A rapid fluorescent indicator displacement assay and principal component/cluster data analysis for determination of ligand–nucleic acid structural selectivity

Rafael del Villar-Guerra, Robert D. Gray, John O. Trent* and Jonathan B. Chaires*

James Graham Brown Cancer Center, University of Louisville, 505 S. Hancock St., Louisville, KY 40202, USA

ABSTRACT

We describe a rapid fluorescence indicator displacement assay (R-FID) to evaluate the affinity and the selectivity of compounds binding to different DNA structures. We validated the assay using a library of 30 well-known nucleic acid binders containing a variety chemical scaffolds. We used a combination of principal component analysis and hierarchical clustering analysis to interpret the results obtained. This analysis classified compounds based on selectivity for AT-rich, GC-rich and G4 structures. We used the FID assay as a secondary screen to test the binding selectivity of an additional 20 compounds selected from the NCI Diversity Set III library that were identified as G4 binders using a thermal shift assay. The results showed G4 binding selectivity for only a few of the 20 compounds. Overall, we show that this R-FID assay, coupled with PCA and HCA, provides a useful tool for the discovery of ligands selective for particular nucleic acid structures.

INTRODUCTION

Nucleic acids are dynamic biomolecules that can adopt a wide variety of structures (1–3). In addition to the familiar duplex structure, nucleic acids can form a variety of non-canonical triplex (4,5), quadruplex (6) or i-motif (3,7,8) structures. Some of these structures have been detected in living cells (9–11) where they may play key roles in a variety of biological processes, including replication, transcription, oncogene expression, and telomere functions (11–17). Identification of ligands that selectively bind to duplex, triplex, i-motif, and G-quadruplex (G4) structures is an important and challenging goal of drug discovery (12,18–32). G4 structures, in particular, have emerged as promising therapeutic targets for anti-cancer therapies (13,18,27–29,33–42). A major challenge is to identify compounds that bind specifically to one of the several possible G4 structures without binding to duplex DNA. Over the last decades, a number of ligands that bind G4 structures have been described (38,43). Unfortunately, none have successfully progressed completely through clinical trials to become useful drugs.

This failure might be because most G4 binders found to date are polyaromatic compounds with poor drug-like properties and nonselective binding (42). Indeed, few G4 binders have been rigorously evaluated for selectivity with respect to their interaction with other nucleic acid structures, in part because there is no convenient assay for evaluating their structural-selective binding. A number of potential binding assays have been developed including chip-based (44), ligand fishing (45), FRET melting (46) and equilibrium dialysis (21,23,25). However, these methods may be expensive, time consuming or not amenable to high-throughput or even rapid screening.

The fluorescent indicator displacement assay (FID) (27,28,47–56), which relies on displacement of a fluorescent DNA-binding ligand by the test compound, was adapted for the discovery of new G4 binding ligands. However, most previous FID studies evaluated binding only to a single G4 structure, but did not assess binding to other DNA structures or to different G4 folds. Here, we describe an efficient and inexpensive rapid FID assay (R-FID) designed to explore the structural selectivity of binding. The assay measures the affinity and selectivity of a library of compounds against eight different DNA sequences and structures (four duplex, two triplex and two G4 structures). It was tested using a library of 30 compounds with defined DNA binding modes and preferences. We caution that all FID assays are indirect measures of compound binding, requiring competition with the prebound signal ligand. While test compound binding is registered in favorable cases, the competition reaction must always be kept in mind since it may lead to underestimation of binding affinity.

*To whom correspondence should be addressed. Tel: +1 502 852 1172; Email: j.chaires@louisville.edu
Correspondence may also be addressed to John O. Trent. Tel: +1 502 852 2194; Email: john.trent@louisville.edu

Importantly, we also developed a chemometric approach for evaluation of the large data set to facilitate visualization of the affinity, structural selectivity and binding mode of the compound library. Traditional graphical methods for the analysis of large data sets are often ineffective for easily visualizing binding selectivity. Data analysis therefore becomes a bottleneck in drug discovery. We present here a novel application of principal component analysis (PCA) and hierarchical cluster analysis (HCA) as powerful tools for analysis of large FID data sets. We demonstrate that PCA and HCA can be used to visualize binding properties and to clearly identify the binding selectivity of a library of compounds. This visualization provides a direct and easily interpretable method to depict complex, often hidden, relationships between affinity and selectivity.

