2.3 Gel Electrophoresis for DNA Quality Inspection.....	S-11
3. NMR Spectra of Components Used.....	S-13
4. Titration Curves and Binding Affinity Measurements	S-19
4.1 Fluorescence Titration of Dye-DNA.....	S-19
4.2 Job plots of Dye-DNA	S-20
4.3 Affinity Measurement of DNA-Dye	S-21
4.4 Affinity Measurement of Cavitand-Dye	S-22
5. Array Analysis for Differentiation of 11 DNAs.....	S-27
5.1 Bar Plots for Array Signals from 11 DNAs	S-27
5.2 PCA biplot for 25-element Array Signals from 11 DNAs	S-29
6. Array Analysis for Differentiation of 9 Core DNAs.....	S-31
6.1 PCA for 16-element Array Signals from 9 Core DNAs	S-31

6.2 SVM-RFECV Feature Selection for 9 Core DNAs Classification.....	S-36
6.3 Performance metrics of 9 Core DNAs Classification.....	S-37
7. Array Analysis for <i>AG22</i> Series DNAs.....	S-38
7.1 SVM-RFECV Feature Selection for <i>AG22</i> Series DNAs Classification	S-38
7.2 Performance metrics of <i>AG22</i> Series DNAs Classification.....	S-38
7.3 t-distribution Curve of <i>AG22</i> series DNAs	S-39
7.4 Bar Plots for Array Signals from <i>AG22</i> Series DNA Mixture	S-41
8. Array Analysis for DNA Sensing in Diluted Serum.....	S-42
8.1 Bar Plots of Sensor Element for <i>AG22</i> Series in Diluted Serum.....	S-42
8.2 Bar Plots for Array Signals from <i>AG22</i> and <i>32R</i> Series in 10% Serum.....	S-45
8.3 PCA plot of <i>AG22</i> and <i>32R</i> Series in 10% Serum.....	S-46
8.4 Repeats of 3-element Array Sensing in Different Serum.....	S-47
9. UV-Vis Absorption Spectra.....	S-51
9.1 UV-Vis Spectra of <i>32R</i> series with 2-DSMI+AMI.....	S-51
9.2 UV-Vis Spectra of <i>AG22</i> series with 2-DSMI+AMI	S-52
10. References.....	S-53

1. General Information.

Cavitands **TCC**,¹ **CHI**,² **AMI**,³ and fluorophore **PSMI**³ were synthesized according to literature procedures. ¹H and ¹³C spectra were recorded on a Bruker Avance NEO 400 MHz or Bruker Avance 600 MHz NMR spectrometer. The spectrometers were automatically tuned and matched to the correct operating frequencies. Proton (¹H) and carbon (¹³C) chemical shifts are reported in parts per million (δ) with respect to tetramethylsilane (TMS, $\delta=0$), and referenced internally. Deuterated NMR solvents were obtained from Cambridge Isotope Laboratories, Inc., Andover, MA, and used without further purification. All other materials, including **4-DSMI** (*trans*-4-[4-(dimethylamino)-styryl]-1-methyl-pyridinium iodide) 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 structural information of which are given in Table S-1. The concentrations of DNA stock solutions were determined by NanoDrop 2000 (Thermo Fisher Scientific) using the corresponding molar extinction coefficients provided by IDT after background subtraction. DNA stock solutions were diluted with in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4) and re-annealed to form the most stable folding topology, in which the DNA solutions were heated at 95 °C for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before the experiments. Single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) and pooled human serum off the clot (ISER50ML-36670) were purchased from Innovative Research. Fetal bovine serum (10437028) was purchased from Gibco. The serum was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer. Fluorescence measurements were performed with a BioTek Synergy H1 Hybrid Multi-Mode Microplate Reader at Fluorescence Endpoint or Spectral scanning read mode with the Ex/Em wavelengths at 480/600 nm (**4-DSMI**), 500/600 nm (**PSMI**), 480/580 nm (**2-DSMI**), 540/600 nm (**DTMI**), 440/580 nm (**2-SMIQ**), with default Gain value=100 unless specifically emphasized. UV-Vis absorbance measurements were performed with an Agilent Technologies Cary 60 UV/Vis spectrophotometer using Brandtech ultra-micro cuvettes (path length = 10 mm). 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). Feature selection and classification were

performed with Python 3.9 (64-bit), using StandardScaler for data standardization, Recursive feature elimination with cross-validation (RFECV) to select the number of features, Support Vector Machine (SVM) (kernel='linear') as the supervised classification estimator, RFECV(estimator=svm.SVC(kernel='linear'), step=1, cv=StratifiedKFold(n_splits=4, shuffle=True), scoring='accuracy', min_features_to_select=1). Performance metrics for the classification evaluation were calculated by using RepeatedStratifiedKFold (n_splits=4, n_repeats=3) for cross validation. The correlation heatmap of selected features was computed using pandas.DataFrame.corr(method='pearson'). PCA was applied for orthogonal linear transformation and dimensionality reduction, and SVM decision region boundary of PCA plot was generated using plot_decision_regions.

Fluorescence measurements.

1) Array constituents. The fluorescence assay was carried out by making the mixture solution of 0.625 μM fluorescent dye: **4-DSMI/PSMI/2-DSMI/DTMI/2-SMIQ**, 0.625 μM cavitand: **TCC/CHI/AMI/AMD** or no cavitand, 0.1 μM DNA in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). The mixture was incubated for 15 min at room temperature, then added in the 96-well plate with a volume = 100 μL before the fluorescence signal (F) was recorded. Fluorescence response (F/F_0) are normalized to the response of cavitand–dye in the absence of DNA, that is F_0 is defined as the fluorescence recorded for that concentration of host and guest when $[\text{DNA}] = 0 \mu\text{M}$.

2) Titrations. Dye-DNA Titration: The fluorescence titration curves were obtained by using 0-20 μM Dye and 0.1 μM DNA or no DNA. Fluorescence response (F/F_0) are normalized to the response of dye in the absence of DNA, that is F_0 is defined as the fluorescence recorded for that concentration of dye when $[\text{DNA}] = 0 \mu\text{M}$. **Binding Affinity Measurement:** Fluorescence response curves of dye upon titration of hosts were obtained by using 0.625 μM dye, 0-50/100 μM cavitand or 0-5/10 μM DNA. Fluorescence response (F/F_0) are normalized to the response of dye in the absence of cavitand, that is F_0 is defined as the fluorescence recorded for that concentration of dye when $[\text{Host}]$ or $[\text{DNA}] = 0 \mu\text{M}$. The binding affinities were achieved by fitting F/F_0 data by Hill 1 function in Growth/Sigmoidal category from Origin software.

Hill 1 function is a more general form of Hill function. The equation is:

$$y = \text{START} + (\text{END} - \text{START}) \frac{x^n}{k^n + x^n}$$

x is the ligand concentration, k is half-maximal concentration constant, n is Hill coefficient.

3) Job Plot. Job plots were obtained by measuring the fluorescence intensity of solutions containing dye and DNA at the total concentration of 5 μM with the dye mole fraction (X_{Dye}) varied between 0 and 1.

UV-Vis Absorbance Spectra.

Absorbance spectra were obtained with solutions containing 5 μM dye/5 μM DNA/10 μM cavitand, in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). The spectra were presented with baseline-correction in which the background signal from the buffer was subtracted.

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 K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4) 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 gradient native polyacrylamide gel electrophoresis (PAGE) gel (4%-20%). 5 μL of a 2 μM DNA solution in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4) with or without diluted 15% serum was loaded to the gel. DNA was denatured at 95°C for 5 min, cooled on ice for 10 min and then at room temperature for 30 min before loading. The gel was run at 120 V for 60 min at room temperature in 1×TBE buffer, and stained with SYBR Gold (1.5:10000 dilution) before imaged using the UV transilluminator (SPECTROLINE).

2. DNA Sequences and Characterization

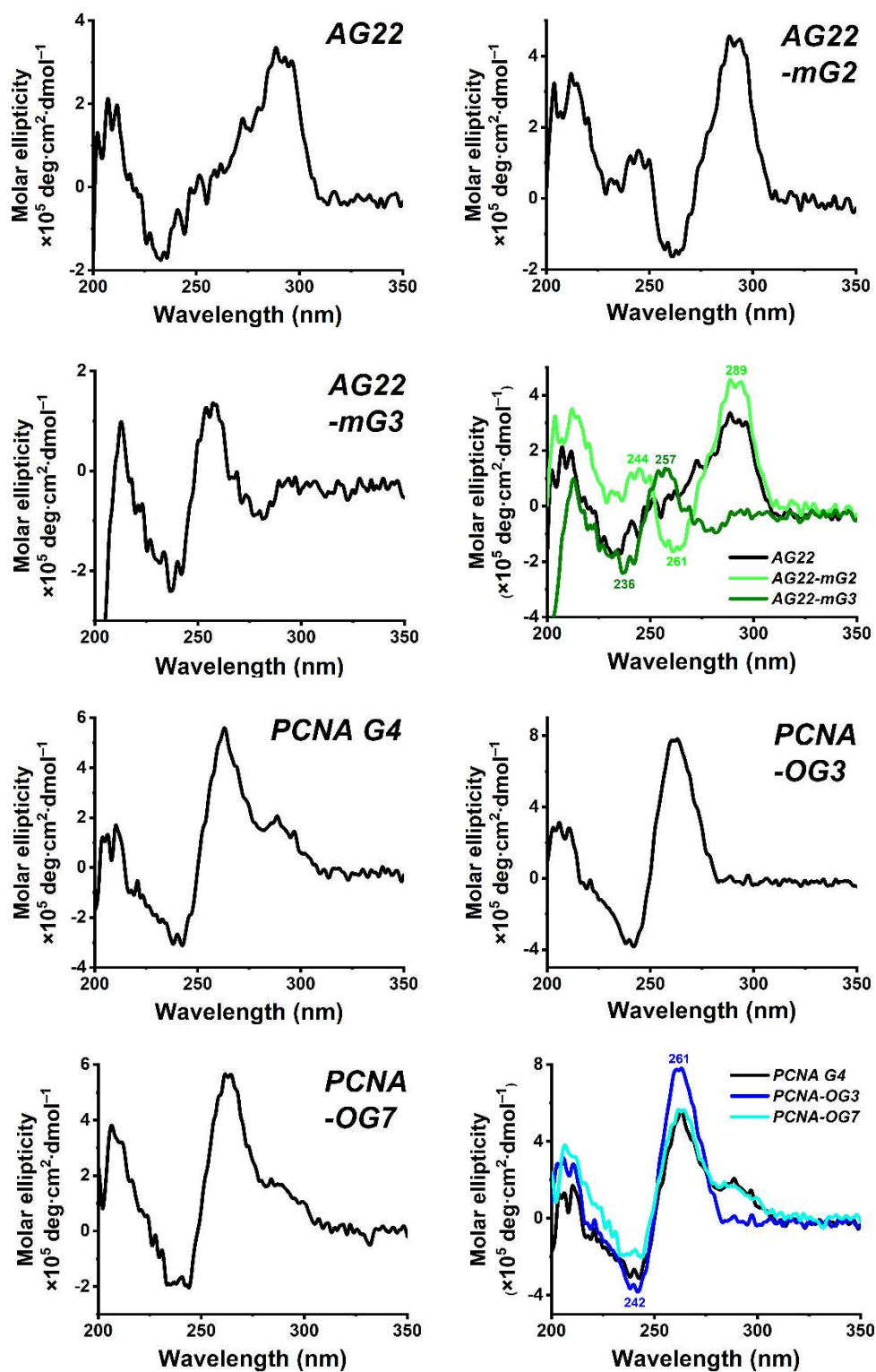
2.1 DNA Sequences

Table S-1. DNA sequences used in this project.

Group	Name	Sequence	G4 Type in Ref.	G4 Type in Exp.	Bases	Ref
Group1 6mG=O ⁶ - Methylguanine	<i>AG22</i>	AGGGTTAGGGTTAGGGT TAGGG	Na ⁺ : Antiparallel K ⁺ : Hybrid	Hybrid	22	4
	<i>AG22-mG2</i>	A-6mG- GGTTAGGGTTAGGGTTA GGG	Na ⁺ : Antiparallel K ⁺ : Antiparallel	Antiparallel	22	
	<i>AG22-mG3</i>	AG-6mG- GTTAGGGTTAGGGTTAG GG	Na ⁺ : Antiparallel K ⁺ : Hybrid	Parallel	22	
Group2 ^{ox} G=8- oxoguanine	<i>PCNA G4</i>	CAGGGCGACGGGGGCG GGGCGGGGCG	Parallel	Parallel with a small Hybrid peak	26	5
	<i>PCNA-OG3</i>	CA ^{ox} GGGCGACGGGGGCG GGGCGGGGCG	Parallel, stability↓	Parallel (Higher than original)	26	
	<i>PCNA-OG7</i>	CAGGGC ^{ox} GACGGGGGCG GGGCGGGGCG	Parallel	Parallel with a small Hybrid peak	26	
Group3 ^{ox} G=8- oxoguanine	<i>32R</i>	AGGGCGGTGTGGGAAGA GGGAAGAGGGGGAGG	Parallel	Parallel	32	6
	<i>32R-OG6</i>	AGGGC ^{ox} GGTGTGGGAAG AGGGAAGAGGGGGAGG	Parallel with a small hybrid peak, stability↓	Parallel, lower than original & blue shift	32	
	<i>32R-OG11</i>	AGGGCGGTGT ^{ox} GGGAAG AGGGAAGAGGGGGAGG	Parallel, stability↓	Parallel, lower than original & blue shift	32	
Control	MYOG-3332	AGGGTGGGCTGGGAGGT	Parallel	Antiparallel	17	7
	HT-T5 original	TTGGGTAGGGTTAGGG TTAGGGA	Hybrid	Hybrid	24	8

2.2 CD Spectra for DNA Folding Confirmation

2.2.1 DNA in K⁺ buffer



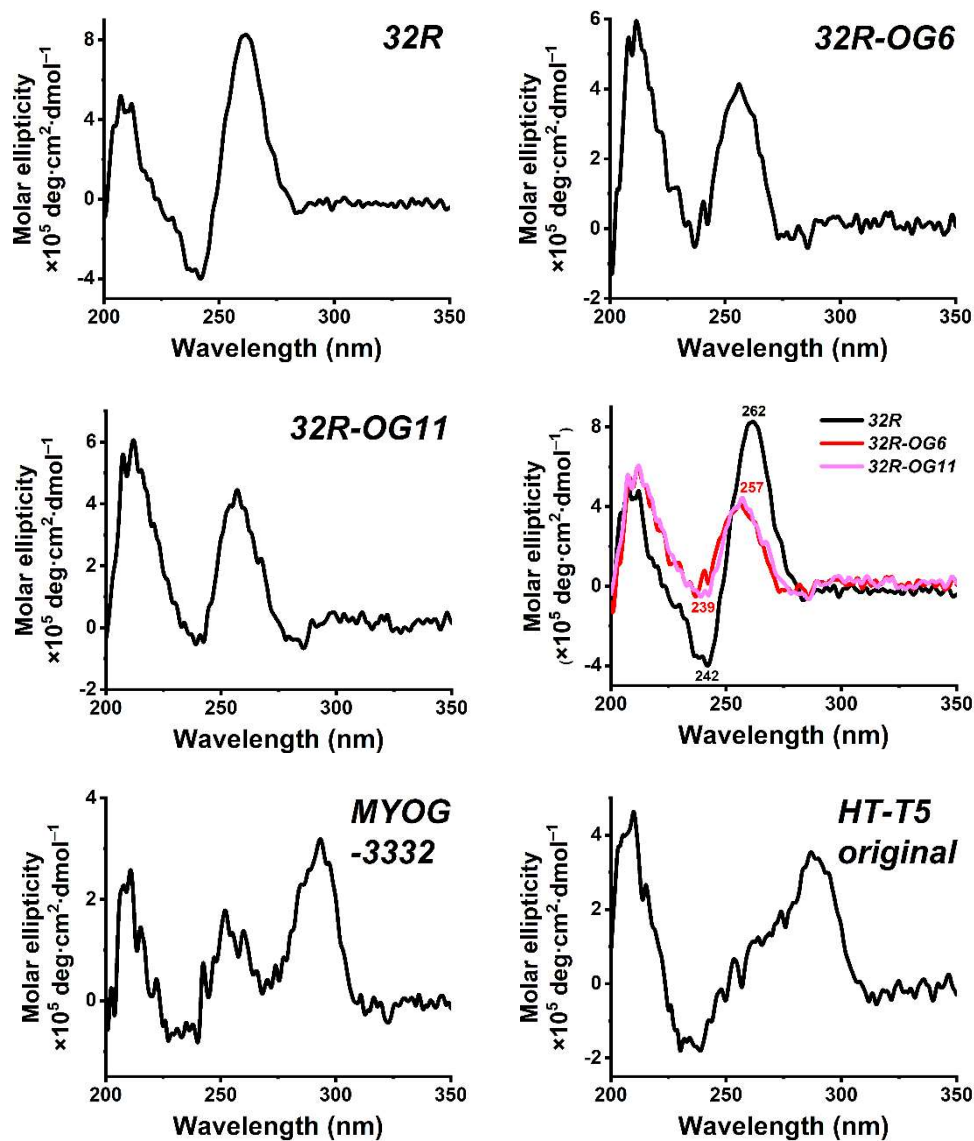


Figure S-1. CD spectra with baseline correction of 10 μM G4 DNA in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). DNA was denatured at 95 $^\circ\text{C}$ for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before the experiment.

2.3.2 DNA 32R with Dye 2-DSMI and Cavitand AMI in K⁺ buffer

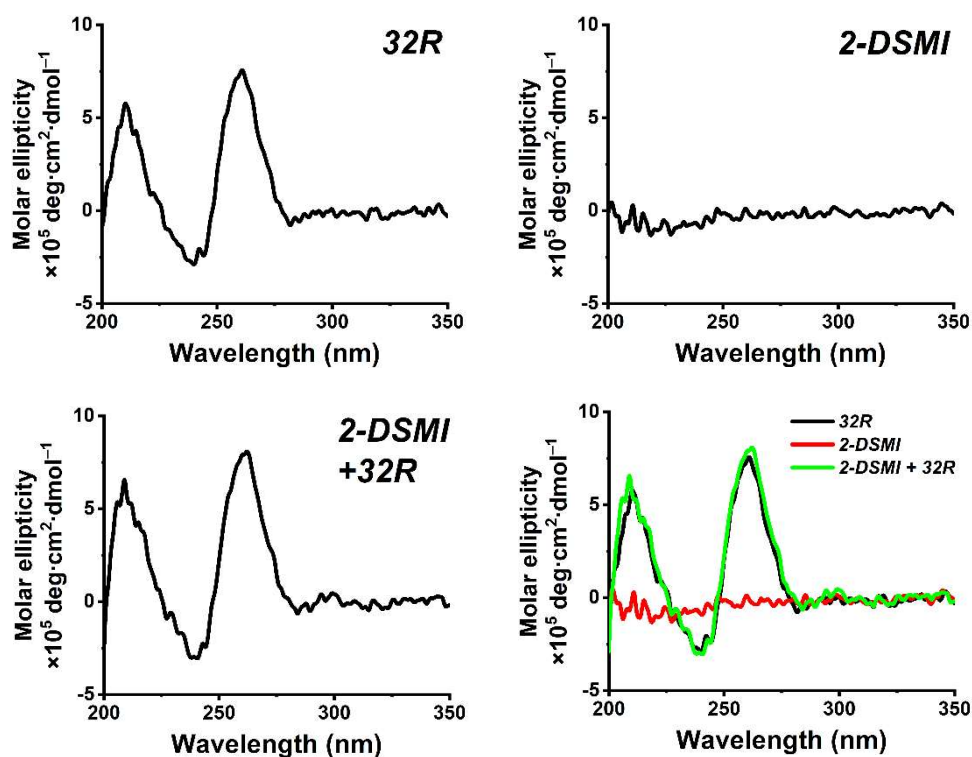


Figure S-2. CD spectra with baseline correction of 10 μM 32R DNA, 10 μM 2-DSMI, and 10 μM 2-DSMI + 10 μM 32R in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4). DNA was denatured at 95 °C for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before the experiment.

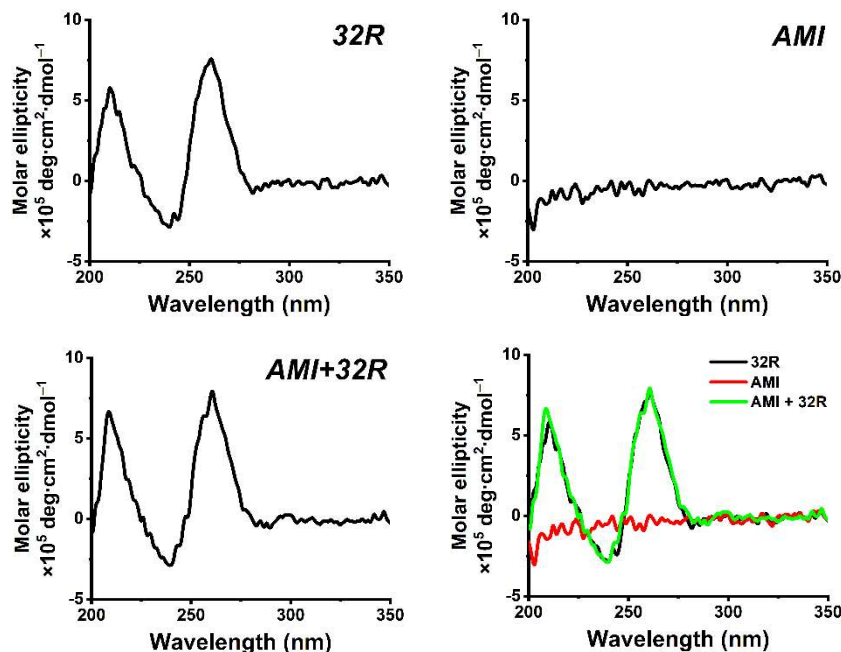


Figure S-3. CD spectra with baseline correction of 10 μM 32R DNA, 10 μM AMI, and 10 μM AMI + 10 μM 32R in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). DNA was denatured at 95 $^\circ\text{C}$ for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before the experiment.