One may question why yet another variation of the FID assay needed. As part of a drug discovery platform it is essential, once a ligand is found that hits the desired target, to evaluate its binding specificity. We envision this FID assay as a rapid secondary screen for that purpose. Primary high-throughput screening of large compound libraries for target hits can be reliably done by thermal shift assays, particularly differential scanning fluorometry. The FID assay we describe will fill an important need by providing rapid specificity screen to identify the best candidates to continue along the drug discovery platform.

MATERIALS AND METHODS

Preparation of oligonucleotides samples

The sequences, names of the oligonucleotides and buffers used in this study are given in Table 1. Oligonucleotides were purchased as desalted lyophilized powders from Integrated DNA Technologies, Inc. (Coralville, IA) and used without further purification. Stock solutions were prepared at 2 mM concentration in Milli-Q water and stored at 4°C for a least 24 h before use. Further dilutions to working concentrations were prepared using the following buffers optimized for specific DNA structures: Buffer A (6 mM K_2HPO_4, 4 mM KH_2PO_4, 15 mM KCl, pH 7.2), Buffer B (1 mM K_2HPO_4, 9 mM KH_2PO_4, 15 mM KCl, pH 6.2), Buffer C (6 mM K_2HPO_4, 4 mM KH_2PO_4, 15 mM KCl, 1 mM $MgCl_2$, pH 7.2). To induce quadruplex, triplex or duplex formation, stocks solutions of each oligonucleotide were diluted with the appropriate buffer to ∼200 μM and annealed by heating in a boiling water bath for 20 min followed by overnight cooling to room temperature. The annealed samples were stored at 4°C. DNA strand concentrations were estimated from the absorbance at 260 nm after thermal denaturation for 5 min at 90°C using molar extinction coefficients supplied by the manufacture. Oligonucleotide solutions that contained 1% (v/v) dimethyl sulfoxide (DMSO) were prepared by adding the necessary volume of DMSO to the previously annealed nucleic acid solution. All nucleic acid structures were characterized by circular dichroism and UV absorbance spectroscopies, and by thermal denaturation (see Supplementary Material).

Sample preparation

Thiazole Orange (TO) and DMSO (99.9% purity) were purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. The test compounds were purchased from commercial sources and were used without further purification. Stock solutions of test compounds were prepared by dissolving a weighed amount in DMSO to give a concentration of 1–10 mM and stored in the dark at −20°C to prevent light-induced degradation. Working solutions of test compounds were prepared immediately before use at 15 μM concentration in the appropriate buffer containing 1% (v/v) DMSO. Serial dilutions of the stock ligand solutions contained 1% (v/v) DMSO. Stock solutions for FID screening were made in the same buffer used for the DNA preparation (Table 1). Oligonucleotide stock solutions (6 μM) were prepared in the appropriate buffer solutions with 1% (v/v) of DMSO using the pre-folded oligonucleotide stock solution (200 μM) prepared as described above.

Thiazole orange displacement assay

FID assays were conducted in duplicate at room temperature (20°C) without control with a Tecan Safire[2] microplate reader (Tecan US, Durham, NC, USA) with the following parameters: excitation and emission bandwidth of 5 nm, gain of 125, integration time of 200 μs and 16 reads. Emission spectra were measured at 1 nm intervals from 510 to 750 nm with an excitation wavelength of 500 nm. Assays were carried out in 96-well NBS™ black, flat bottom polystyrene microplates (Cat# 3650, Corning Inc., NY, USA). Each assay solution contained 1 μM TO, 2 μM oligonucleotide and 5 μM test compound in 150 μl. The instrumental gain was adjusted using the fluorescence of the most fluorescent sample (1 μM TO and 2 μM AT triplex 5). A representation of a 96-well plate assay is depicted in Supplementary Figure S1. This configuration allowed testing five compounds and eight oligonucleotide structures per plate along with appropriate controls.

The percentage TO displacement (%FID) was calculated from the corrected fluorescence emission intensity at 527 nm (λ_{ex} = 500 nm) using Equation (1). The fluorescence intensity F was corrected by subtracting the fluorescence signal of the test compound/DNA mixture and that of unbound TO:

$$\begin{aligned}
\%\text{FID} &= 100 - \left(100 \times \frac{F}{F_0}\right) \\
F &= F_{(\text{Ligand+DNA+TO})} - F_{(\text{Buffer+TO})} - F_{(\text{DNA+Ligand})} \\
F_0 &= F_{(\text{DNA+TO})} - F_{(\text{Buffer+TO})}
\end{aligned} \tag{1}$$

The %FID was calculated using the fluorescence intensity at the TO emission maximum of 527 nm rather than using the integrated intensity over the wavelength range 510–750 nm as has been used in previous implementations of the FID method (28,47,49,54,57). We observed that interference can be introduced when the integrated intensity is used instead of the single wavelength intensity, especially when fluorescent compounds are tested (Supplementary Figures S2–3).