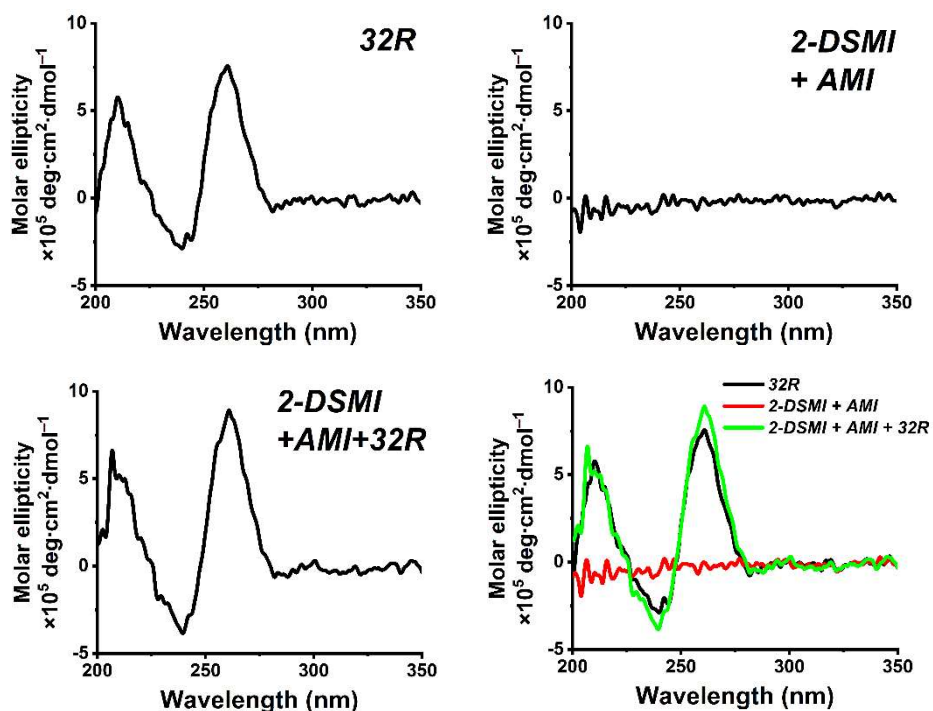


Figure S-4. CD spectra with baseline correction of 10 μM 32R DNA, 10 μM 2-DSMI+AMI, and 10 μM 2-DSMI+AMI + 10 μM 32R in K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). DNA was denatured at 95 $^\circ\text{C}$ for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before the experiment.

2.3 Gel Electrophoresis for DNA Quality Inspection

2.3.1 DNA in K⁺ buffer

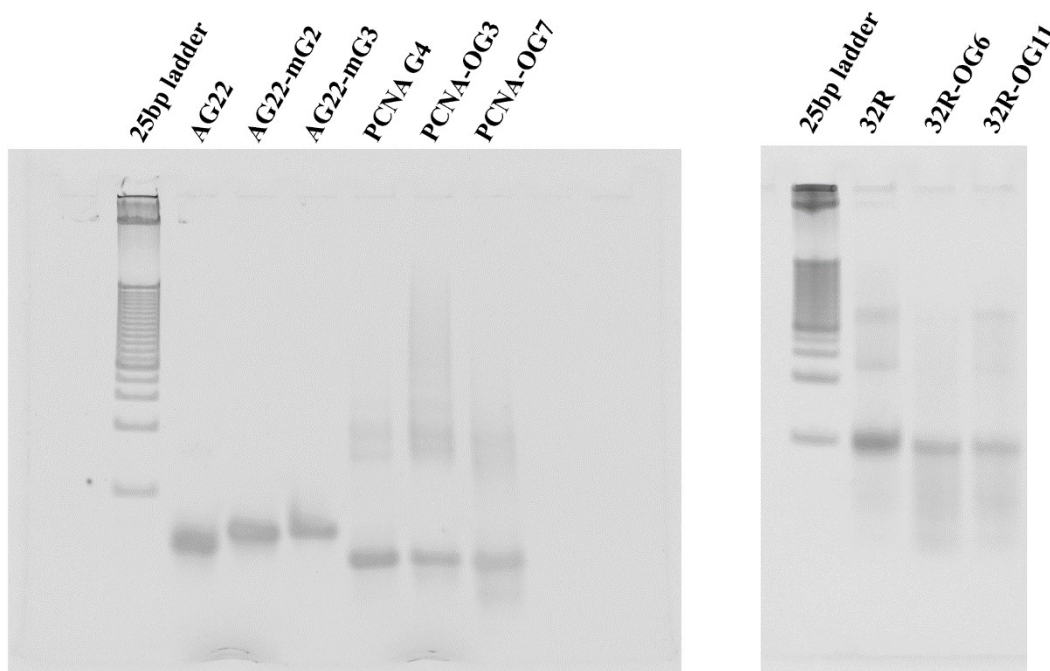


Figure S-5. The gradient native polyacrylamide gel electrophoresis (PAGE) gel (4%-20%) results of DNA sequences. The gel was loaded with 5 μ l of a 2 μ M DNA dissolved in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4), which had been denatured at 95 °C for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min. The gel was run at 120 V for 60 min at room temperature in 1 $\times$ TBE buffer and stained with SYBR Gold (1.5:10000 dilution) before imaging using a UV transilluminator (SPECTROLINE).

2.3.2 AG22 series incubated with 15% serum in K⁺ buffer for 3 h

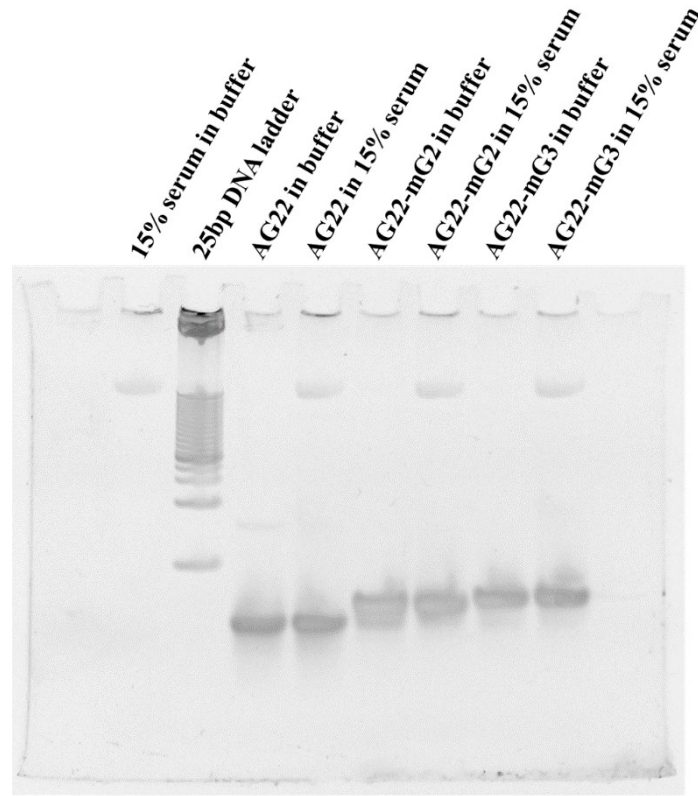


Figure S-6. The gradient native polyacrylamide gel electrophoresis (PAGE) gel (4%-20%) results of DNA sequences in 15% serum. The gel was loaded with 5 μ l of a 2 μ M DNA incubated in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4) with or without 15% serum for 3 h. DNA had been denatured at 95 °C for 5 min, cooled on ice for 10 min and then held at room temperature for 30 min before loading. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer. The gel was run at 120 V for 60 min at room temperature in 1 $\times$ TBE buffer and stained with SYBR Gold (1.5:10000 dilution) before imaging using a UV transilluminator (SPECTROLINE).

3. NMR Spectra of Components Used

Synthesis of New Molecules

(E)-2-(4-(dimethylamino)styryl)-1-methylpyridin-1-ium iodide (2-DSMI):

1,2-Dimethylpyridinium iodide (235 mg, 1.00 mmol) and 4-dimethylaminobenzaldehyde (149 mg, 1.00 mmol) were dissolved in ethanol (5 mL) inside a round bottom flask. While stirring, one drop of piperidine was added and the resulting solution was refluxed for 12 hours. The reaction was cooled, then diluted with water (10 mL). The resulting precipitate was filtered, rinsed with water and cold ethanol, then dried under vacuum to yield (E)-2-(4-(dimethylamino)styryl)-1-methylpyridin-1-ium iodide (343 mg, 94% yield) as a bright red powder. ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 8.75 (dd, $J = 6.4, 1.5$ Hz, 1H), 8.44 (dd, $J = 8.6, 1.5$ Hz, 1H), 8.34 (td, $J = 8.4, 1.4$ Hz, 1H), 7.92 (d, $J = 15.7$ Hz, 1H), 7.70 (m, 3H), 7.24 (d, $J = 15.7$ Hz, 1H), 6.79 (d, $J = 8.9$ Hz, 2H), 4.29 (s, 3H), 3.04 (s, 6H). ^{13}C NMR (100 MHz, $\text{DMSO}-d_6$) δ 153.21, 152.10, 145.27, 144.14, 143.06, 130.70, 123.53, 123.00, 122.25, 111.75, 110.57, 45.64, 40.12. (ESI-MS: m/z $\text{C}_{16}\text{H}_{19}\text{N}_2^+$ calculated: 239.3349, found: $(\text{M})^+$ 239.1548. UV/Vis: Exc. $\lambda_{\text{max}} = 440$ nm, Em. $\lambda_{\text{max}} = 585$ nm.

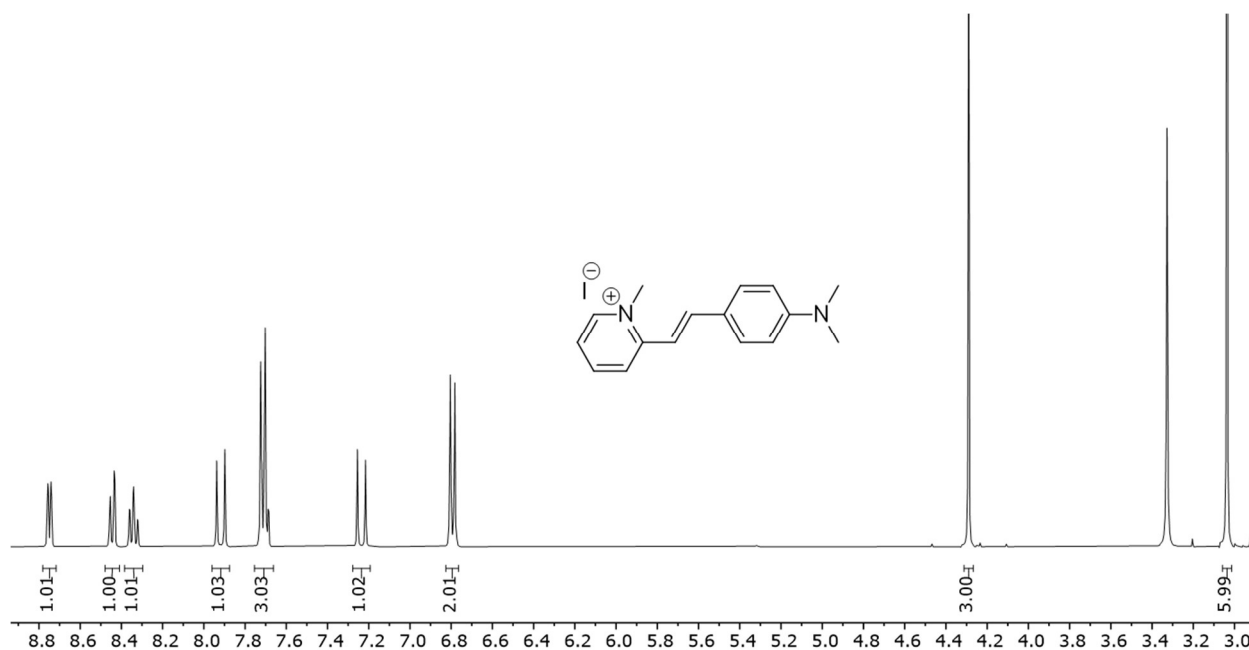


Figure S-7. ^1H NMR spectrum of **2-DSMI** (400 MHz, $\text{DMSO}-d_6$, 298K).

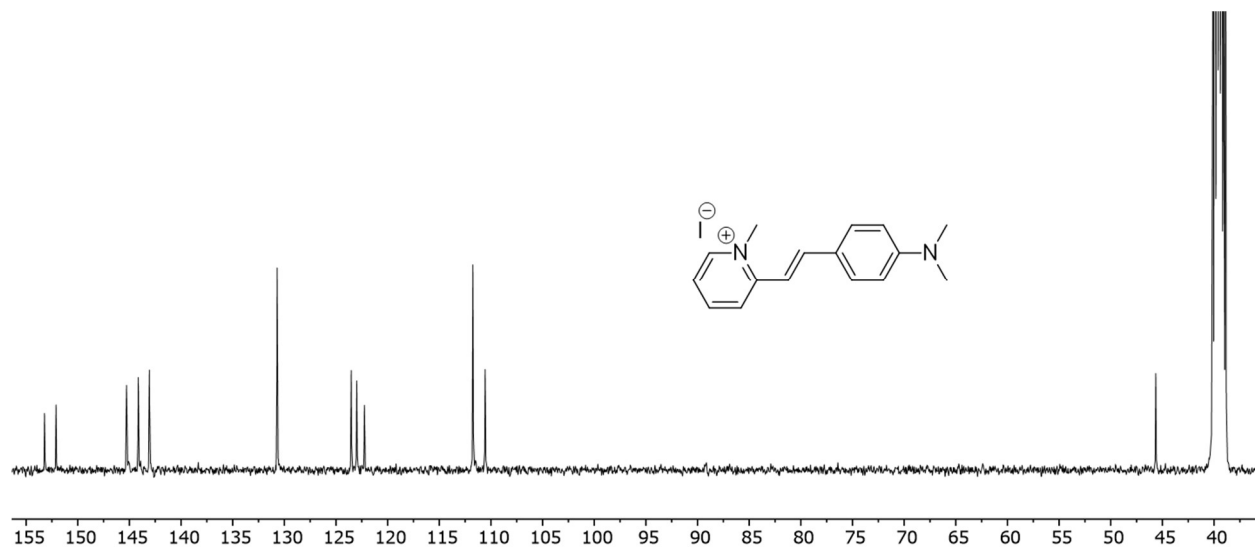


Figure S-8. ^{13}C NMR of **2-DSMI** (100 MHz, $\text{DMSO-}d_6$, 298K).

(E)-2-(4-(dimethylamino)styryl)-3-methylbenzothiazol-3-ium iodide (DTMI):

2-methylbenzothiazole (200 μL , 1.60 mmol) was dissolved in ethanol (5 mL), iodomethane (1 mL) was added to the reaction mixture while stirring and the reaction was refluxed for 12 hours. The solution was diluted with diethyl ether (10 mL) and the resulting precipitate was filtered, then rinsed with diethyl ether and dried under vacuum to yield 2,3-dimethylbenzothiazol-3-ium iodide (398 mg, 87%) as a white solid. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 8.43 (dd, $J = 8.0, 1.3$ Hz, 1H), 8.29 (dt, $J = 8.5, 1.2$ Hz, 1H), 7.90 (dt, $J = 8.5, 1.3$ Hz, 1H), 7.81 (dd, $J = 8.0, 1.2$ Hz, 1H), 4.20 (s, 3H), 3.17 (s, 3H).

2,3-dimethylbenzothiazol-3-ium iodide (290mg, 1.00 mmol) and 4-(dimethylamino)benzaldehyde (149 mg, 1.00 mmol) were dissolved in ethanol (5 mL) inside a round bottom flask. While stirring, one drop of piperidine was added and the resulting solution was refluxed for 12 hours. The reaction was cooled, then diluted with water (10 mL). The resulting precipitate was filtered, rinsed with water and cold ethanol, then dried under vacuum to yield (E)-2-(4-(dimethylamino)styryl)-3-methylbenzothiazol-3-ium iodide (386 mg, 92% yield) as a dark purple powder. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 8.32 (dd, $J = 7.9, 1.5$ Hz, 1H), 8.11 (dd, $J = 7.7, 1.1$, 1H), 8.09 (d, $J = 15.3$, 1H), 7.93 (d, $J = 8.8$ Hz, 2H), 7.80 (td, $J = 7.9, 1.1$ Hz, 1H), 7.70 (td, $J = 7.7, 1.5$ Hz, 1H), 7.65 (d,

$J = 15.3$ Hz, 1H), 6.87 (d, $J = 8.8$ Hz, 2H), 4.24 (s, 3H), 3.13 (s, 6H). ^{13}C NMR (100 MHz, DMSO- d_6) δ 150.61, 142.45, 133.31, 129.36, 127.94, 127.29, 124.28, 121.95, 116.43, 112.46, 106.76, 40.51, 36.01. ESI-MS: m/z $\text{C}_{18}\text{H}_{19}\text{N}_2\text{S}^+$ calculated: 295.4213, found: (M) $^+$ 295.1270. UV/Vis: Exc. $\lambda_{\text{max}} = 510$ nm, Em. $\lambda_{\text{max}} = 600$ nm.

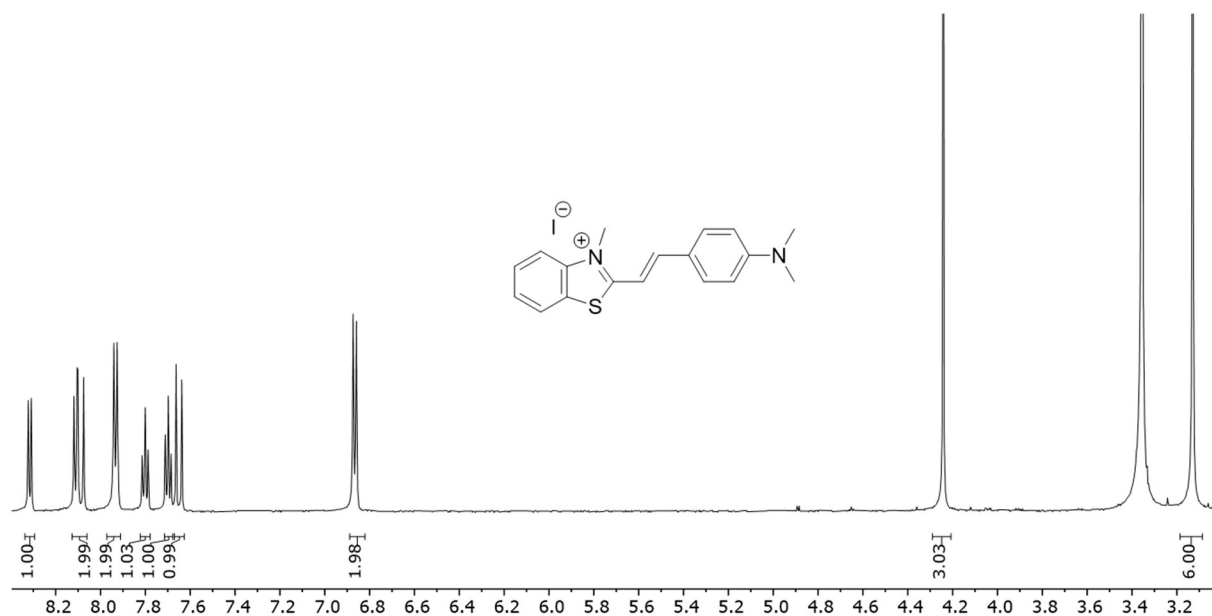


Figure S-9. ^1H NMR spectrum of **DTMI** (400 MHz, DMSO- d_6 , 298K).

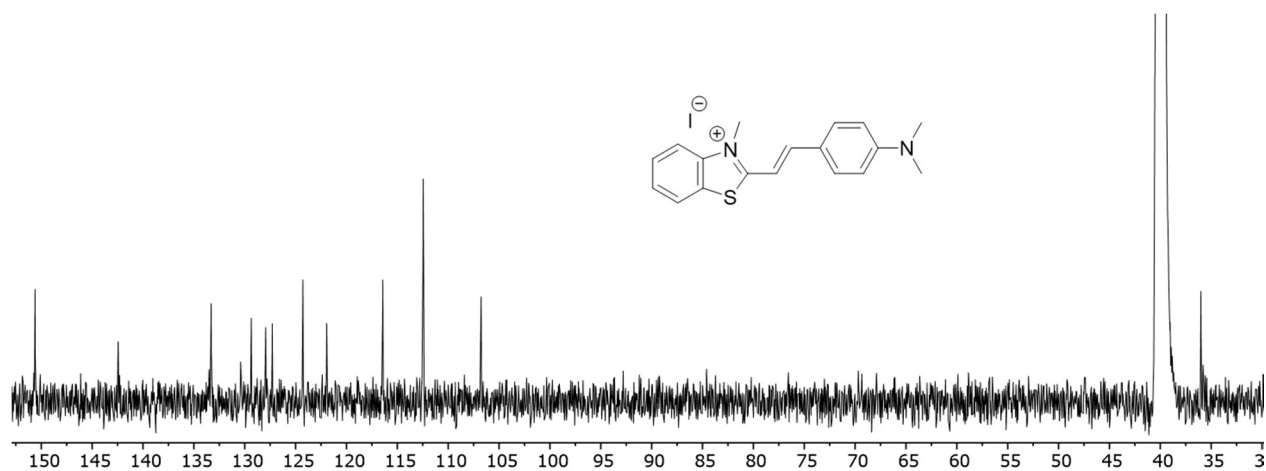


Figure S-10. ^{13}C NMR of **DTMI** (100 MHz, DMSO- d_6 , 298K).

(E)-1-methyl-2-(4-(methylthio)styryl)quinolin-1-ium iodide (2-SMIQ):

2-methylquinoline (250 μ L, 1.88 mmol) was dissolved in ethanol (3 mL), iodomethane (0.5 mL) was added to the reaction mixture while stirring and the reaction was refluxed for 12 hours. The solution was diluted with diethyl ether (10 mL) and the resulting precipitate was filtered, then rinsed with diethyl ether and dried under vacuum to yield 1,2-dimethylquinolin-1-ium iodide (457 mg, 85%) as a yellow solid. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.09 (d, J = 8.5 Hz, 1H), 8.59 (d, J = 9.1 Hz, 1H), 8.40 (d, J = 8.0 Hz, 1H), 8.23 (t, J = 8.0 Hz, 1H), 8.12 (d, J = 8.6 Hz, 1H), 7.99 (t, J = 7.5 Hz, 1H), 4.44 (s, 3H), 3.07 (s, 3H).