Table 1. Oligonucleotides sequence and their molar extinction coefficients used in this study

ID	Name	Sequence[a]	ε ($M^{-1}cm^{-1}$) 260 nm[b]	Buffer[c]
1	AT Duplex	5′-ATA TAT ATC CCC ATA TAT AT-3′	205700	A
2	GC Duplex	5′-GCG CGC GCT TTT GCG CGC GC-3′	166300	A
3	A/T Duplex	5′-AAA AAA AAC CCC TTT TTT TT-3′	191300	A
4	H-Tel Duplex	5′-GGG TTA GGG TTT TCC CTA ACC C-3′	202200	A
5	AT Triplex	5′-AAA AAA AAC CCC TTT TTT TTC CCC TTT TTT TT-3′	284900	C
6	GC Triplex	5′-CCC CCC CCT TTT GGG GGG GGT TTT CCC CCC CC-3′	261600	B
7	H-Tel i-Motif	5′-CCC TAA CCC TAA CCC TAA CCC T-3′	193700	B
8	H-Tel Quadruplex	5′-AGG GTT AGG GTT AGG GTT AGG G-3′	228500	A
9	c-Myc Quadruplex	5′-TGA GGG TGG GTA GGG TGG GTA A-3′	228700	A
10	c-Myc i-Motif	5′-TTC CCC ACC CTC CCC ACC CTC CCC TAA-3′	222200	B
11	A/U RNA duplex	5′-rArArA rArArA rArArC rCrCrC rUrUrU rUrUrU rUrU-3′	202900	C
12	DNA-RNA duplex	5′-GGG TTA GGG TTT TrCrC rCrUrA rArCrC rC-3′	202900	A

[a]The strand orientation is 5′ to 3′ from left to right, and 'r' denotes a ribonucleotide.
[b]Molar nearest neighbor extinction coefficients of the single-strand at 260 nm.
[c]Buffer conditions: A = 6 mM K_2HPO_4, 4 mM KH_2PO_4, 15 mM KCl, pH 7. 2; B = 1 mM K_2HPO_4, 9 mM KH_2PO_4, 15 mM KCl, pH 6.2; C = 6 mM K_2HPO_4, 4 mM KH_2PO_4, 15 mM KCl, 1 mM $MgCl_2$, pH 7. 2.

Principal component analysis (PCA) and cluster analysis

Multivariate data analysis, principal component analysis, and cluster analysis of the FID assay data were performed with R software version 3.2.3 (10 December 2015), using the R package FactoMineR (58). These programs are freely available from the Comprehensive R Archive Network (CRAN) at http://cran.r-project.org.

Hierarchical clustering on the principal components (59) was performed on the principal components using the Euclidean distance as a measure of similarity between individuals with the Ward criterion as agglomeration method. The initial clusters obtained by hierarchical clustering were further consolidated by partitional clustering using the *k*-means algorithm. The results of the cluster analysis are presented as dendrograms in which the horizontal axis represents the compounds and the vertical axis represents the degree of similarity of the individuals.

RESULTS AND DISCUSSION

Oligonucleotide design and characterization

A library of 12 oligonucleotides that form representative duplex, triplex, i-motif and G4 structures was designed for use in the new FID assay. Unimolecular structures were selected to avoid potential problems related to concentration and the molecularity of folding. Hairpin duplex structures are represented by deoxyribo-oligonucleotides 1–4 and the RNA oligonucleotides 11–12. These structures represent B-form DNA with AT- and GC-rich base compositions, an A-form RNA and a DNA–RNA hybrid. Sequences 5 and 6 form AT and GC intramolecular triplex DNA structures, respectively. These triplex structures (H-DNA structures) are important in several biological processes (60). Oligonucleotides 7 and 10 represent i-motif structures (7,61) corresponding to the human telomeric sequence (62) and the c-MYC NHE III1 wild-type C-rich promoter sequence (3). Oligonucleotides 8 and 9 form representative G4 structures. H-Tel (8) is the human telomeric repeat sequence (63,64) that is important in cancer (16,39) and which is polymorphic depending on environmental conditions (63,65–67) An antiparallel hybrid ('3+1') structure is formed under the conditions of the assay. c-Myc (9) is a sequence variant of

the wild-type NHE III1 sequence of the c-Myc promoter that forms a homogeneous parallel G-quadruplex structure in potassium solution (68).