1,2-dimethylquinolin-1-ium iodide (150 mg, 0.50 mmol) and 4-(methylthio)benzaldehyde (70 μ L, 0.50 mmol) were dissolved in ethanol (5 mL) inside a round bottom flask. While stirring, one drop of piperidine was added and the resulting solution was refluxed for 12 hours. The reaction was cooled, then diluted with water (10 mL). The resulting precipitate was filtered, rinsed with water and cold ethanol, then dried under vacuum then recrystallized with toluene to yield (E)-1-methyl-2-(4-(methylthio)styryl)quinolin-1-ium iodide (134 mg, 61% yield) as a dark purple powder. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.05 (d, J = 8.9 Hz, 1H), 8.57 (dd, J = 9.0, 6.1 Hz, 2H), 8.35 (dd, J = 8.1, 1.6 Hz, 1H), 8.25 – 8.14 (m, 2H), 8.00 – 7.92 (m, 3H), 7.90 (d, J = 11.9 Hz, 1H), 7.42 (d, J = 8.5 Hz, 2H), 4.56 (s, 3H), 2.57 (s, 3H). ^{13}C NMR (150 MHz, $\text{DMSO-}d_6$) δ 156.69, 147.13, 144.38, 144.09, 139.71, 135.35, 131.68, 130.55, 130.17, 129.44, 128.21, 126.03, 121.48, 119.82, 118.53, 40.40, 14.55. ESI-MS: m/z $\text{C}_{19}\text{H}_{18}\text{N}_2\text{S}^+$ calculated: 292.1154, found: $(\text{M})^+$ 292.0802. UV/Vis: Exc. λ_{max} = 410 nm, Em. λ_{max} = 580 nm.

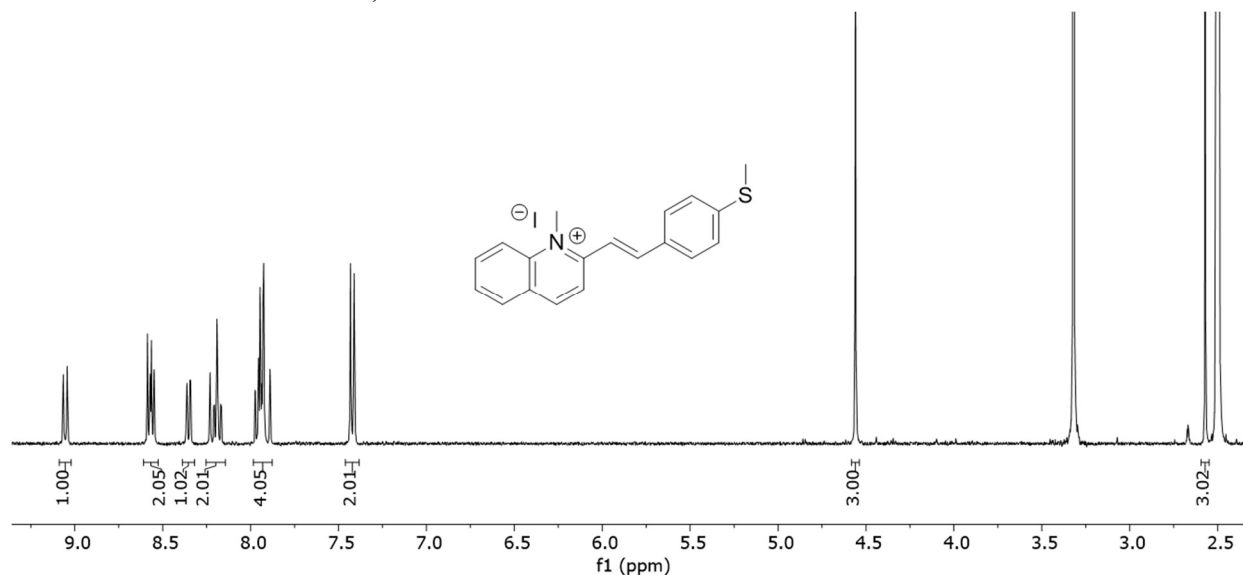


Figure S-11. ^1H NMR spectrum of **2-SMIQ** (400 MHz, $\text{DMSO-}d_6$, 298K).

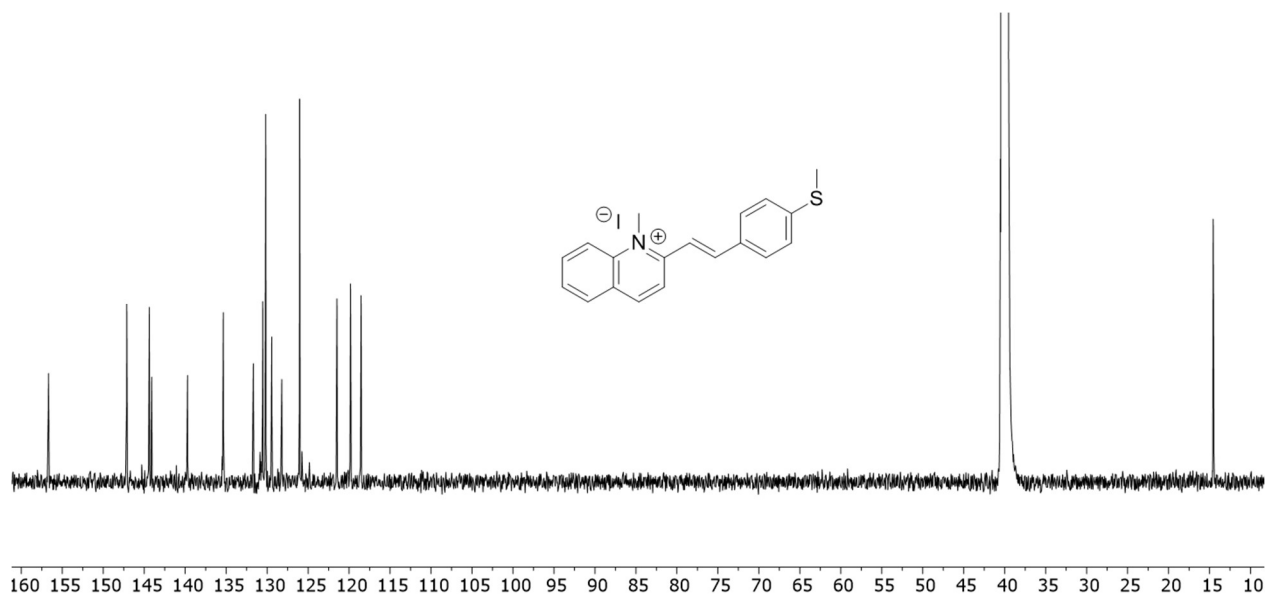
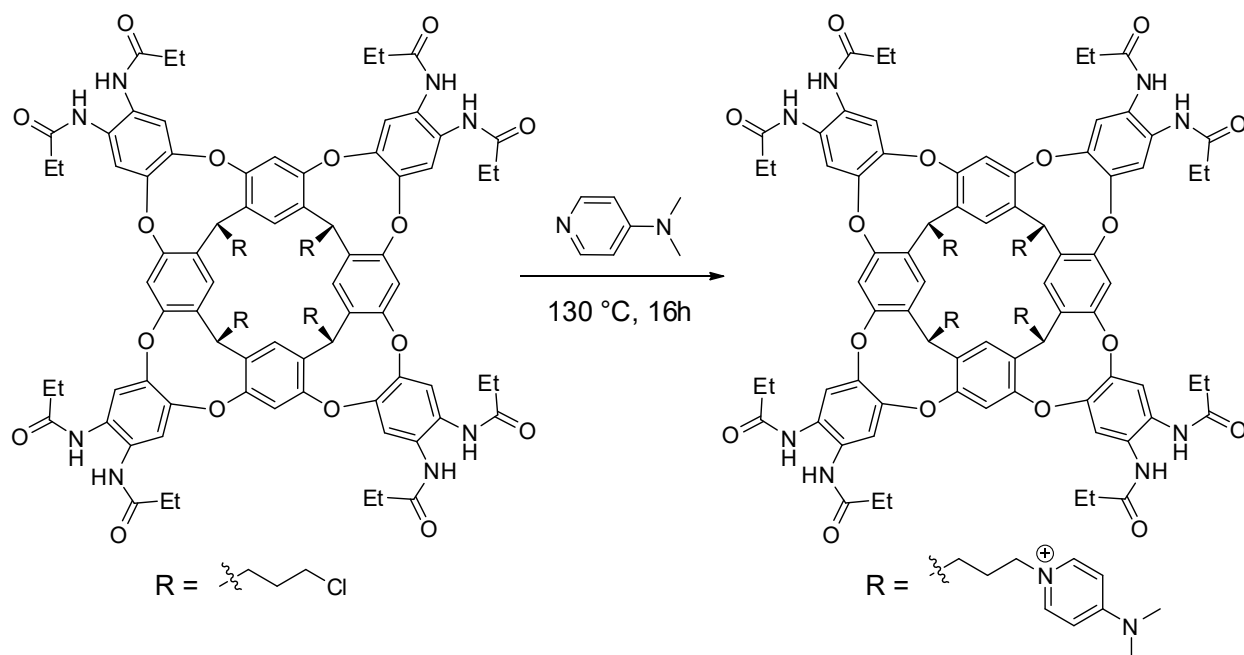


Figure S-12. ^{13}C NMR of **2-SMIQ** (150 MHz, $\text{DMSO-}d_6$, 298K).



4-dimethylaminopyridine-amide cavitand (AMD):

Chloro-amide cavitand (50 mg, 0.031 mmol) was placed in a pear-shaped flask and excess 4-dimethylaminopyridine (148 mg, 1.22 mmol) was placed on top, the reaction was melted at 130 °C for 16 h. The reaction was cooled and cold acetone (2 mL) was added to form a pale beige precipitate which was filtered and collected. The solid was then refluxed in acetone (3 mL) for 16 h. The reaction was again cooled and the solid was filtered and dried resulting in 4-

dimethylaminopyridine-amide cavitand (47 mg, 67% yield) as a beige solid. ^1H NMR (400 MHz, Acetone) δ 8.45 (d, J = 6.8 Hz, 3H), 7.92 (s, 1H), 7.66 (s, 1H), 7.41 (s, 1H), 7.03 (d, J = 6.8 Hz, 4H), 4.58 (s, 2H), 3.24 (s, 2H), 3.20 (d, J = 6.0 Hz, 14H), 2.78 (s, 1H), 2.44 – 2.36 (m, 5H), 1.96 – 1.79 (m, 5H), 1.13 (t, J = 7.6 Hz, 10H). ESI MS: m/z $\text{C}_{120}\text{H}_{140}\text{N}_{16}\text{O}_{16}^{4+}$ calculated: 515.266, found: 515.493. The product was too insoluble in NMR solvents to allow acquisition of a ^{13}C spectrum in any reasonable amount of time.

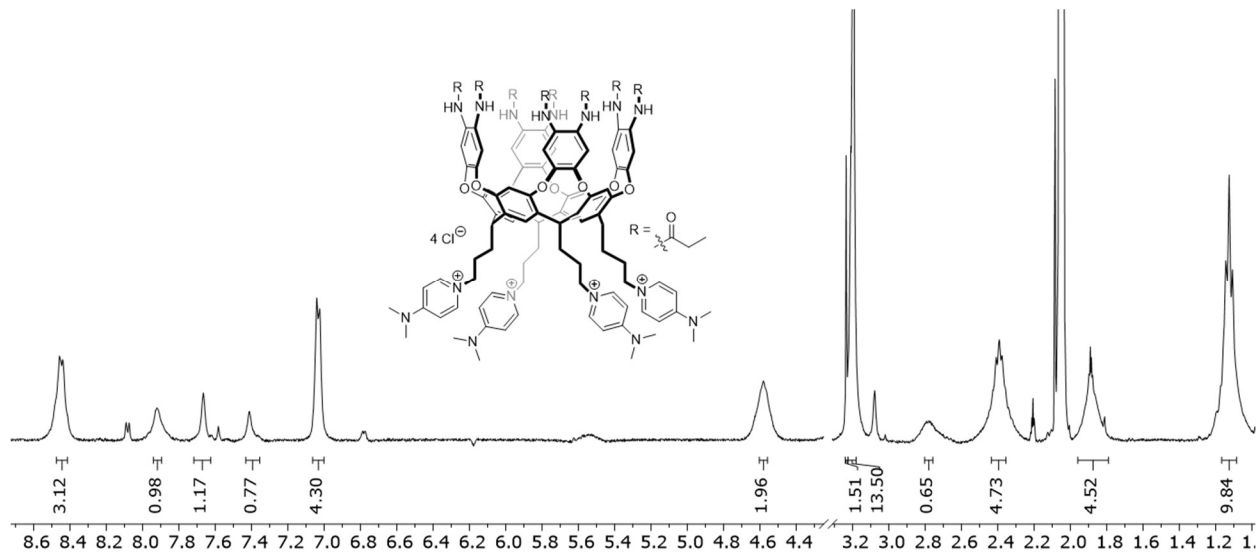


Figure S-13. ^1H NMR spectrum of **AMD** cavitand (400 MHz, D_2O , 298K).

4. Titration Curves and Binding Affinity Measurements

4.1 Fluorescence Titration of Dye-DNA

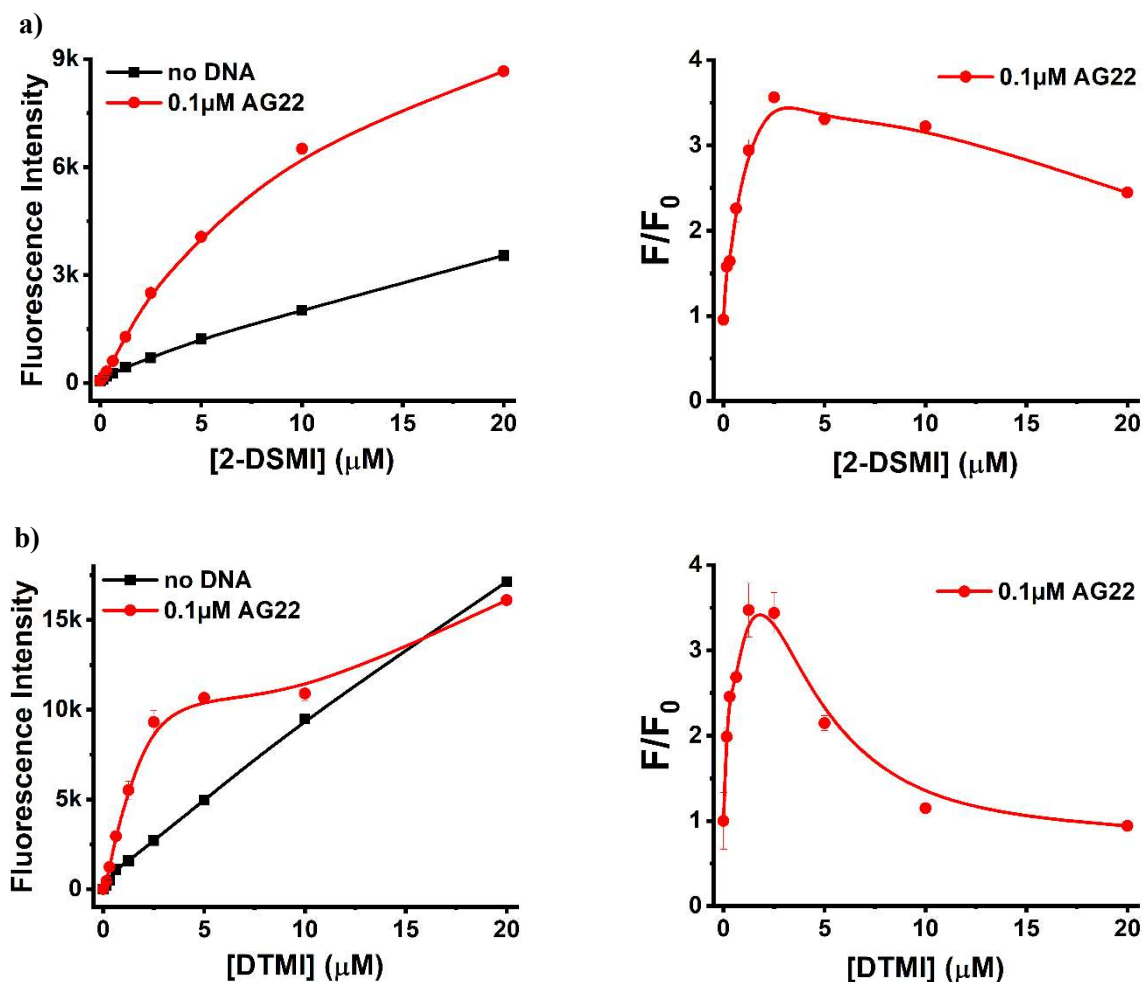


Figure S-14. Fluorescence response curves of DNA AG22 with increasing concentration (0-20 μM) of Dye a) 2-DSMI and b) DTMI. Left: plots using the raw fluorescence counts; Right: plots using the fluorescence normalized against that of the dye (F_0 being the dye fluorescence in the absence of DNA). [Dye] = 0-20 μM , [AG22] = 0.1 μM , K^+ buffer (10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4). 2-DSMI Ex/Em = 480/580 nm, DTMI Ex/Em = 540/600 nm.

4.2 Job plots of Dye-DNA

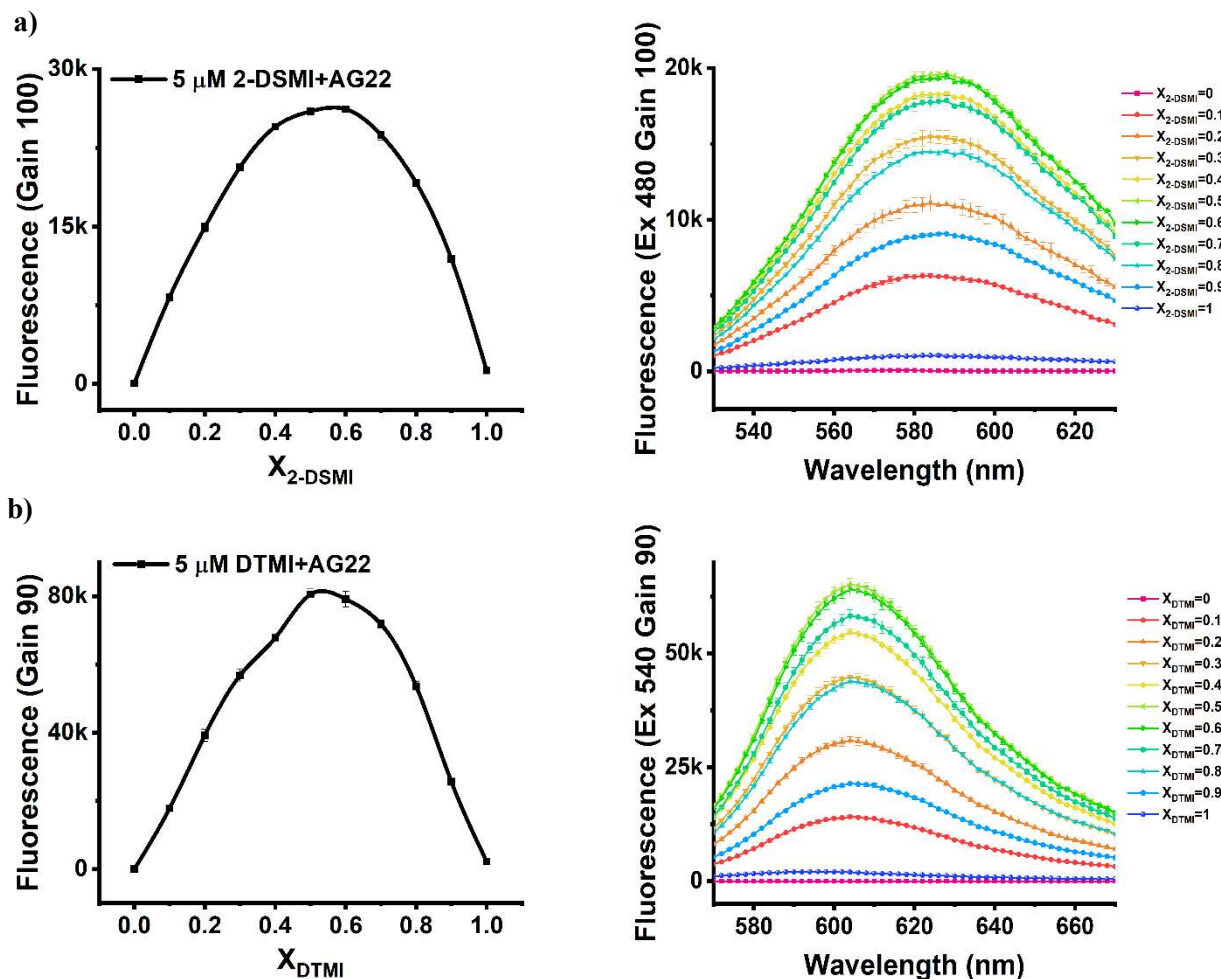


Figure S-15. Fluorescence Job plots of DNA *AG22* with Dye a) **2-DSMI** and b) **DTMI**. Fluorescence of solutions containing Dye and *AG22* at the total concentration of $5 \mu\text{M}$ with the Dye mole fraction (X_{Dye}) varied between 0 and 1 was measured. Left: Fluorescence changes measured at selected excitation and emission wavelengths using Endpoint mode; Right: Emission spectra at the selected excitation wavelength using Spectrum mode. K^+ buffer ($10\text{mM } \text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, $1\text{mM } \text{EDTA}$, $\text{pH } 7.4$). **2-DSMI** Ex/Em = $480/580 \text{ nm}$, Gain = 100; **DTMI** Ex/Em = $540/600 \text{ nm}$, Gain = 90. (The default Gain value of all the fluorescence measurement = 100 unless specifically emphasized.)

4.3 Affinity Measurement of DNA-Dye

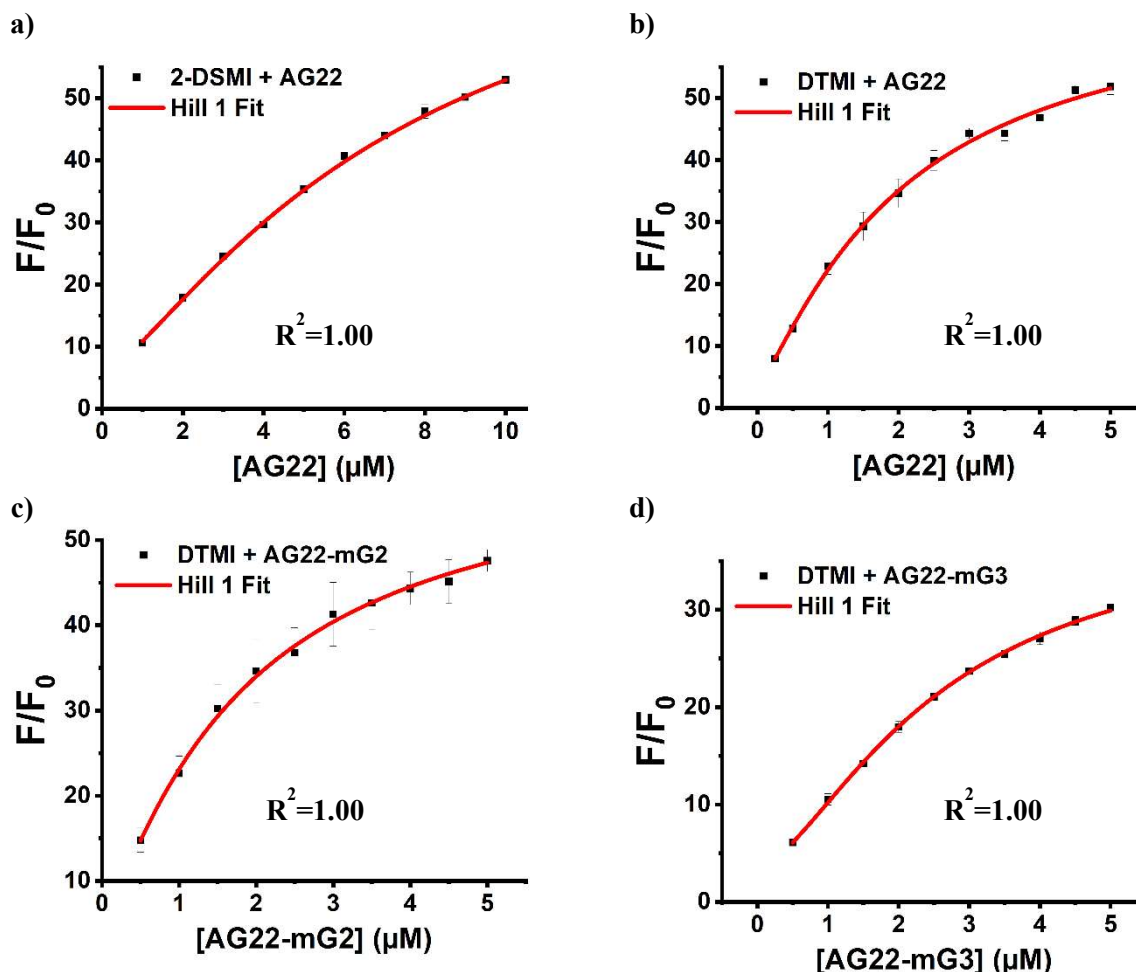


Figure S-16. Affinity measurement of DNA with Dyes via fluorescence. a) **2-DSMI** + *AG22*; b) **DTMI** + *AG22*; c) **DTMI** + *AG22-mG2*; d) **DTMI** + *AG22-mG3*. $[Dye] = 0.625 \mu M$, Buffer: 10 mM KH_2PO_4/K_2HPO_4 , 1 mM EDTA, pH 7.4, **2-DSMI** Ex/Em = 480nm/580 nm, **DTMI** Ex/Em = 540nm/600 nm.

Table S-2. The binding affinities of DNA-Dye: *AG22* with **2-DSMI/DTMI** were obtained using Hill 1 fitting of data from Figure S-16.

k (μM)	2-DSMI	DTMI
<i>AG22</i>	7.5 ± 1.7	2.1 ± 0.5

Table S-3. The binding affinities of DNA-Dye: *AG22/AG22-mG2/AG22-mG3* with **DTMI** were obtained using Hill 1 fitting of data from Figure S-16.

k (μM)	<i>AG22</i>	<i>AG22-mG2</i>	<i>AG22-mG3</i>
DTMI	2.1 ± 0.5	1.9 ± 0.2	2.5 ± 0.1

4.4 Affinity Measurement of Cavitand-Dye

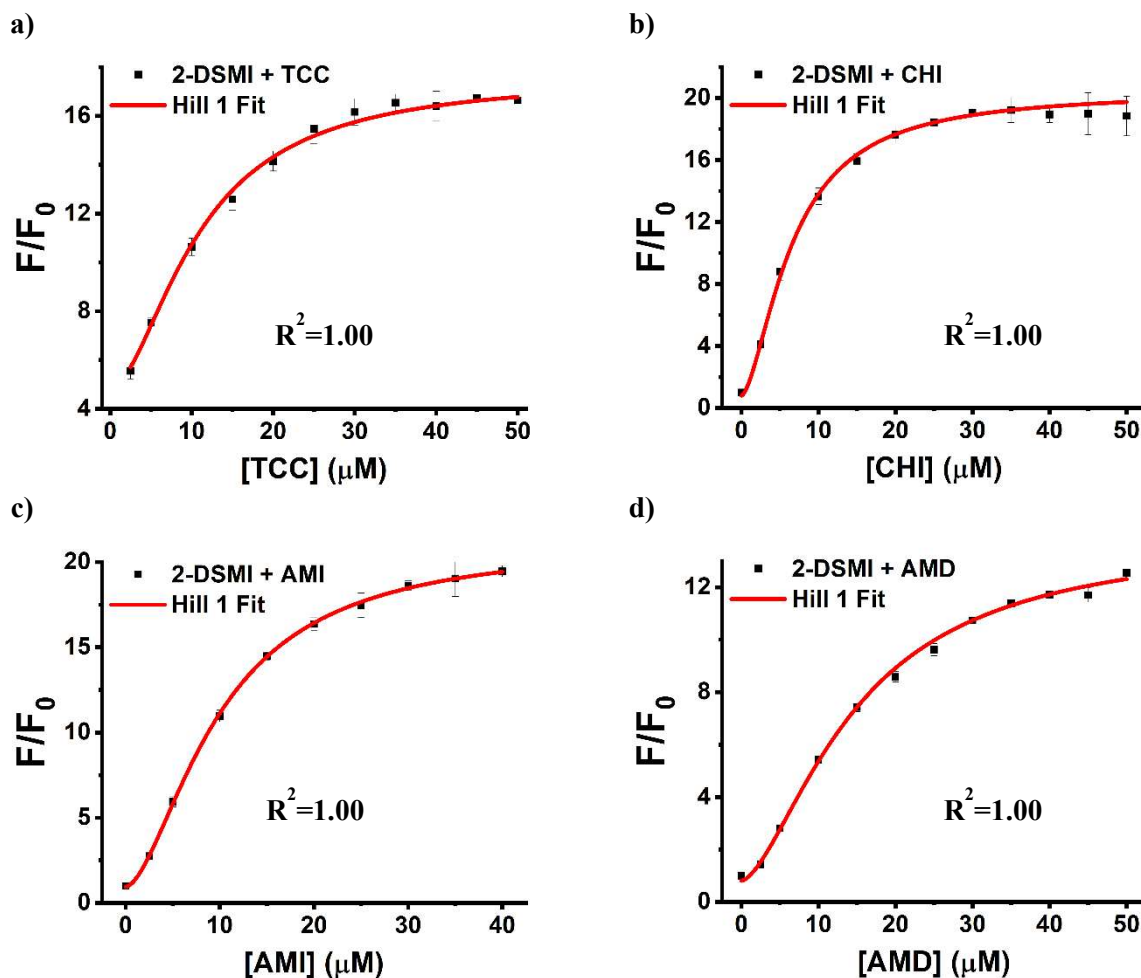


Figure S-17. Affinity measurement of 2-DSMI with Hosts via fluorescence. a) 2-DSMI + TCC; b) 2-DSMI + CHI; c) 2-DSMI + AMI; d) 2-DSMI + AMD. [2-DSMI] = $0.625\mu\text{M}$, Buffer: 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 1 mM EDTA, pH 7.4, Ex/Em = 480nm/580 nm.

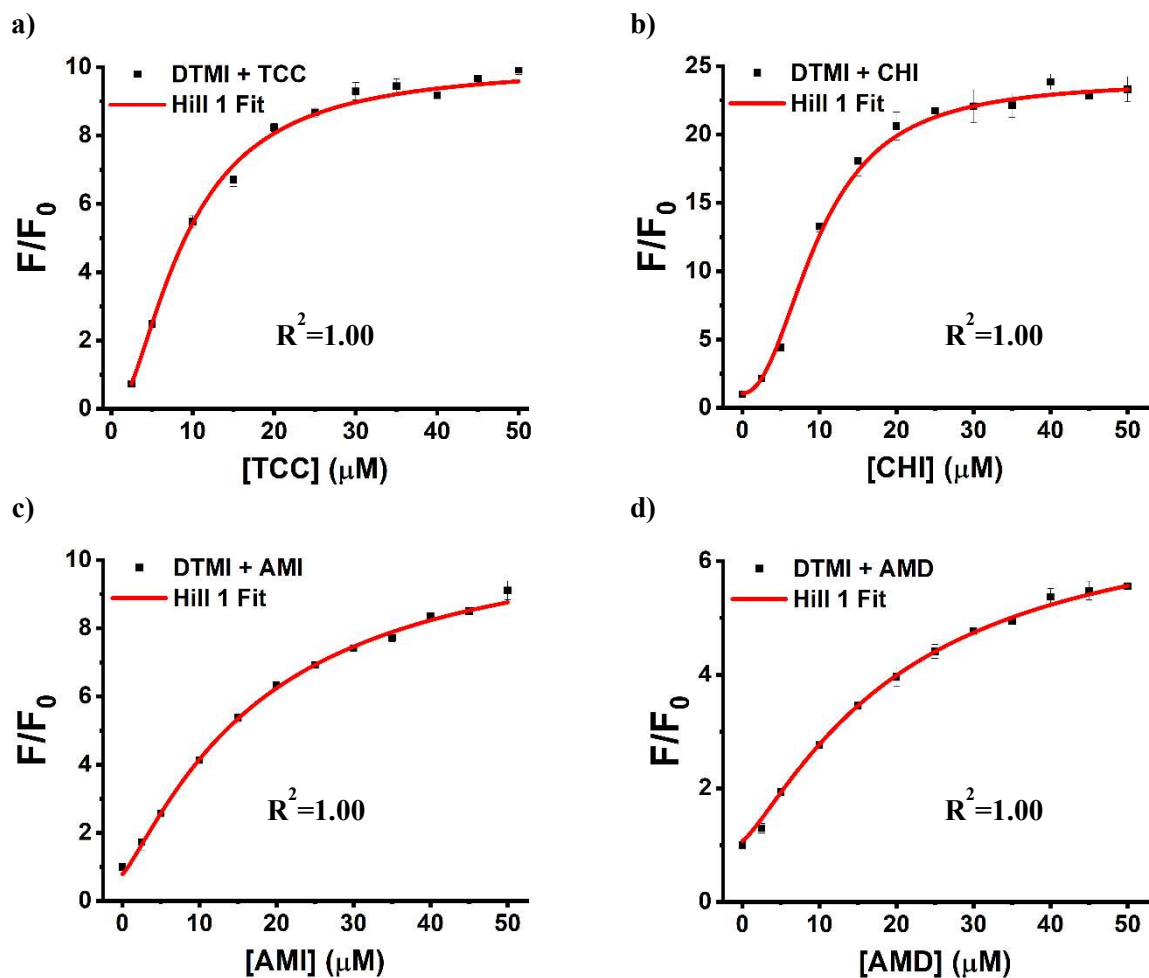


Figure S-18. Affinity measurement of **DTMI** with Hosts via fluorescence. a) **DTMI + TCC**; b) **DTMI + CHI**; c) **DTMI + AMI**; d) **DTMI + AMD**. [DTMI] = $0.625\mu\text{M}$, Buffer: 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 1 mM EDTA, pH 7.4, Ex/Em = 540nm/600 nm.

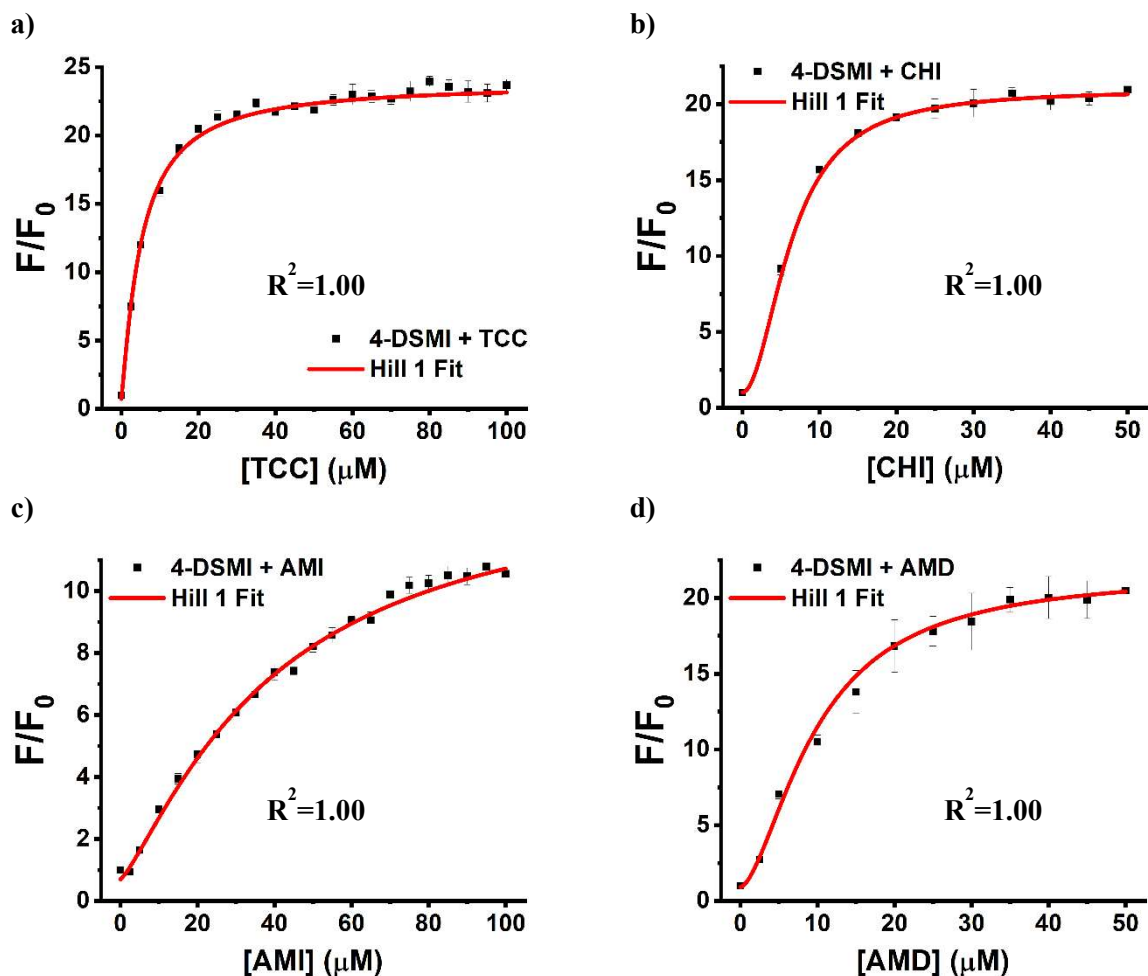


Figure S-19. Affinity measurement of 4-DSMI with Hosts via fluorescence. a) 4-DSMI + TCC; b) 4-DSMI + CHI; c) 4-DSMI + AMI; d) 4-DSMI + AMD. [4-DSMI] = $0.625\mu\text{M}$, Buffer: 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 1 mM EDTA, pH 7.4. The data in a), b), and c) were replotted from ref 9, whose raw data were recorded in a Perkin Elmer Wallac 1420 Victor 2 Microplate Reader with the Ex/Em wavelengths at 485/605 nm. The data in d) were performed with a BioTek Synergy H1 Hybrid Multi-Mode Microplate Reader at Ex/Em = 480nm/600 nm. (The default plate reader used in all the fluorescence measurement is BioTek Synergy H1 Hybrid Multi-Mode Microplate Reader unless specifically emphasized.)

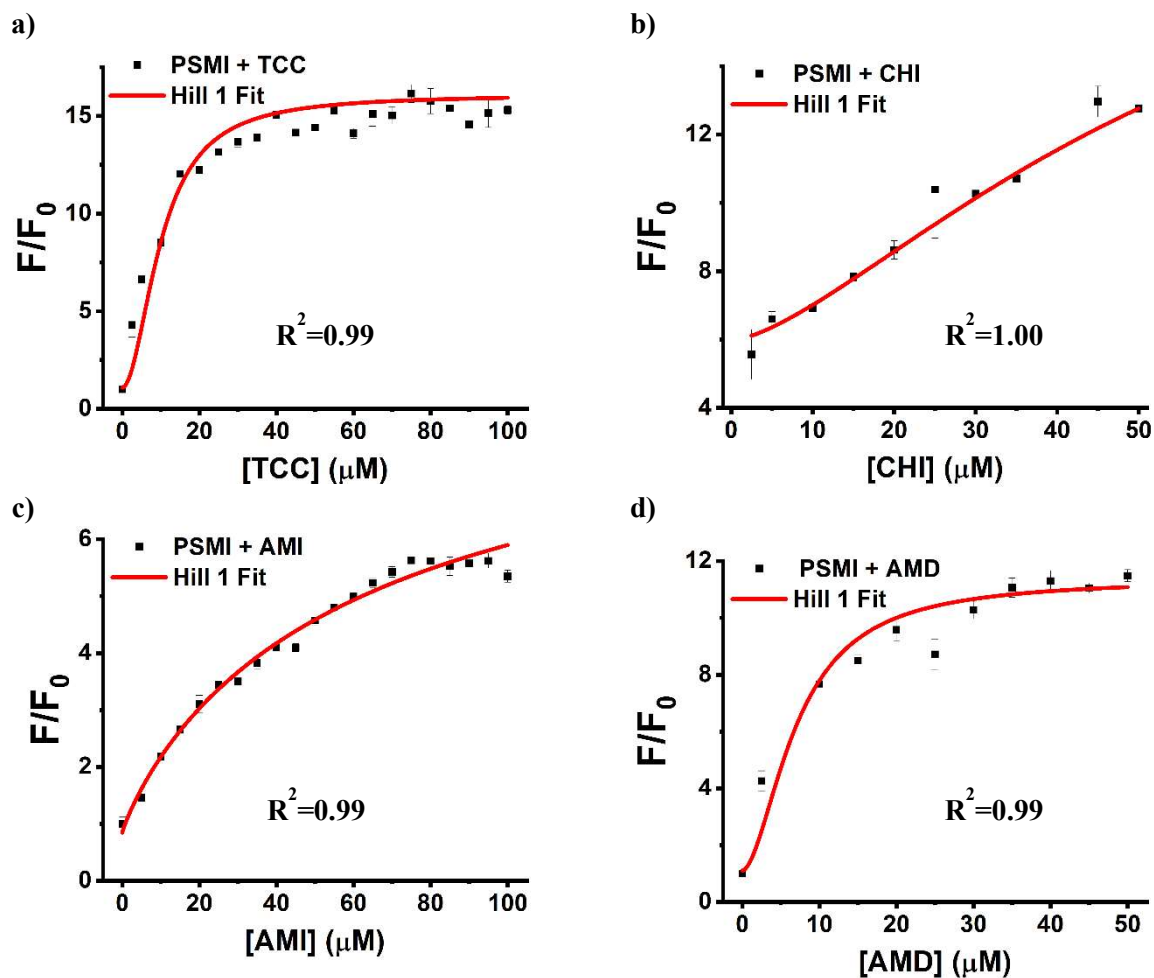


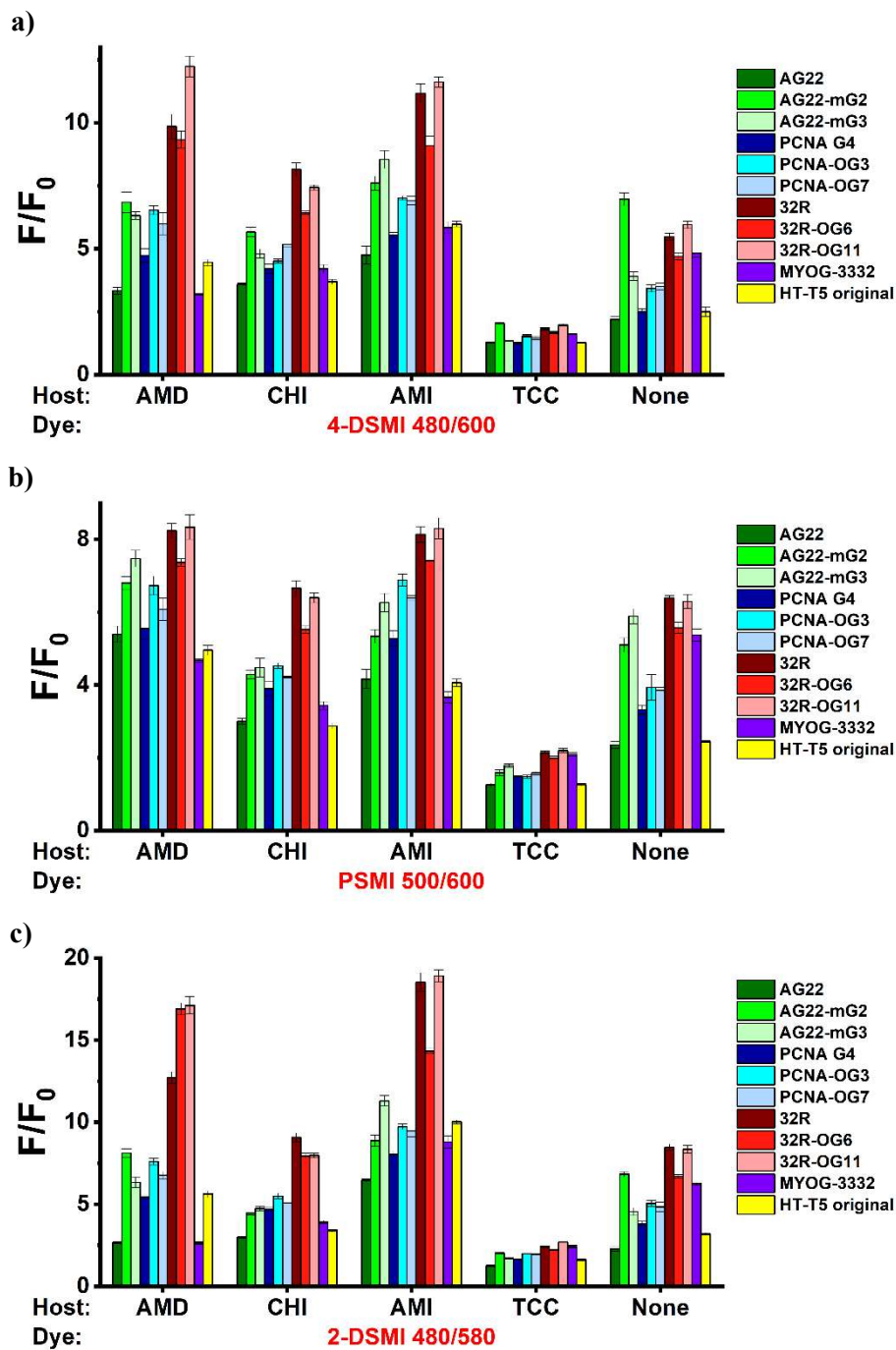
Figure S-20. Affinity measurement of **PSMI** with Hosts via fluorescence. a) **PSMI + TCC**; b) **PSMI + CHI**; c) **PSMI + AMI**; d) **PSMI + AMD**. [PSMI] = $0.625\mu\text{M}$, Buffer: 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 1 mM EDTA, pH 7.4. The data in a) and c) were replotted from ref 9, raw data of a), b), and c) were recorded in a Perkin Elmer Wallac 1420 Victor 2 Microplate Reader with the Ex/Em wavelengths at 485/605 nm. The data in d) were performed with a BioTek Synergy H1 Hybrid Multi-Mode Microplate Reader at Ex/Em = 500nm/600 nm.