Each of these nucleic acid structures have different stabilities and require optimization of experimental conditions to ensure their formation. The influence of buffer composition, pH, DMSO and TO on the folding and stability of the structures was evaluated using UV-Vis absorption, circular dichroism (CD) spectroscopy (69–74), thermal difference spectra (TDS) (75), and UV/CD thermal denaturation (Supplementary Figures S4–S15). A summary of the physical and spectroscopic characteristics of the oligonucleotides under different buffer conditions is given in Supplementary Table S1.

The effect of TO and DMSO (used as a solvent for the ligands) on the stability and folding of the triplex oligonucleotides was also evaluated (Supplementary Figures S16–S18). CD and melting experiments showed that TO and DMSO stabilized the triplex structures without changing their conformation. In summary, these experiments show that all of the test oligonucleotides formed the desired structure and were sufficiently stable ($T_m > 25°C$) to be used in the FID assay.

The affinity of TO for the oligonucleotides in the structure array was determined by fluorimetric titrations carried out under the optimized buffer conditions (76). The results are summarized in Figure 1 and Supplementary Table S2. The corresponding binding isotherms and emission spectra are in Supplementary Figures S4–S15, panels H and I. The degree of fluorescence enhancement of TO binding depended on the DNA structure (Supplementary Table S2). In summary, the titrations showed that TO binds with similar affinity ($K_a \sim 9 \times 10^5$ to 6×10^6 M^{-1}) to all of the oligonucleotide structures with the exceptions of i-motifs 7 and 10 and RNA duplexes 11 and 12, where lower affinities ($K_a < 6 \times 10^5$ M^{-1}) were observed. Because of their low TO affinity, these structures were omitted from the test panel. A simple 1:1 DNA:TO model provided adequate fits to all of our titration data for the protocol used, except for c-Myc G4 where a secondary binding process was observed at high c-Myc:TO ratios. This weak apparent binding may result from TO-mediated quadruplex dimerization as recently
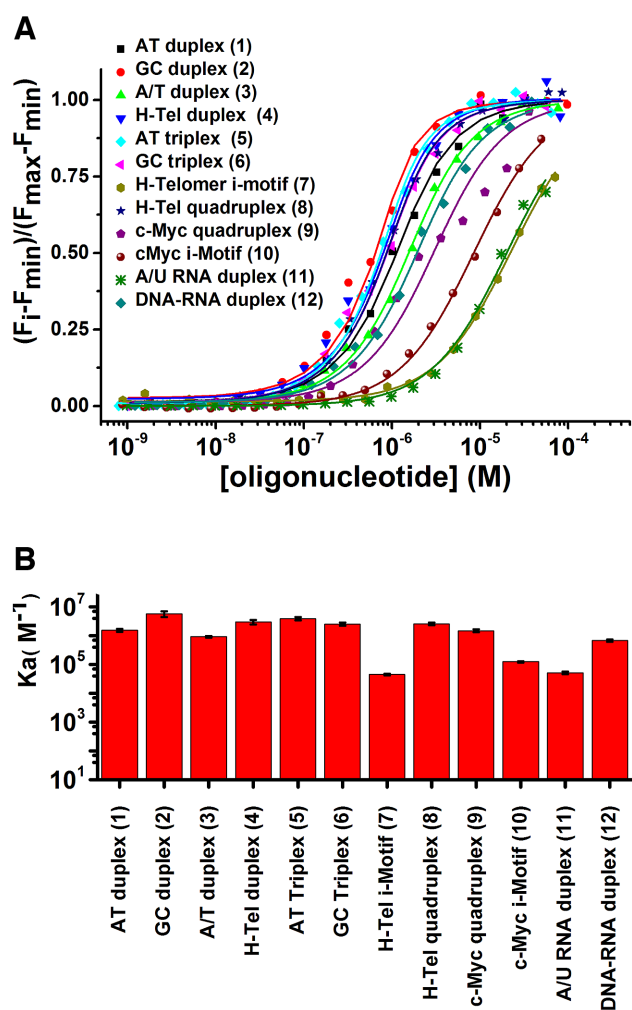
Figure 1. (**A**) Saturation curves for fluorescence titrations of TO with different oligonucleotides. (**B**) Bar graph showing binding constants of TO for different DNA structures used.

reported for some viral G4s (77). Since the contribution of the low affinity binding was negligible under the conditions of the FID assay, this low-affinity process was ignored in evaluating equilibrium constant for TO:c-Myc G4 interaction.