Table S-4. The binding affinities of Cavitand-Dye: **2-DSMI/DTMI/4-DSMI/PSMI** with different hosts **TCC/CHI/AMI/AMD** were obtained using Hill 1 fitting of data from Figure S-17, S-18, S-19, and S-20, respectively.

k (μM)	4-DSMI	PSMI	2-DSMI	DTMI
TCC	5.2 $\pm$ 2.0	10.1 $\pm$ 7.6	10.9 $\pm$ 0.7	8.9 $\pm$ 0.6
CHI	6.4 $\pm$ 2.6	57.3 $\pm$ 33.0	6.6 $\pm$ 0.8	9.9 $\pm$ 0.5
AMI	39.8 $\pm$ 6.2	66.1 $\pm$ 37.5	9.9 $\pm$ 0.8	18.2 $\pm$ 3.1
AMD	9.8 $\pm$ 0.9	7.1 $\pm$ 3.5	14.9 $\pm$ 0.8	20.9 $\pm$ 3.3

5. Array Analysis for Differentiation of 11 DNAs

5.1 Bar Plots for Array Signals from 11 DNAs



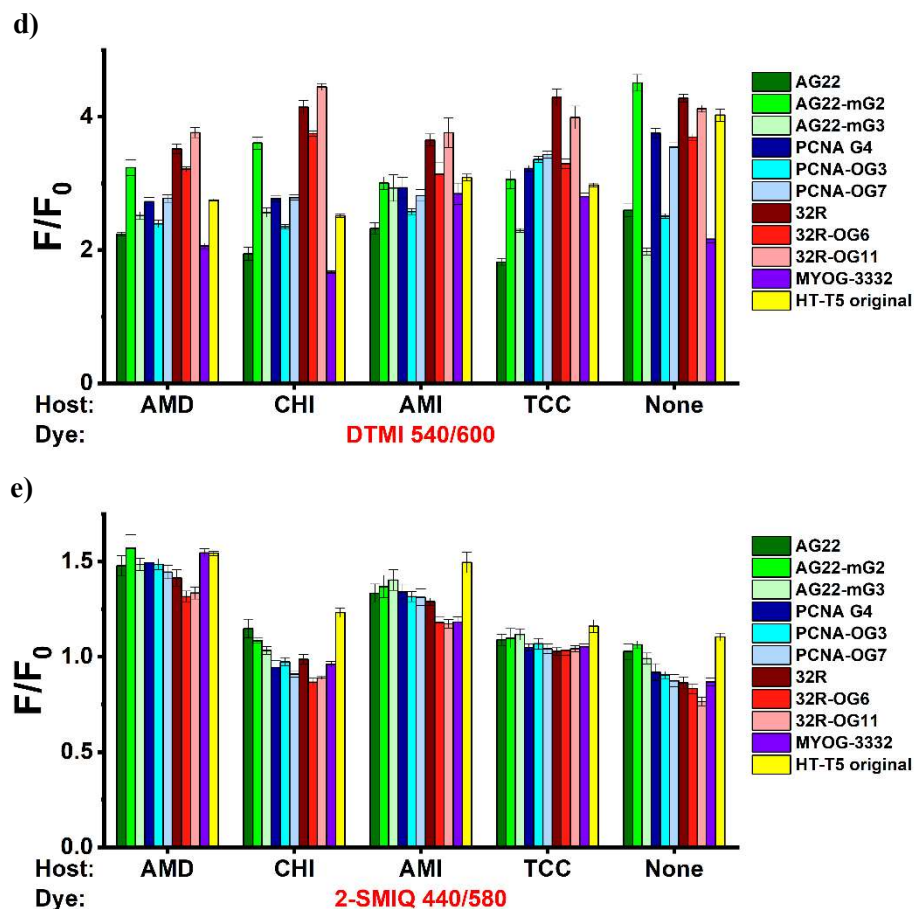


Figure S-21. Full fluorescence response plots of 11 DNA sequences, obtained with the full 25-element array: 5 dyes a) **4-DSMI**, b) **PSMI**, c) **2-DSMI**, d) **DTMI**, and e) **2-SMIQ** with AMD/CHI/AMI/TCC/No cavitand. [Dye] = 0. 625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). **4-DSMI** Ex/Em = 480/600 nm, **PSMI** Ex/Em = 500/600 nm, **2-DSMI** Ex/Em = 480/580 nm, **DTMI** Ex/Em = 540/600 nm, **2-SMIQ** Ex/Em = 440/580 nm.

5.2 PCA biplot for 25-element Array Signals from 11 DNAs

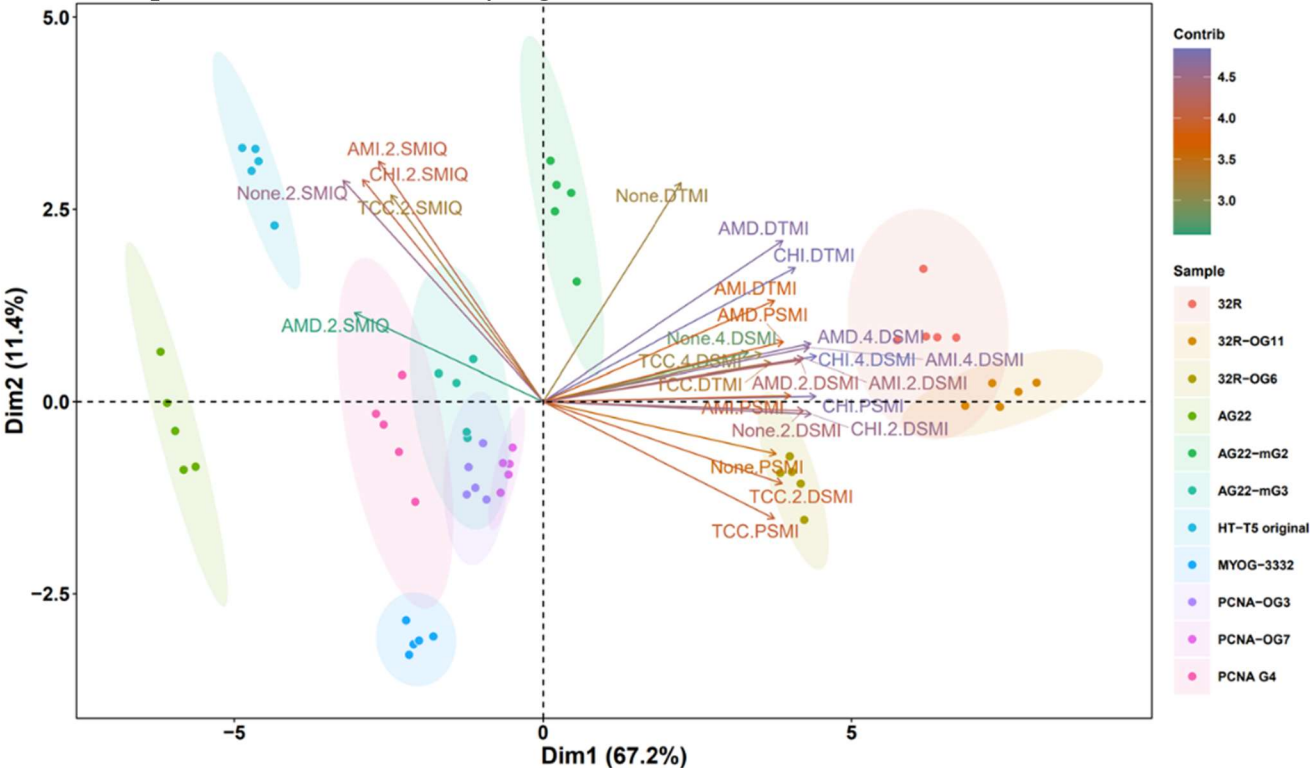


Figure S-22. PCA biplot (combining both PCA scores plot and loading plot) of 11 DNA sequences using the full 25-element array from Figure S-21. Loadings are gradient-colored according to the contribution of each variable. Ellipses indicate 95% confidence intervals.

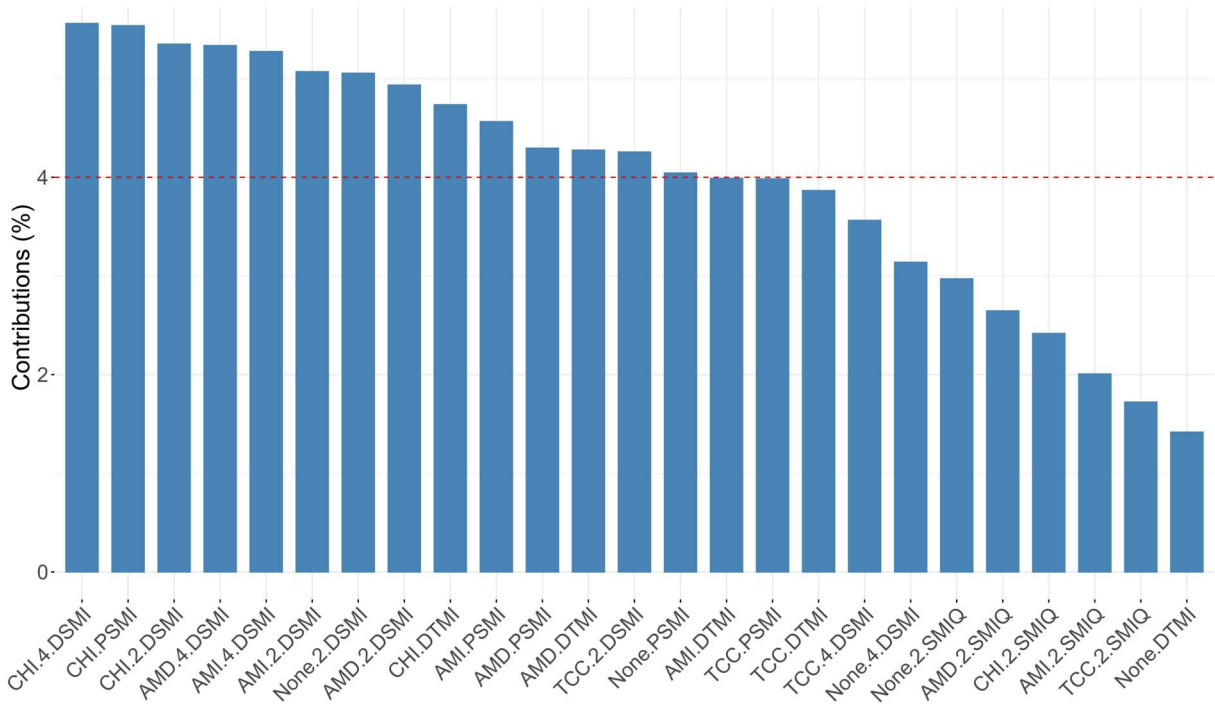


Figure S-23. Contribution of variables to PC 1 in the 25-element array for sensing 11 G4s from Figure S-22. The red dashed line indicates the expected average contribution.

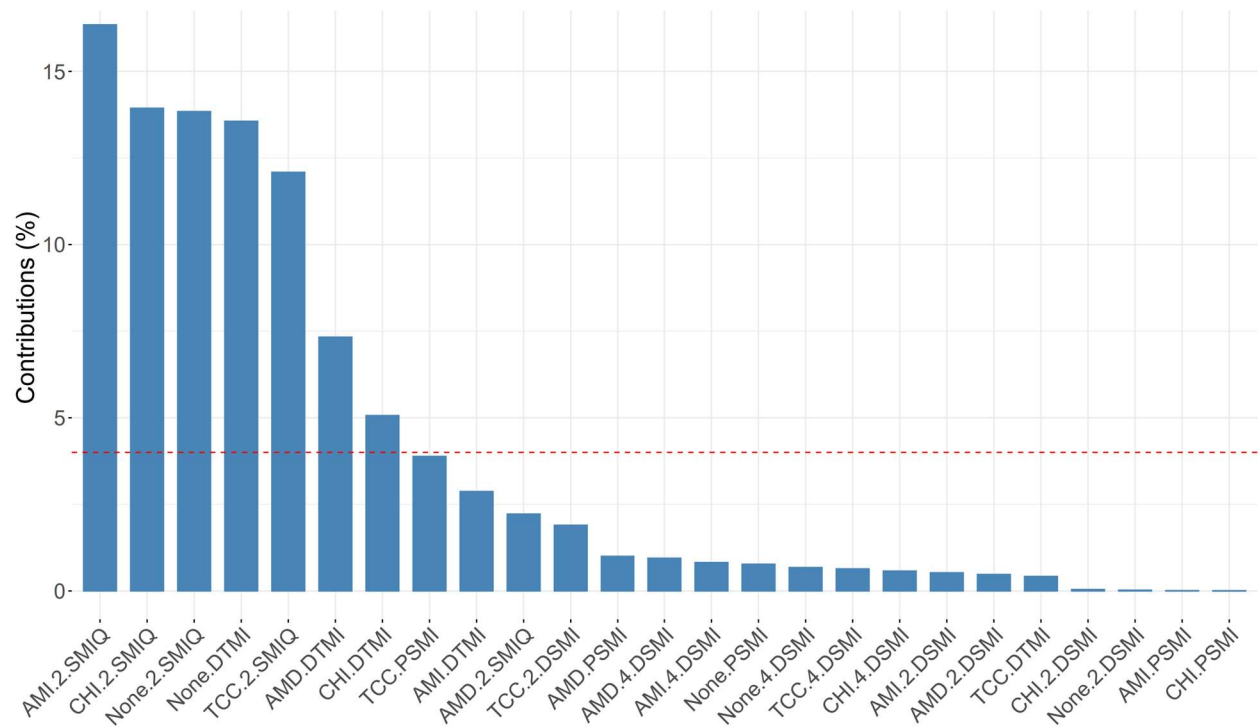


Figure S-24. Contribution of variables to PC 2 in the 25-element array for sensing 11 G4s from Figure S-22. The red dashed line indicates the expected average contribution.

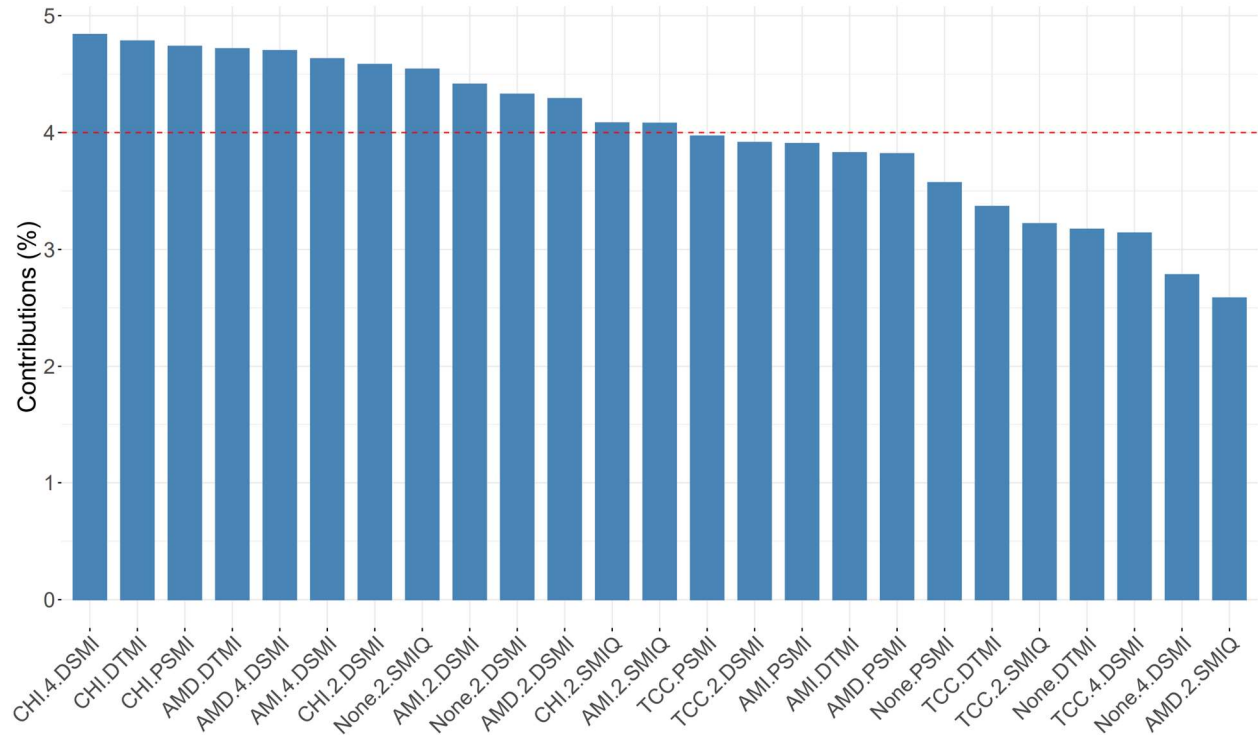


Figure S-25. Total contribution of variables to PC 1 & PC 2 in the 25-element array for sensing 11 G4s from Figure S-22. The red dashed line indicates the expected average contribution.

6. Array Analysis for Differentiation of 9 Core DNAs

6.1 PCA for 16-element Array Signals from 9 Core DNAs

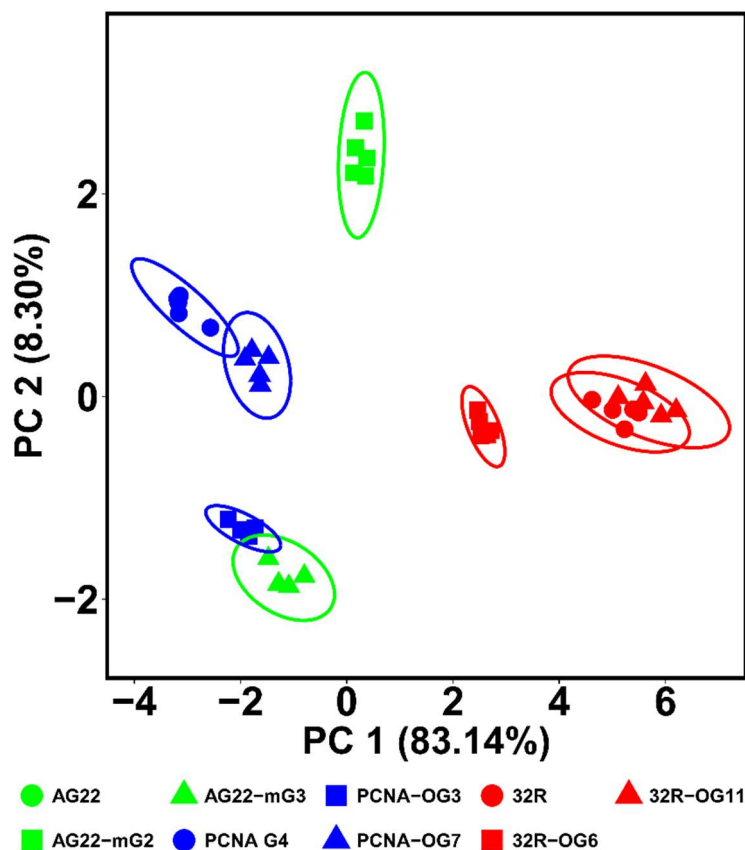


Figure S-26. PCA plot of 9 core DNA sequences, obtained with the 16-element array: 4 dyes **4-DSMI**, **PSMI**, **2-DSMI**, and **DTMI** with **AMD/CHI/AMI/No** cavitand. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). **4-DSMI** Ex/Em = 480/600 nm, **PSMI** Ex/Em = 500/600 nm, **2-DSMI** Ex/Em = 480/580 nm, **DTMI** Ex/Em = 540/600 nm. Ellipses indicate 95% confidence.

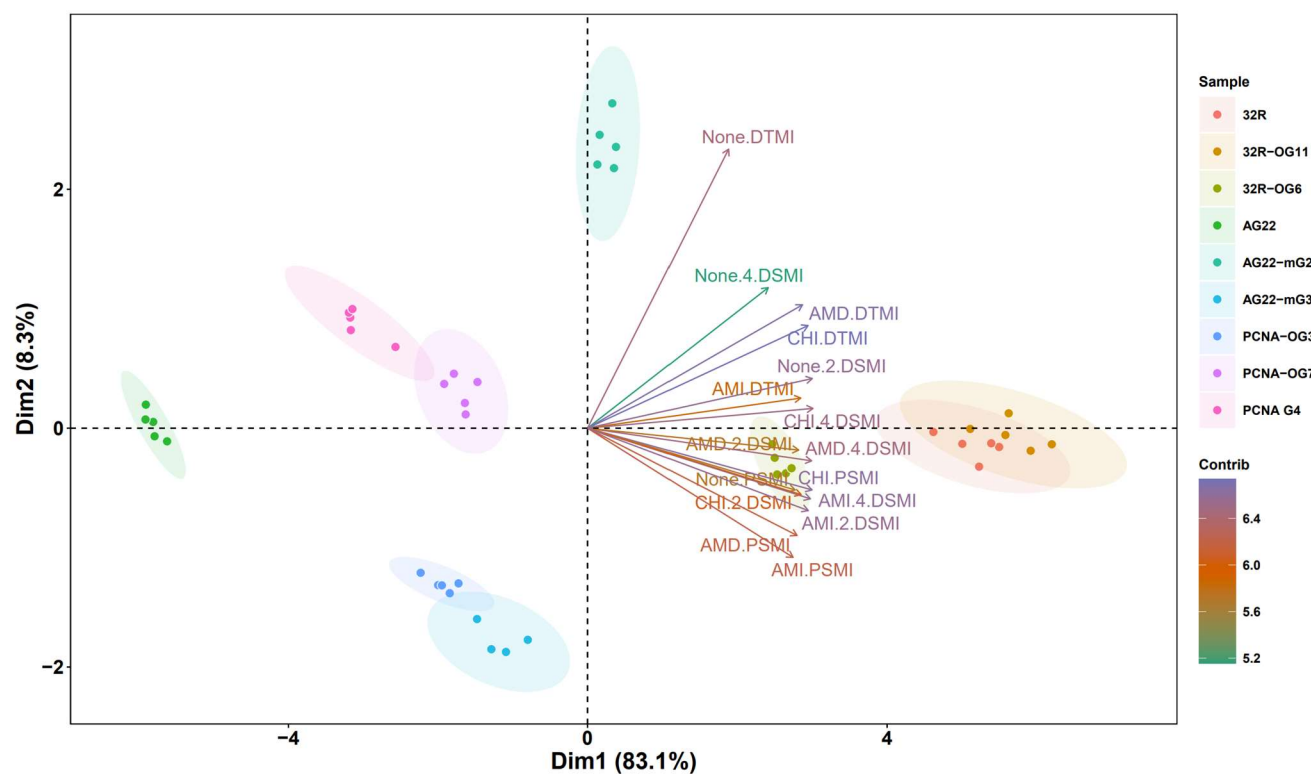


Figure S-27. PCA biplot (combining both PCA scores plot and loading plot) of 9 DNA sequences using the 16-element array from Figure S-26. Loadings are gradient-colored according to the contribution of each variable. Ellipses indicate 95% confidence intervals.