FID assay design

Assay conditions were optimized to give high sensitivity, a high signal-to-noise ratio, and to minimize the effects of different binding stoichiometries. Corrections were applied to account for any interference from intrinsic ligand fluorescence or from inner filter effects. To obtain the maximum sensitivity, the gain of the microplate reader was set to maximize the fluorescence of AT triplex 5 which exhibited the largest change in emission intensity. The reagent concentrations of 1 μM TO and 2 μM nucleic acid were selected because these concentrations gave ∼75% saturation (Figure 1), a point in the binding isotherm sensitive to competitive displacement and which minimizes potential problems from the binding of multiple probe molecules to the structure (26,56). The lower oligonucleotide/TO ratio used in this

work differs from previous assays where excess TO was used (27,28,49,55,56). The lower ratio ensures that only ligands with high affinity will be detected. Supplementary Figure S19 shows a simulation of the expected TO displacement as a function of ligand binding affinity. This simulation uses the exact concentrations in our assay and the experimentally determined TO binding constants for each structure and assumes 1:1 binding. The simulations show that under the conditions of our assay the onset of TO displacement requires a ligand affinity of $>10^5$ M^{-1}, and complete displacement requires a ligand affinity of $>10^8$ M^{-1}.

The effect of test compound concentration was also examined. The results show that good discrimination between oligonucleotides in the library was obtained at a compound concentration of 5 μM (Supplementary Figure S20).

We also assessed the inner filter effect of fluorescent ligands ethidium bromide, doxorubicin, NMM580, TmPyP2 and TmPyP4 and a non-fluorescent ligand, netropsin, which was selected as a negative control because its absorption spectrum does not overlap with the absorption or emission spectra of TO. The results show that the absorbance of these compounds at 5 μM is <0.06 at the wavelengths of TO excitation (500 nm) and emission (527 nm) (Supplementary Figures S21 and S22). Therefore, the inner filter effect is negligible. For instance, 5 μM doxorubicin or TmPyP2 have absorbances at 500 and 528 nm that are similar (Supplementary Figure S21) but their %FID profiles for individual duplex structures differs (Figure S20C and E). This shows that the difference in binding selectivity for these structures results from different affinities rather than an inner filter effect. In addition, the importance of assessing the spectroscopic properties of potential ligands for possible interference with the FID assay, is illustrated by a %FID > 100% for TmPyP4 (Supplementary Figure S20F). This apparent anomaly most likely results from fluorescence of the ligand at the emission wavelength of TO (Supplementary Figure S22).

FID assay: reference library

We validated the FID assay for nucleic acid structural selectivity with a reference library of 30 nucleic acid ligands tested against eight oligonucleotide structures. The ligand library contains various intercalators (e.g. ethidium bromide), minor groove binders (e.g. netropsin), triplex binders (e.g. coralyne) and G4 ligands (e.g. NMM) (Supplementary Table S3, Supplementary Figure S23). In addition, a variety of other chemical structures including heterocycles, carbohydrates, and peptides, each with different physical, chemical, spectroscopic and binding properties, were in the reference library.

The results of the FID assay of the 30 ligands binding to eight nucleic acid sequences and structures (four duplex, two triplex and two quadruplex) are presented as a bar graph in Figure 2. The data report the results of 240 individual binding interactions. Positive and a few negative values of %FID were observed. Positive values can be explained by a fluorescence decrease resulting from displacement of TO from the nucleic acid by the competitor ligand. In these cases, %FID is directly related to the affinity of a compound for a given nucleic acid structure. The higher the %FID, the
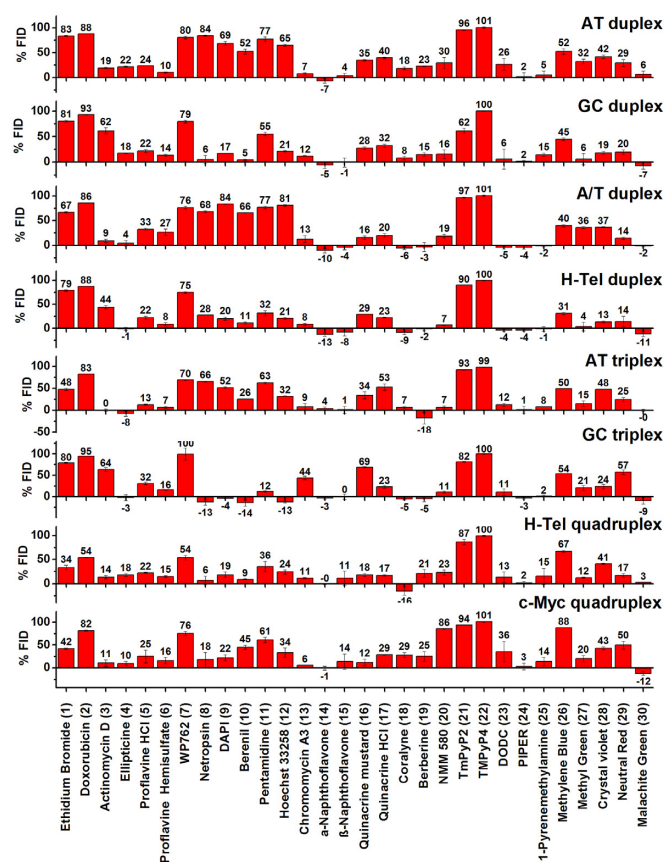
Figure 2. Results of FID assay performed with 30 nucleic acid binders and eight oligonucleotide structures. Each bar represents %FID. The error bars correspond to standard deviation of the %FID values from two different plate measurements.

higher affinity of the compound for a nucleic acid structure. For example, netropsin showed 84%FID for AT duplex 1 compared to 6% for GC duplex 2. This is consistent with the known selectivity of netropsin for AT over GC duplexes (23,25).

Negatives values of %FID are more difficult to rationalize (54). A number of factors might account for TO fluorescence enhancement leading to a negative %FID: (a) interaction between TO and the test compound; (b) increased binding affinity of TO induced by the test compound; c) similar photo-physical behavior of the compound with TO. To further investigate the negative %FID observed with α-naphtoflavone, coralyne and berberine interacting with the A/T duplex, the AT triplex and H-tel G4 structures, we determined absorbance, excitation and emission spectra of TO in the presence and absence of ligand under the same experimental conditions used for the FID (Figures S24–S28). Both α-naphthoflavone (Supplementary Figure S24) and coralyne (Supplementary Figures S25–S26) enhanced TO fluorescence both in the presence and absence of oligonucleotide. This accounts for the observed negative %FID and suggests an interaction between these compounds and TO. For berberine (Supplementary Figures S27–S28), significant differences were observed in the fluorescence spectra when the experiment was conducted in a quartz cu-

vette compared to a 96-well plate. The %FID observed for berberine using a quartz cuvette was –1%, in contrast to – 15% observed when a 96-well plate was used. This suggests non-specific binding of berberine to the plastic plate.

Statistical data analysis: principal component and cluster analysis

Traditional methods of reporting FID HTS data such as bar graphs are cumbersome for detailed specificity analysis since it is difficult to grasp global trends. A major goal of the present study was to develop an efficient, quantitative approach for more effectively evaluating ligand–nucleic acid interactions. The methods of principle component analysis and hierarchal cluster analysis provide the tools for such a method.

PCA is a multivariate statistical method to reduce the dimensionality of data without significant loss of information using derived variables (principal components) (78–81). This is a non-supervised classification method that performs unbiased clustering of the data. PCA is particularly useful for finding hidden patterns and aiding in interpretation of high dimension, complex data sets (82). PCA projects the multi-dimensional data onto a new lower dimension subspace. The axes in this new subspace, known as principal components (PCs), are linear combinations of the original variables with their origin located at the center of the multi-dimensional data. The PCs are independent (orthogonal) to each other and retain as much of the information in the original variables as possible. The first principal component (PC1) corresponds to the axis that contains the greatest variation of the data. The second principal component (PC2) is orthogonal to the first and oriented in the direction of the second greatest amount of variation of the data, and so on.

In addition to PCA, we used HCA as an unsupervised classification method to find clusters of compounds. HCA calculates the distance between observations in multidimensional space and forms clusters of individuals based on the similarity of their variables.