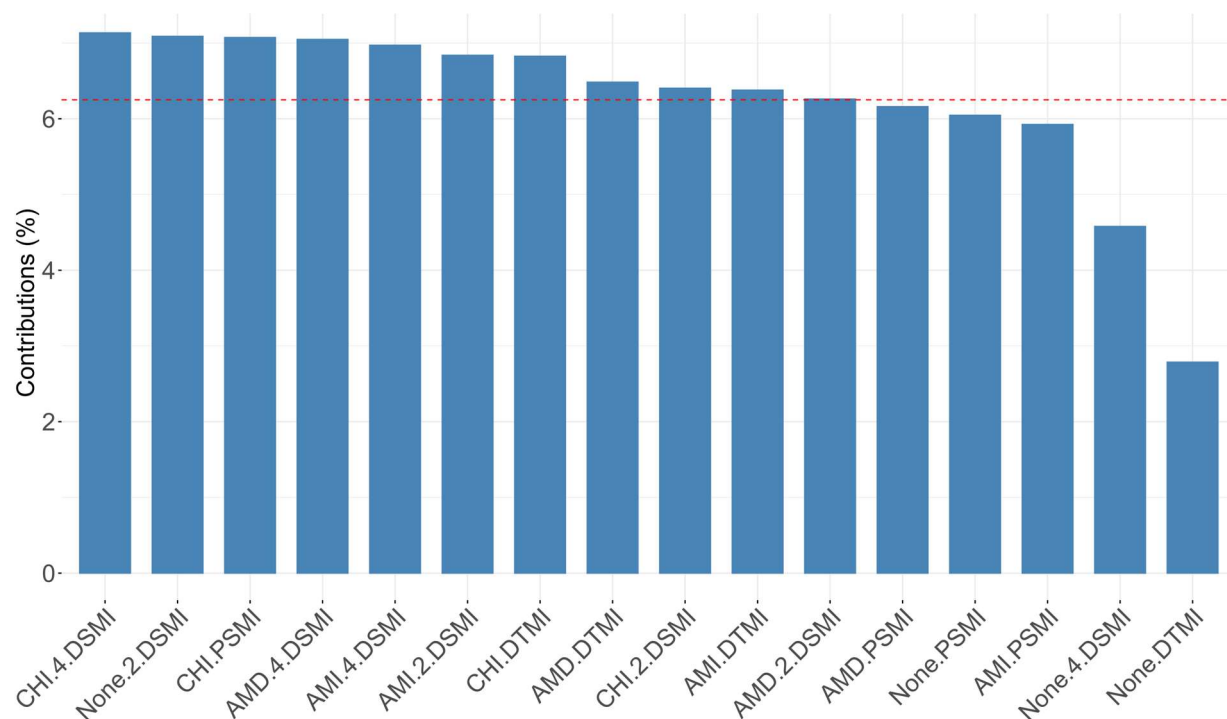


Figure S-28. Contribution of variables to PC 1 in the 16-element array for sensing 9 G4s from Figure S-27. The red dashed line indicates the expected average contribution.

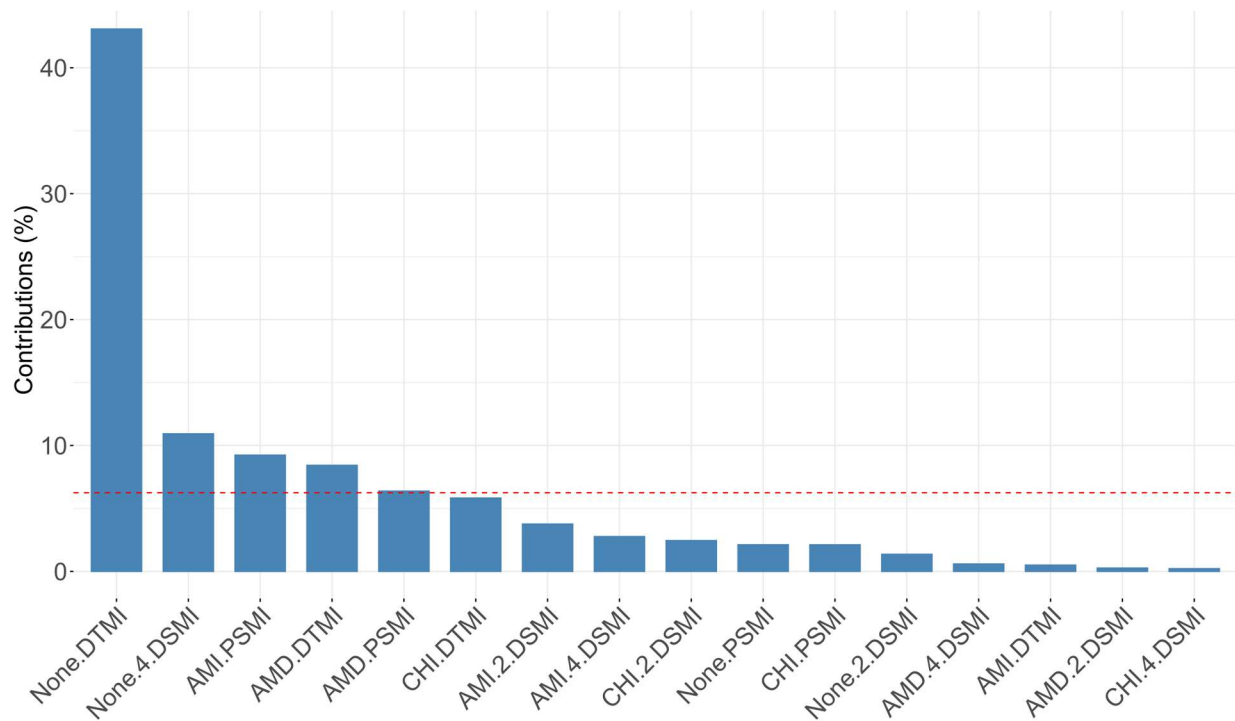


Figure S-29. Contribution of variables to PC 2 in the 16-element array for sensing 9 G4s from Figure S-27. The red dashed line indicates the expected average contribution.

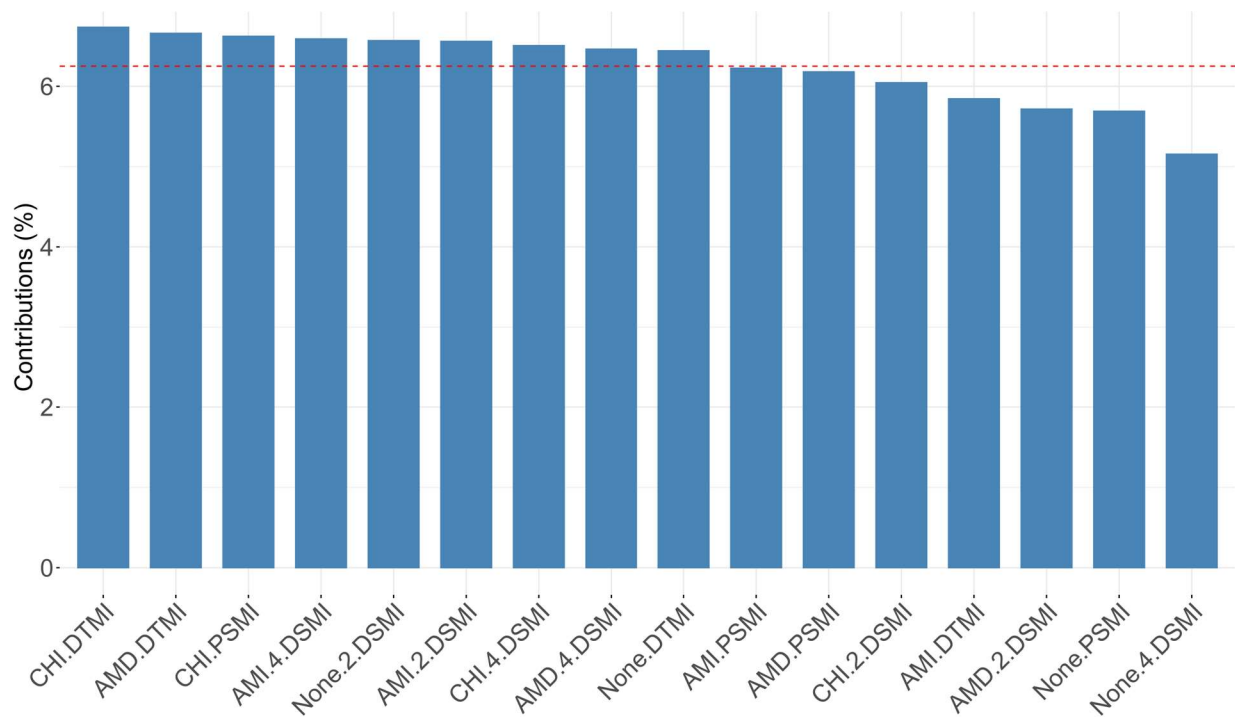


Figure S-30. Total contribution of variables to PC 1 & PC 2 in the 16-element array for sensing 9 G4s from Figure S-27. The red dashed line indicates the expected average contribution.

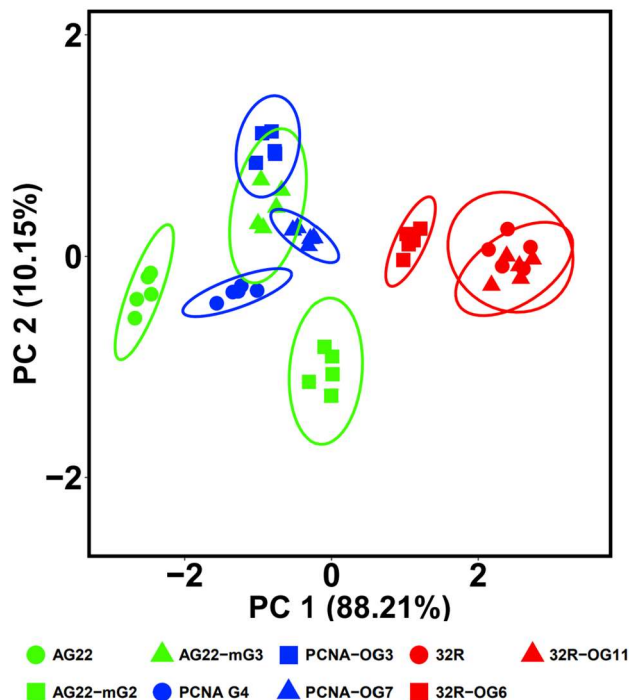


Figure S-31. PCA plot of 9 core DNA sequences, obtained with the manually selected 3-element array other than the SVM-RFECV selected ones, which show different loading vectors in Figure S-27: **PSMI+AMI**, **4-DSMI+CHI**, **DTMI+AMD**. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). **4-DSMI** Ex/Em = 480/600 nm, **PSMI** Ex/Em = 500/600 nm, **DTMI** Ex/Em = 540/600 nm. Ellipses indicate 95% confidence.

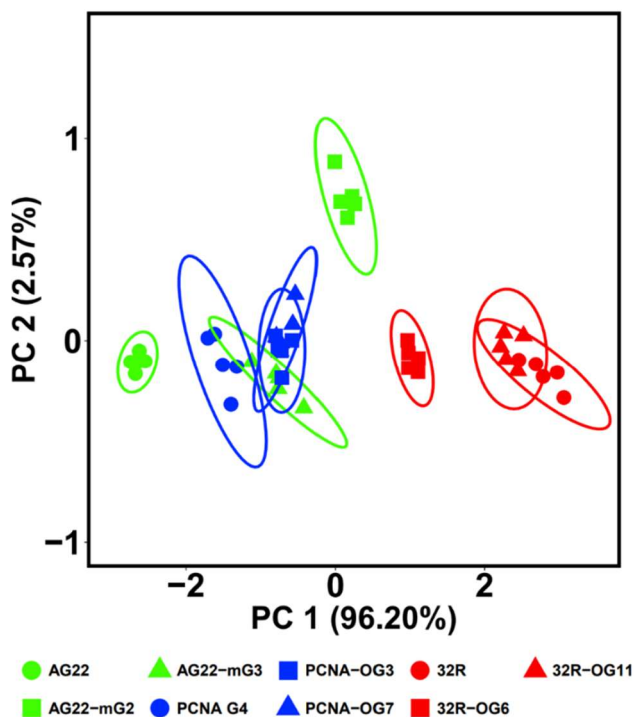


Figure S-32. PCA plot of 9 core DNA sequences, obtained with the top 3 elements in the contribution to PC 1 (Figure S-28): **4-DSMI+CHI**, **2-DSMI+None**, **PSMI+CHI**. [Dye] = 0.625 μ M, [Host] = 0.625 μ M,

[DNA] = 0.1 μ M, in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4). **4-DSMI** Ex/Em = 480/600 nm, **PSMI** Ex/Em = 500/600 nm, **2-DSMI** Ex/Em = 480/580 nm. Ellipses indicate 95% confidence.

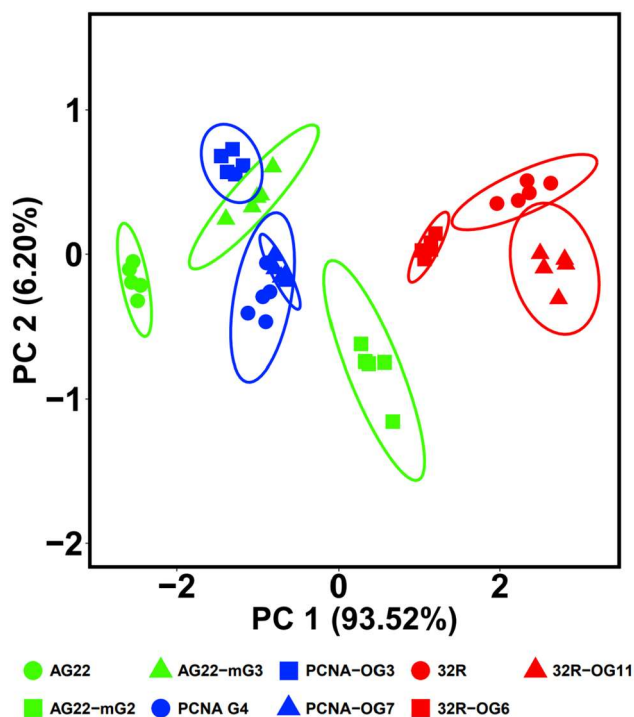


Figure S-33. PCA plot of 9 core DNA sequences, obtained with the top 3 elements in the total contribution to PC 1 & PC 2 (Figure S-30): **PSMI+CHI**, **DTMI+AMD**, **DTMI+CHI**. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4). **PSMI** Ex/Em = 500/600 nm, **DTMI** Ex/Em = 540/600 nm. Ellipses indicate 95% confidence.

6.2 SVM-RFECV Feature Selection for 9 Core DNAs Classification

Table S-5. SVM-RFECV Rank List of 16-element Array for Classification of 9 Core DNAs.

Dye	Host	Rank	Select
4-DSMI 480/600	AMD	4	FALSE
	CHI	7	FALSE
	AMI	1	TRUE
	None	3	FALSE
PSMI 500/600	AMD	13	FALSE
	CHI	14	FALSE
	AMI	2	FALSE
	None	5	FALSE
DTMI 540/600	AMD	6	FALSE
	CHI	10	FALSE
	AMI	9	FALSE
	None	1	TRUE
2-DSMI 480/580	AMD	1	TRUE
	CHI	8	FALSE
	AMI	12	FALSE
	None	11	FALSE

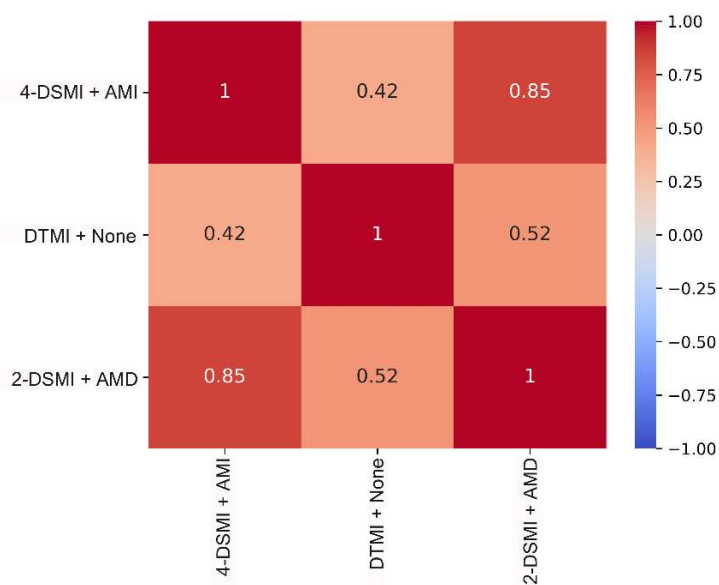


Figure S-34. Correlation of SVM-RFECV selected 3 features: **4-DSMI+AMI** Ex/Em = 480/600 nm, **DTMI+None** Ex/Em = 540/600 nm, and **2-DSMI+AMD** Ex/Em = 480/580 nm in the data set of 9 core DNAs.

6.3 Performance Metrics of 9 Core DNAs Classification

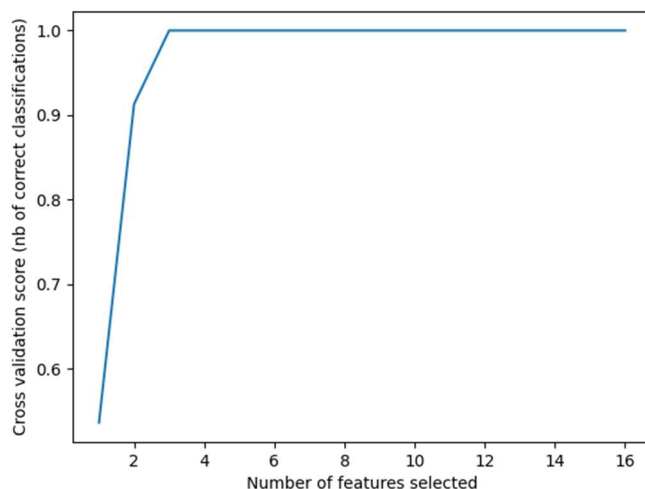


Figure S-35. The cross-validation scores correspond to the increasing numbers of features from 16-element array for 9 core DNAs classification.

Table S-6. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of 9 core DNA using selected 3 features: **4-DSMI+AMI** Ex/Em = 480/600 nm, **DTMI+None** Ex/Em = 540/600 nm, and **2-DSMI+AMD** Ex/Em = 480/580 nm.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	1.0000 (0.0000)
Sensitivity	1.0000 (0.0000)
Specificity	1.0000 (0.0000)
Precision	1.0000 (0.0000)
F1 Score	1.0000 (0.0000)
AUC	1.0000 (0.0000)

7. Array Analysis for *AG22* Series DNAs

7.1 SVM-RFECV Feature Selection for *AG22* Series DNAs Classification

Table S-7. SVM-RFECV rank list of 16-element array for classification of *AG22* series DNAs.

Dye	Host	Rank	Select
4-DSMI 480/600	AMD	11	FALSE
	CHI	6	FALSE
	AMI	12	FALSE
	None	2	FALSE
PSMI 500/600	AMD	14	FALSE
	CHI	13	FALSE
	AMI	10	FALSE
	None	3	FALSE
DTMI 540/600	AMD	7	FALSE
	CHI	4	FALSE
	AMI	15	FALSE
	None	1	TRUE
2-DSMI 480/580	AMD	8	FALSE
	CHI	5	FALSE
	AMI	1	TRUE
	None	9	FALSE

7.2 Performance metrics of *AG22* Series DNAs Classification

Table S-8. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of *AG22* series DNAs using single feature: **DTMI+None** Ex/Em = 540/600 nm.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	0.8750 (0.1250)
Sensitivity	0.9167 (0.0833)
Specificity	0.9444 (0.0556)
Precision	0.9167 (0.0833)
F1 Score	0.8889 (0.1111)
AUC	0.6597 (0.0230)

Table S-9. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of *AG22* series DNAs using single feature: **2-DSMI+AMI** Ex/Em = 480/580 nm.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	1.0000 (0.0000)
Sensitivity	1.0000 (0.0000)
Specificity	1.0000 (0.0000)
Precision	1.0000 (0.0000)
F1 Score	1.0000 (0.0000)
AUC	1.0000 (0.0000)

7.3 t-distribution Curve of *AG22* series DNAs

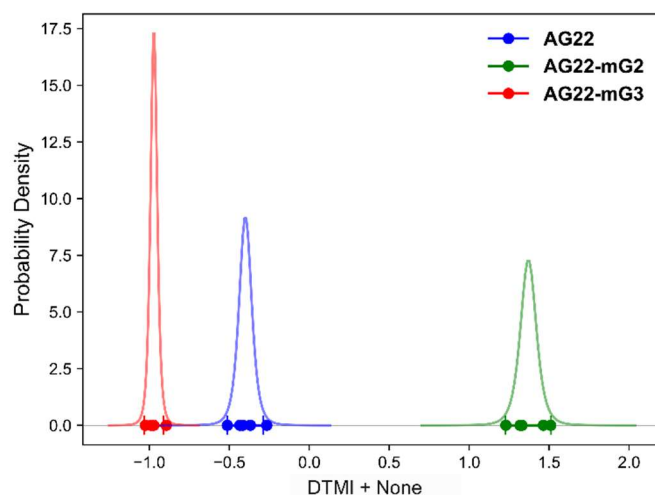


Figure S-36. t-disturbution curve of *AG22* series DNA sequences obtained with single element: **DTMI+None** Ex/Em = 540/600 nm. [Dye] = 0. 625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). Red/blue/green dots = datapoints, curve = t-distribution probability density, vertical markers = 95% confidence intervals.

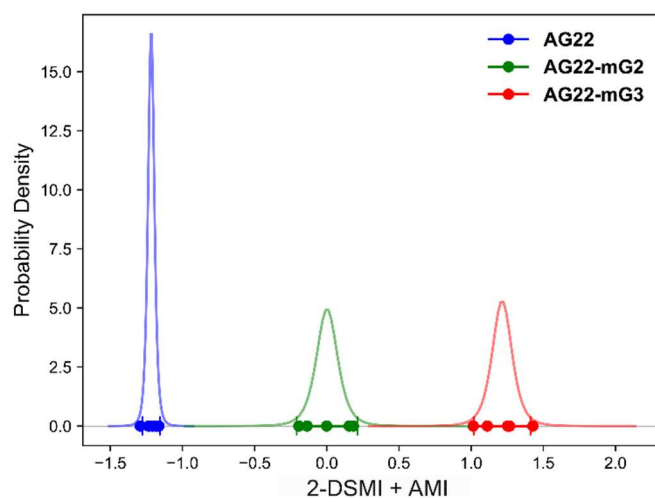


Figure S-37. t-disturbution curve of *AG22* series DNA sequences obtained with single element: **2-DSMI+AMI** Ex/Em = 480/580 nm. [Dye] = 0. 625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). Red/blue/green dots = datapoints, curve = t-distribution probability density, vertical markers = 95% confidence intervals.

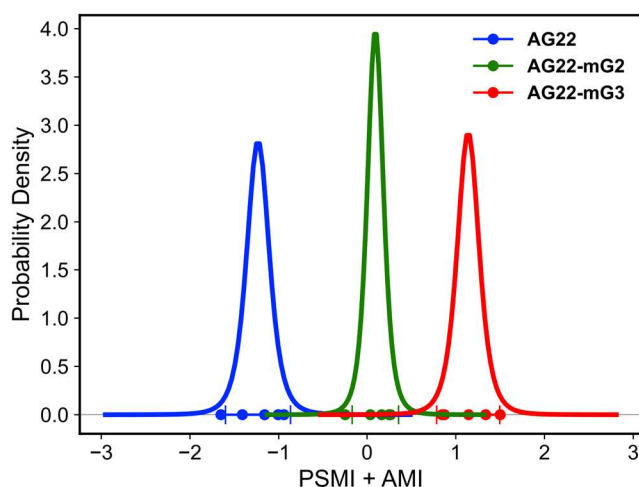


Figure S-38. t-disturbution curve of *AG22* series DNA sequences obtained with single element which was not selected by SVM-RFECV for comparison: **2-PSMI+AMI** Ex/Em = 500/600 nm. [Dye] = 0. 625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, in K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4). Red/blue/green dots = datapoints, curve = t-distribution probability density, vertical markers = 95% confidence intervals.

7.4 Bar Plots for Array Signals from AG22 Series DNA Mixture

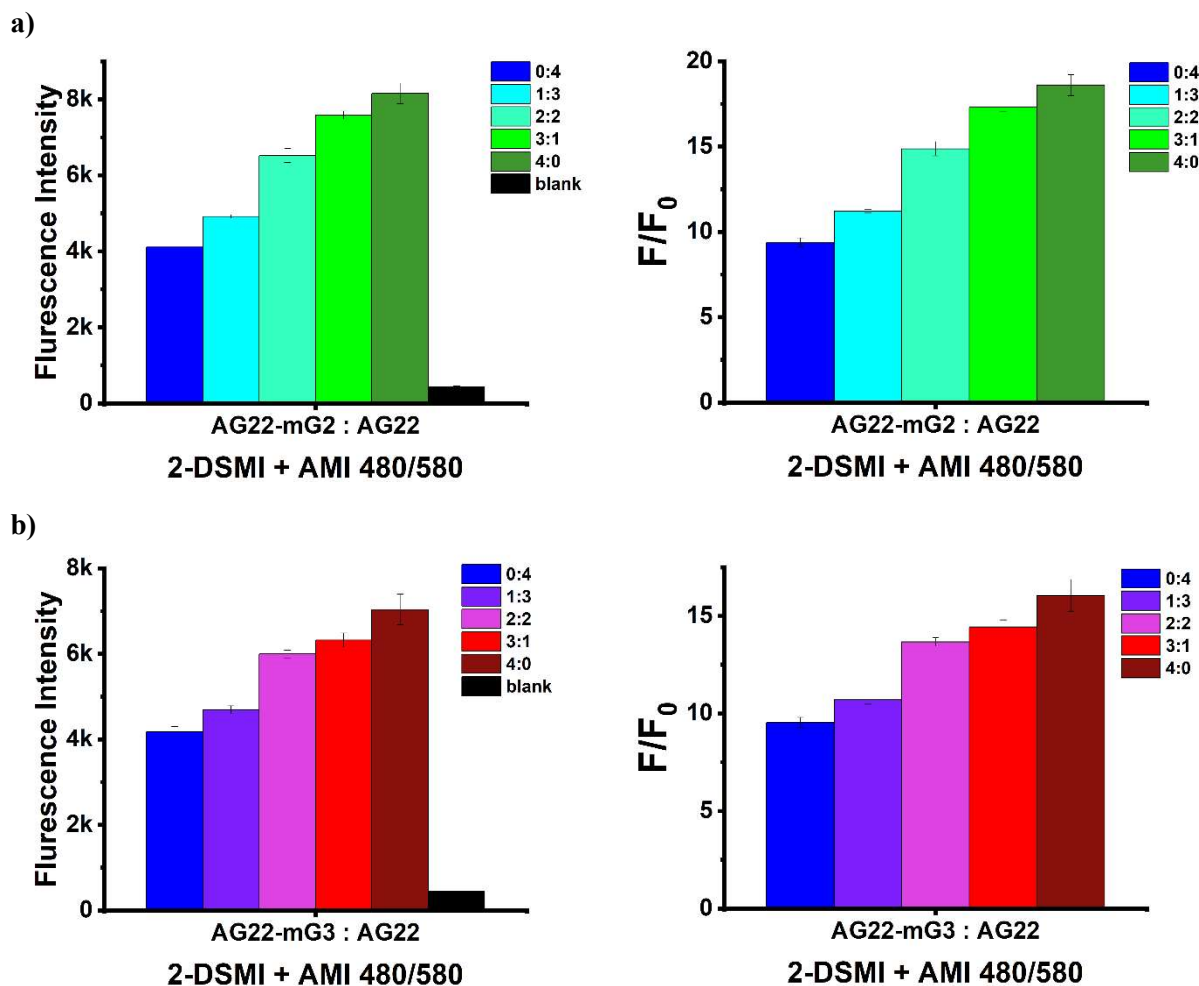


Figure S-39. Fluorescence response plots of mixture of AG22 with O⁶-methylguanine modified sequences at the total concentration of 0.5 μ M, obtained with single element 2-DSMI + AMI Ex/Em = 480/580 nm. a) Mixture of AG22-mG2:AG22 = 0:4, 1:3, 2:2, 3:1, and 4:0; b) Mixture of AG22-mG3:AG22 = 0:4, 1:3, 2:2, 3:1, and 4:0. Left: the raw fluorescence counts (F); Right: F/F₀ plots normalized to the response of Host:guest in the absence of DNA (F₀). [Dye] = 0.625 μ M, [Host] = 0.625 μ M, total [DNA] = 0.5 μ M, buffer 10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4.

8. Array Analysis for DNA Sensing in Diluted Serum

8.1 Bar Plots of Sensor Element for AG22 Series in Diluted Serum

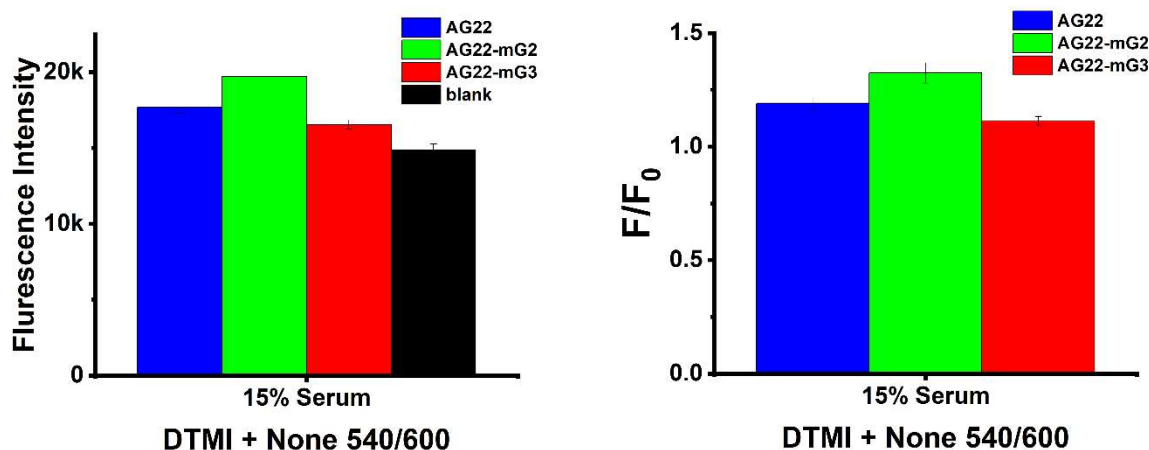


Figure S-40. Fluorescence response plots of AG22 series DNA sequences in 15% serum, obtained with single element **DTMI+None** Ex/Em = 540/600 nm. Left: the raw fluorescence counts (F); Right: F/F₀ plots normalized to the response of Host:guest with 15% serum in the absence of DNA (F₀). [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, 15% serum in 10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer.

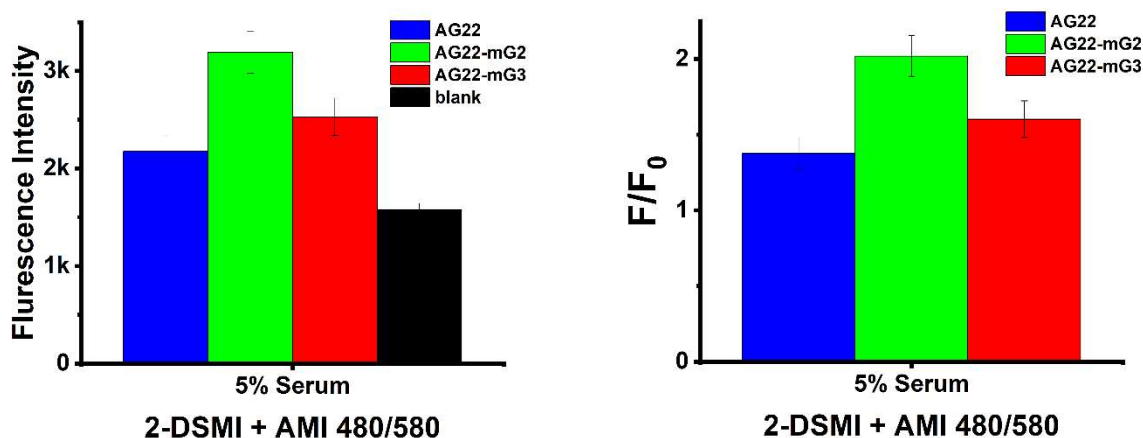


Figure S-41. Fluorescence response plots of AG22 series DNA sequences in 5% serum, obtained with single element **2-DSMI + AMI** Ex/Em = 480/580 nm. Left: the raw fluorescence counts (F); Right: F/F₀ plots normalized to the response of Host:guest with 5% serum in the absence of DNA (F₀). [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, 5% serum in 10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer.

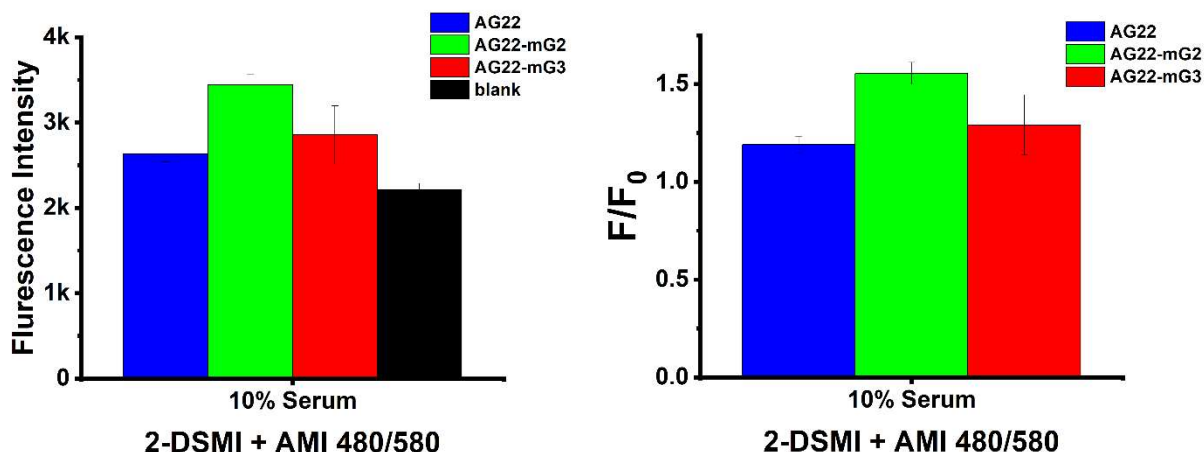


Figure S-42. Fluorescence response plots of *AG22* series DNA sequences in 10% serum, obtained with single element **2-DSMI + AMI** Ex/Em = 480/580 nm. Left: the raw fluorescence counts (F); Right: F/F_0 plots normalized to the response of Host:guest with 10% serum in the absence of DNA (F_0). [Dye] = 0.625 μM , [Host] = 0.625 μM , [DNA] = 0.1 μM , 10% serum in 10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K^+ buffer.

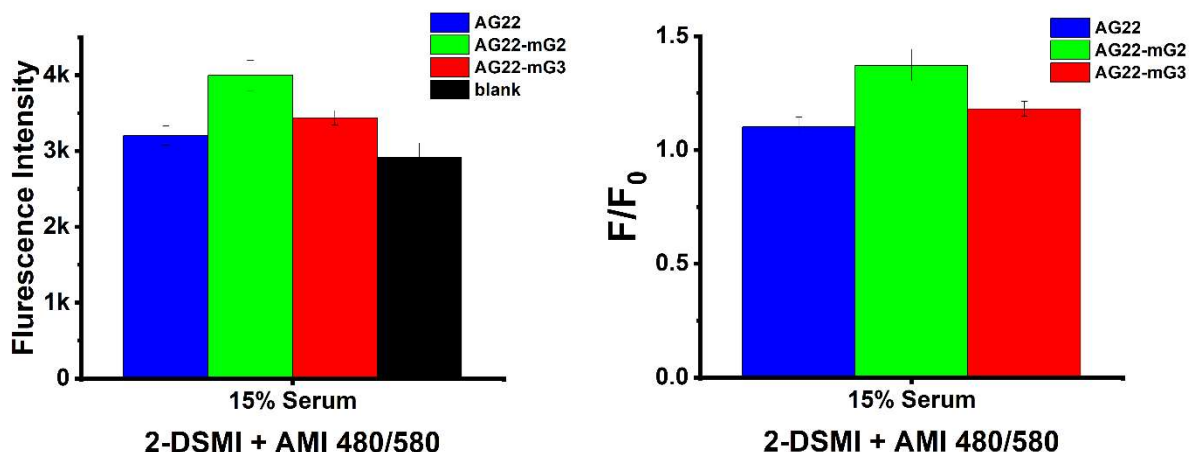


Figure S-43. Fluorescence response plots of *AG22* series DNA sequences in 15% serum, obtained with single element **2-DSMI + AMI** Ex/Em = 480/580 nm. Left: the raw fluorescence counts (F); Right: F/F_0 plots normalized to the response of Host:guest with 15% serum in the absence of DNA (F_0). [Dye] = 0.625 μM , [Host] = 0.625 μM , [DNA] = 0.1 μM , 15% serum in 10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K^+ buffer.

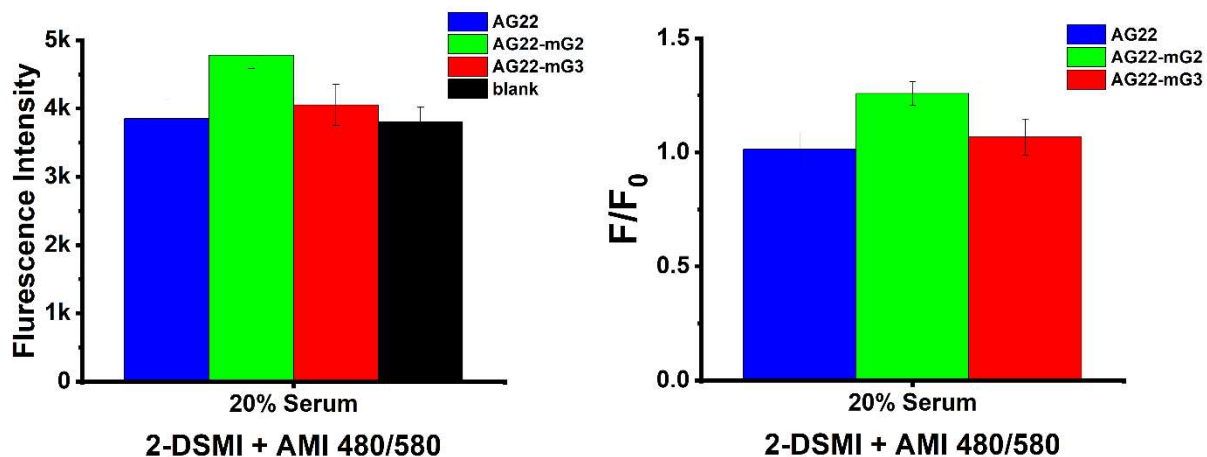


Figure S-44. Fluorescence response plots of *AG22* series DNA sequences in 20% serum, obtained with single element **2-DSMI + AMI** Ex/Em = 480/580 nm. Left: the raw fluorescence counts (F); Right: F/F_0 plots normalized to the response of Host:guest with 20% serum in the absence of DNA (F_0). [Dye] = 0.625 μM , [Host] = 0.625 μM , [DNA] = 0.1 μM , 20% serum in 10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K^+ buffer.

8.2 Bar Plots for Array Signals from AG22 and 32R Series in 10% Serum

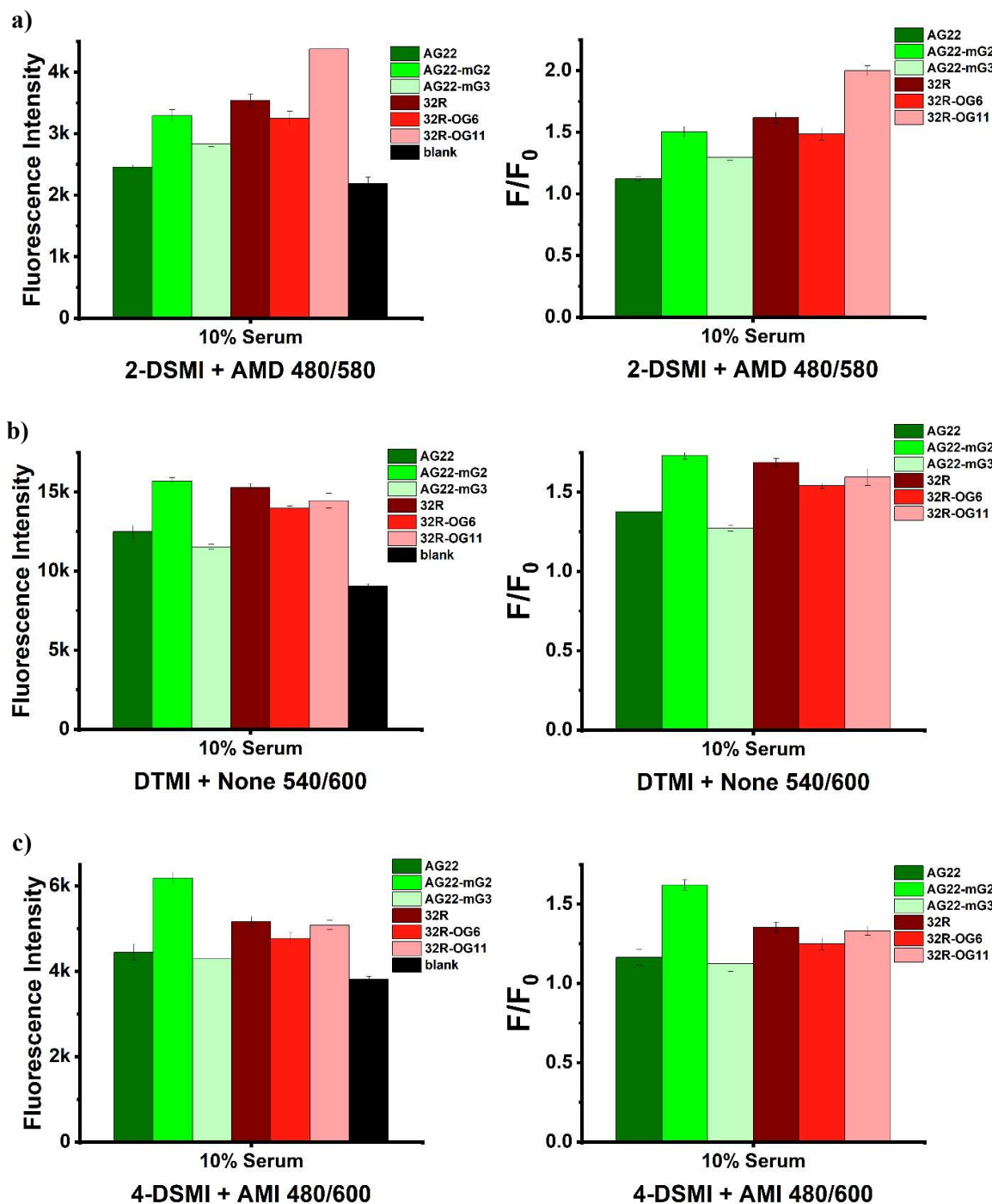


Figure S-45. Fluorescence response plots of AG22 and 32R series DNA sequences in 10% serum, obtained with the 3-element array: a) **2-DSMI+AMD** Ex/Em = 480/580 nm; b) **DTMI+None** Ex/Em = 540/600 nm; c) **4-DSMI+AMI** Ex/Em = 480/600 nm. Left: the raw fluorescence counts (F); Right: F/F₀ plots normalized to the response of Host:guest with 10% serum in the absence of DNA (F₀). [Dye] = 0.625 μM, [Host] = 0.625 μM, [DNA] = 0.1 μM, 10% serum in 10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed

for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer.