The results of PCA and HCA performed on the FID data of our reference library of 30 nucleic acids binders and 8 oligonucleotide structures are shown as biplot graphs, PC1–PC2 (Figure 3, left panel) and PC2–PC3 (Figure 3, right panel) (83). The variables (oligonucleotides) are represented by arrows and the individuals (compounds) by points colored according to their cluster. (We note that these plots were constructed using mean values of %FID. As one referee suggested, unaveraged primary data could be used in these graphs to show the possible variability in the clusters.) The results of HCA performed on the PCs, after consolidation by an additional partitional clustering with a K-means algorithm (84), are shown as dendrograms in Supplementary Figures S29 and S30. The horizontal axis represents the compounds and the vertical axis measures the distance or similarity between individuals in the clusters. The more similar the affinity and selectivity of the individual ligands for the different DNA structures, the lower are the nodes in the tree. The levels at which the hierarchical tree was cut to initialize the k-means algorithm are represented by rectangles.
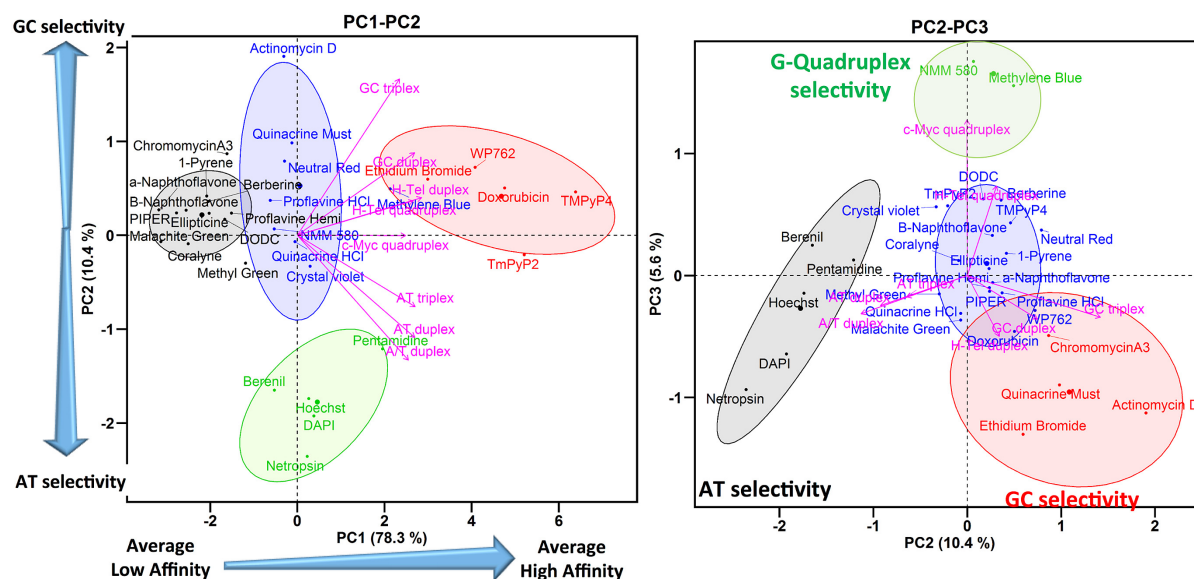
Figure 3. PCA and HCA of FID results for 30 nucleic acid binders and 8 oligonucleotide structures. *Left*: biplot of PC1 and PC2 *Right*: biplot of PC2 and PC3. The center of each cluster is represented by larger point size.

Interpretation of the PC1–PC2 biplot

Four well-separated clusters are observed in the biplot of PC1–PC2 (Figure 3, left), which explains 89% of the variance of the data. Cluster 1 (black) includes the compounds that exhibit low or null capacity for displacement of TO (low %FID) from all the nucleic acid structures. Cluster 2 (blue) and cluster 3 (green) contain compounds situated close to the origin of PC1 and which showed medium affinity. Cluster 3 (green) contains minor groove binders, pentamidine, netropsin, berenil, DAPI and Hoechst, that have a selectivity for AT sequences (23,25). Cluster 4 (red) contains DNA intercalators, ethidium bromide, doxorubicin, WP762 and porphyrins TmPy4, TmPy2, that exhibit high values of %FID for almost all of the DNA structures.