8.3 PCA plot of *AG22* and *32R* Series in 10% Serum

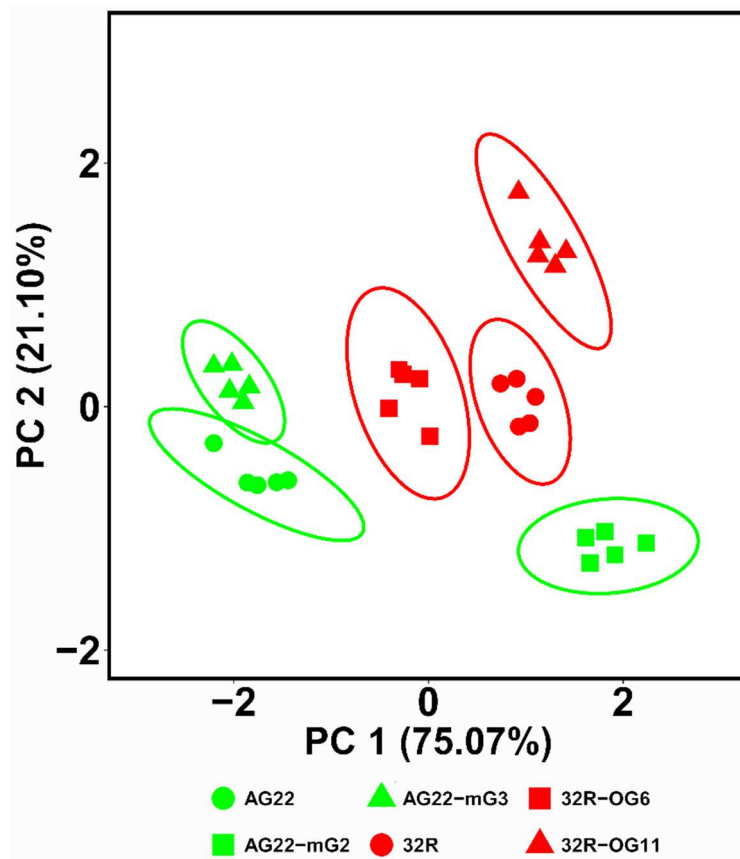


Figure S-46. PCA plot of *AG22* and *32R* series DNA sequences in 10% serum, obtained with the 3-element array: **2-DSMI+AMD** Ex/Em = 480/580 nm; **DTMI+None** Ex/Em = 540/600 nm; **4-DSMI+AMI** Ex/Em = 480/600 nm. Ellipses indicate 95% confidence. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, 10% serum in 10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4. The single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K⁺ buffer.

Table S-10. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of 6 DNAs: *AG22* and *32R* series in 10% single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29) using selected 3 features: **4-DSMI+AMI** Ex/Em = 480/600 nm, **DTMI+None** Ex/Em = 540/600 nm, and **2-DSMI+AMD** Ex/Em = 480/580 nm.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	1.0000 (0.0000)
Sensitivity	1.0000 (0.0000)
Specificity	1.0000 (0.0000)
Precision	1.0000 (0.0000)

F1 Score	1.0000 (0.0000)
AUC	1.0000 (0.0000)

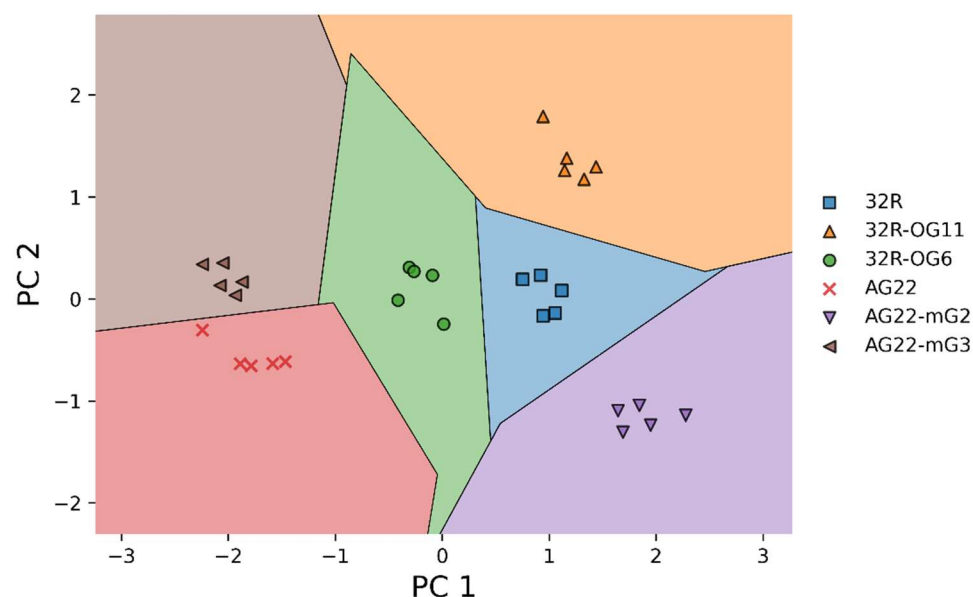


Figure S-47. SVM decision region boundary of PCA plot using the data from Figure S-46 for the classification of *AG22* and *32R* series DNA sequences in 10% single donor human serum off the clot (ISERS50ML-36688-23, Black Male, 29).

8.4 Repeats of 3-element Array for AG22 and 32R Series Sensing in Different Serum

8.4.1 Pooled human serum off the clot (ISER50ML-36670)

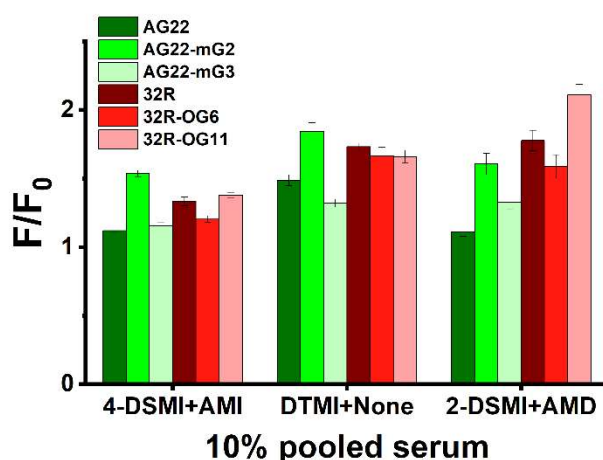


Figure S-48. Bar plot of *AG22* and *32R* series DNA sequences in 10% serum, obtained with the 3-element array: **2-DSMI+AMD** Ex/Em = 480/580 nm; **DTMI+None** Ex/Em = 540/600 nm; **4-DSMI+AMI** Ex/Em = 480/600 nm. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, 10% serum in 10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4. The pooled human serum off the clot (ISER50ML-36670) was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K^+ buffer.

Table S-11. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of 6 DNAs: *AG22* and *32R* series in 10% pooled human serum off the clot (ISER50ML-36670) using the data from Figure S-48.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	1.0000 (0.0000)
Sensitivity	1.0000 (0.0000)
Specificity	1.0000 (0.0000)
Precision	1.0000 (0.0000)
F1 Score	1.0000 (0.0000)
AUC	1.0000 (0.0000)

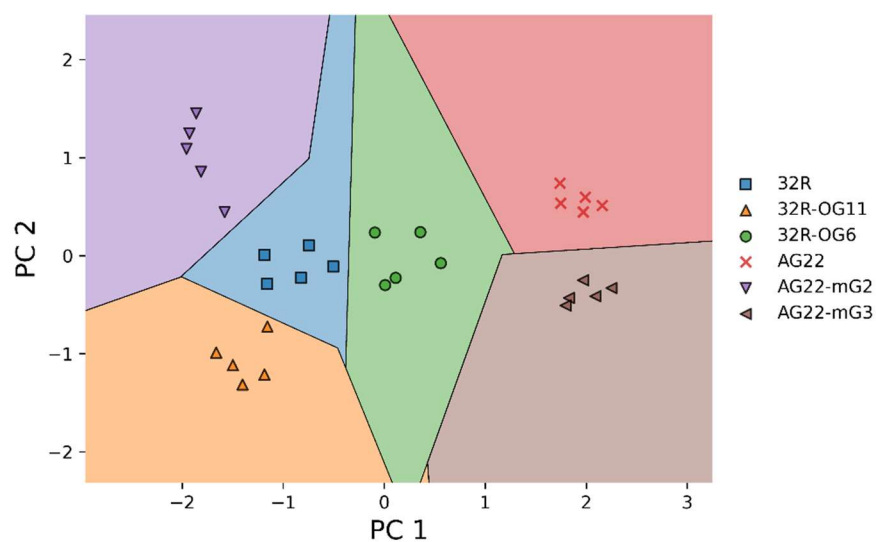


Figure S-49. SVM decision region boundary of PCA plot using the data from Figure S-48 for the classification of *AG22* and *32R* series DNA sequences in 10% pooled human serum off the clot (ISER50ML-36670).

8.4.2 Fetal bovine serum (10437028)

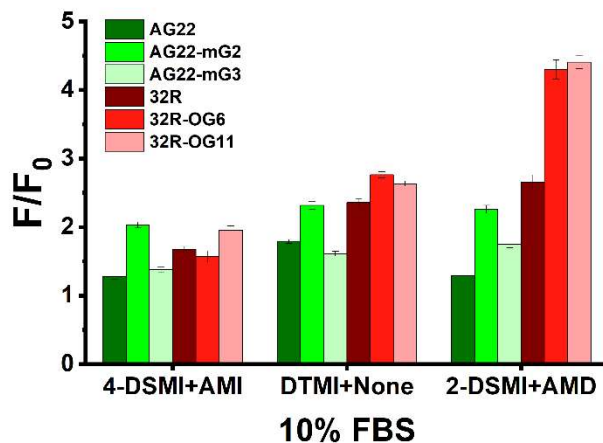


Figure S-50. Bar plot of *AG22* and *32R* series DNA sequences in 10% FBS, obtained with the 3-element array: **2-DSMI+AMD** Ex/Em = 480/580 nm; **DTMI+None** Ex/Em = 540/600 nm; **4-DSMI+AMI** Ex/Em = 480/600 nm. [Dye] = 0.625 μ M, [Host] = 0.625 μ M, [DNA] = 0.1 μ M, 10% serum in 10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4. FBS was centrifuged at 10K speed for 15 min to form a clear liquid; this process was repeated for 3 times, then the serum was diluted in K^+ buffer.

Table S-12. Performance metrics of 3 repeated 4-fold cross validation with SVM as the estimator for classification of 6 DNAs: *AG22* and *32R* series in 10% FBS using the data from Figure S-50.

Evaluation Metrics	Score (standard deviation from 3 repeated running of the 4-fold cross validation)
Accuracy	1.0000 (0.0000)
Sensitivity	1.0000 (0.0000)
Specificity	1.0000 (0.0000)
Precision	1.0000 (0.0000)
F1 Score	1.0000 (0.0000)
AUC	1.0000 (0.0000)

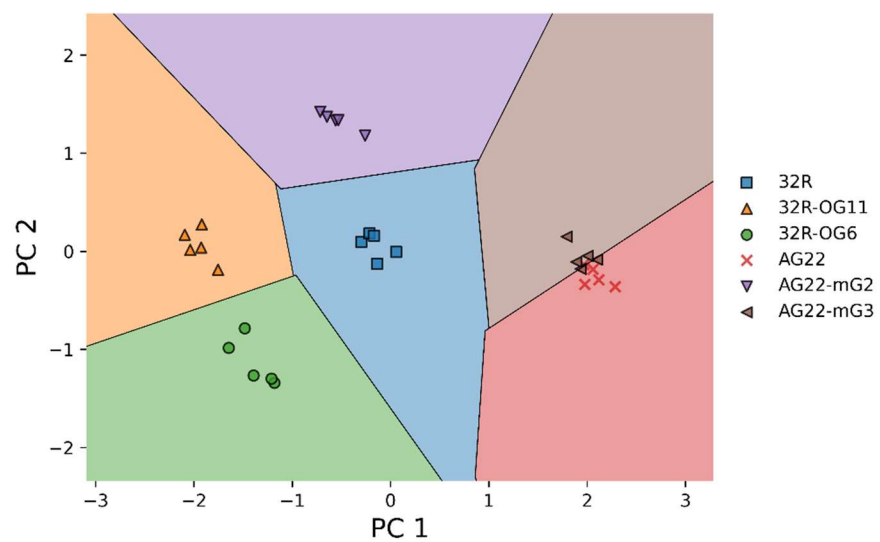


Figure S-51. SVM decision region boundary of PCA plot using the data from Figure S-50 for the classification of *AG22* and *32R* series DNA sequences in 10% FBS.

9. UV-Vis Absorbance Spectra

9.1 UV-Vis Spectra of 32R series with 2-DSMI+AMI

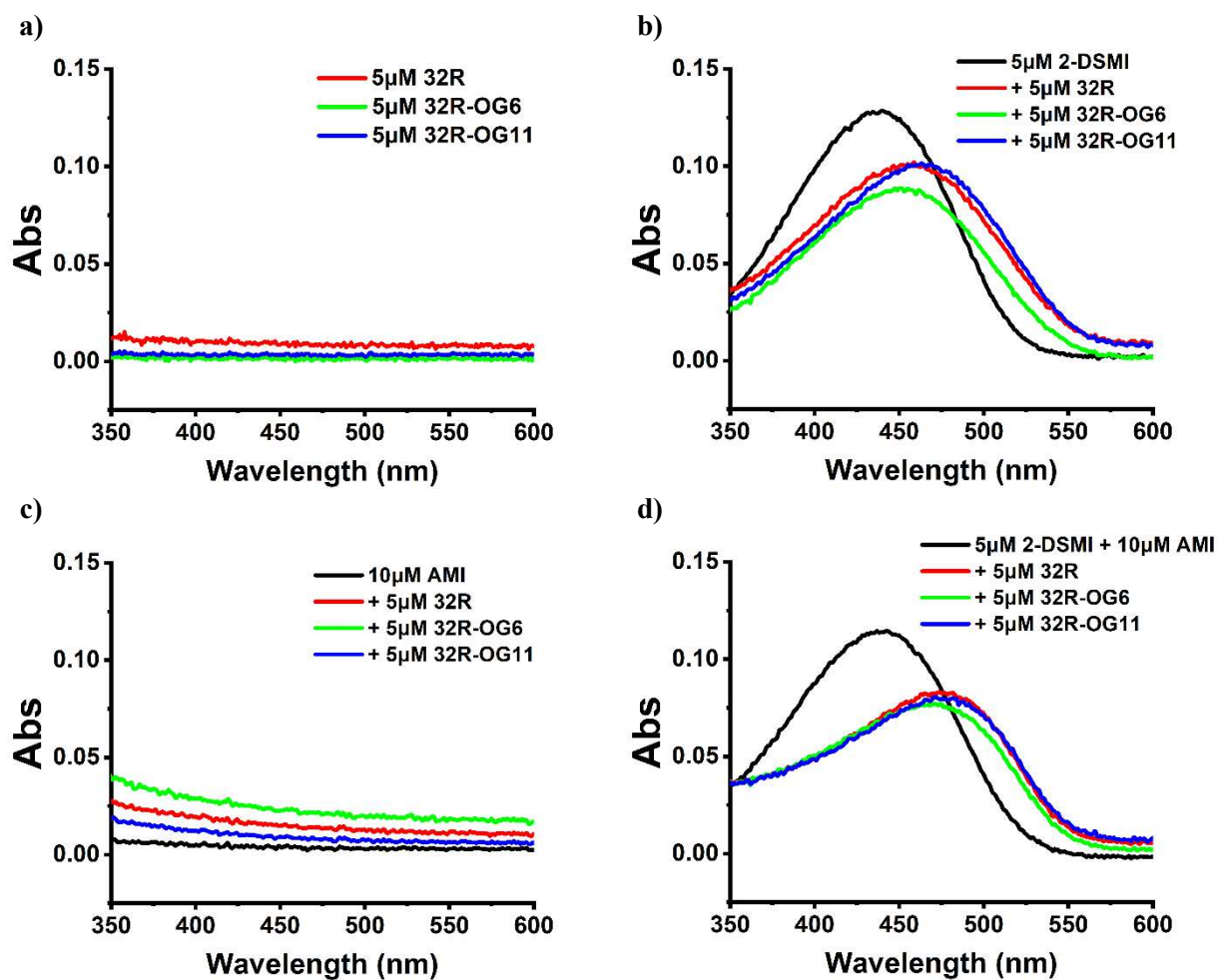


Figure S-52. UV spectra of a) 32R series DNA: 32R/ 32R-OG6/ 32R-OG11, b) 2-DSMI + 32R series DNA, c) AMI + 32R series DNA, and d) 2-DSMI+AMI complex + 32R series DNA. [2-DSMI] = 5 μ M, [AMI] = 10 μ M, [DNA] = 5 μ M, K^+ buffer (10mM K_2HPO_4/KH_2PO_4 , 1mM EDTA, pH 7.4).

9.2 UV-Vis Spectra of AG22 series with 2-DSMI+AMI

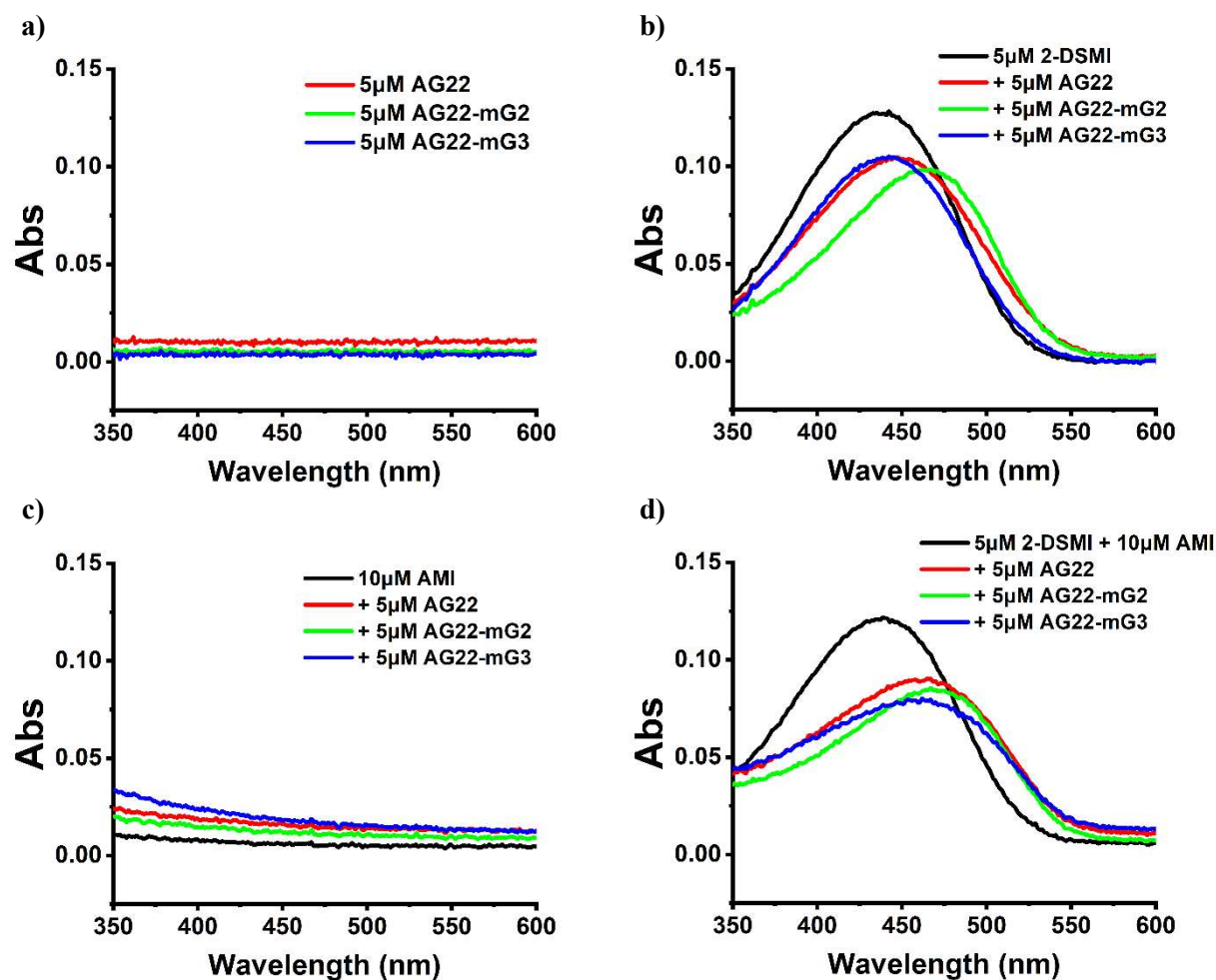


Figure S-53. UV spectra of a) AG22 series DNA: AG22/ AG22-mG2/ AG22-mG3, b) 2-DSMI + AG22 series DNA, c) AMI + AG22 series DNA, and d) 2-DSMI+AMI complex + AG22 series DNA. [2-DSMI] = 5 μ M, [AMI] = 10 μ M, [DNA] = 5 μ M, K⁺ buffer (10mM K₂HPO₄/KH₂PO₄, 1mM EDTA, pH 7.4).

10. References

- (1) 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.* **2004**, *126*, 2870-2876.
- (2) Mosca, S.; Yu, Y.; Rebek, J. Preparative scale and convenient synthesis of a water-soluble, deep cavitand. *Nat. Protoc.* **2016**, *11*, 1371-1387.
- (3) Gill, A. D.; Hickey, B. L.; Wang, S.; Xue, M.; Zhong, W.; Hooley, R. J. Sensing of citrulline modifications in histone peptides by deep cavitand hosts. *Chem. Commun.* **2019**, *55*, 13259–13262.
- (4) Mekmaysy, C. S.; Petraccone, L.; Garbett, N. C.; Ragazzon, P. A.; Gray, R.; Trent, J. O.; Chaires, J. B. Effect of O6-Methylguanine on the Stability of G-Quadruplex DNA. *J. Am. Chem. Soc.* **2008**, *130*, 6710-6711.
- (5) Redstone, S. C. J.; Fleming, A. M.; Burrows, C. J. Oxidative Modification of the Potential G-Quadruplex Sequence in the PCNA Gene Promoter Can Turn on Transcription. *Chem. Res. Toxicol.* **2019**, *32*, 437-446.
- (6) Cogoi, S.; Ferino, A.; Miglietta, G.; Pedersen, E. B.; Xodo, L. E. The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription. *Nucleic Acids Res.* **2017**, *46*, 661-676.
- (7) He, Y.-D.; Zheng, K.-W.; Wen, C.-J.; Li, X.-M.; Gong, J.-Y.; Hao, Y.-H.; Zhao, Y.; Tan, Z. Selective Targeting of Guanine-Vacancy-Bearing G-Quadruplexes by G-Quartet Complementation and Stabilization with a Guanine–Peptide Conjugate. *J. Am. Chem. Soc.* **2020**, *142*, 11394-11403.
- (8) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135*, 5017-5028.
- (9) Chen, J.; Hickey, B. L.; Wang, L.; Lee, J.; Gill, A. D.; Favero, A.; Pinalli, R.; Dalcanale, E.; Hooley, R. J.; Zhong, W. Selective discrimination and classification of G-quadruplex structures with a host–guest sensing array. *Nat. Chem.* **2021**, *13*, 488-495.