Analysis of the loading factors (the magenta arrows) reveals that all of the oligonucleotides have a positive correlation and similar contribution to PC1 as indicated by the length and direction of projection of the variables on to PC1 (Figure 3, left). The biplot PC1–PC2 shows that oligonucleotides with similar AT or GC base composition cluster together along PC2. Low correlation between AT-rich (AT triplex, AT duplex and A/T duplex) and GC-rich (GC triplex and GC duplex) oligonucleotides is observed along PC2. H-Tel duplex, H-Tel G4 and c-Myc G4, with similar AT and GC content (∼50% AT), are situated between the AT-rich and GC-rich oligonucleotides. PC1 therefore can be interpreted in terms of global affinity for different DNA structures and PC2 can be interpreted as showing different ligand affinities for AT- or GC-rich oligonucleotides.

The projection of a compound onto a particular oligonucleotide vector is interpreted as its relative affinity for a particular DNA structure. For instance, the projections of the compounds WP762, methyl blue, NMM580, methyl green onto the GC duplex vector within the PC1–PC2 biplot (Figure 3, left), shows that the order of affinity for GC duplex (2) is approximately WP762 > methyl blue > NMM580 >

methyl green, which agrees with the affinity ranking obtained from the %FID.

In summary, the biplot PC1–PC2 allows easy visualization and classification of compounds based on their affinity for the DNA structures. PC1 separates the compounds based on their global affinity and PC2 separates them based on their AT or GC affinity. In this way, compounds situated at the left, center and right correspond to compounds with high, medium or low affinity for the different DNA structures.

Interpretation of the PC2–PC3 biplot

The PC2–PC3 biplot (Figure 3, right) projection provides visualization of the selectivity of a compound for a particular DNA structure. Three groups of variables hidden in the PC1–PC2 biplot are now observed in the PC2–PC3 biplot (Figure 3, right). Each group of variables points in different directions with an angle between them of ∼120°. The first group contains AT-rich structures (AT duplex, A/T duplex and AT triplex) and points to the bottom left corner of the PC2–PC3 biplot. The second group contains GC-rich structures (GC duplex, GC triplex and H-Tel duplex) and points to the bottom right corner. The third group contains only the G4 structures H-Tel and c-Myc and points in a positive direction of PC3. This orientation of the variables reflects differences in the structures of these oligonucleotides. Therefore, information about the selectivity for a particular DNA structure can be obtained from the location of the compounds in the PC2–PC3 biplot.

The biplot PC2–PC3 shows four well-separated clusters of compounds. Cluster 1 (black) contains the minor groove binders pentamidine, netropsin, berenil, DAPI and Hoechst that are selective for AT in duplexes and triplexes (25). Cluster 2 (red) contains the DNA intercalators ethidium bromide, actinomycin D, doxorubicin, WP762 and also the minor groove ligand chromomycin A3 that prefers GC

base pairs (85). These compounds have higher affinity for GC-rich DNA structures (25). Cluster 3 (green) contains NMM580 and methylene blue that are selective for G4s (21,25,86,87). Cluster 4 (blue) includes the remaining compounds that have low selectivity. To summarize, the PC2–PC3 biplot shows a higher level of discrimination between different DNA structures. PC2 separates compounds based on AT or GC base composition and PC3 discriminates based on G4 binding.

Application to the NCI diversity set III library

In order to demonstrate the utility of this FID assay as a secondary screen in the context of a drug discovery platform, we used the NCI diversity set III library that contains 1598 compounds as a test. As a primary screen of the library, differential scanning fluorometry was used to identify compounds that bound to a FRET-labeled G4 structure, a human telomere sequence 5′AGGG(TTAGGG)$_3$ in the hybrid form. Supplementary Figure S31 shows the distribution of T_m shifts obtained, with several compounds with $\Delta T_m > 10$ degrees, indicative of tight binding. We selected 20 of these with T_m shifts ranging from 6 to 31 degrees (Supplementary Figures S31–S32) for secondary screening using R-FID, based on their apparent affinity and compound availability. The results of this screen are shown in Figure 4.

To our surprise, not all of the 20 compounds that registered as hits in the thermal shift assay were able to displace TO from the unlabeled G4 hybrid structure used in our FID assay, even though the identical sequence was used in both assays. No correlation between the magnitude of ΔT_m and the %FID recorded (Supplementary Figure S33) was observed. To understand this result, several differences between the FID and thermal shift assays need to be emphasized. First, the thermal shift assay requires a FRET labeled G4 structure, which might influence binding, while the FID assay does not. Second, the thermal shift assay uses a large molar excess of ligand over G4 (250:1), while a much lower molar ratio (5:1), is used in the FID assay, perhaps resulting in differing extents of ligand binding. Third, the FID assay is by design a competition binding assay and the apparent ligand affinity is relative to the affinity of the TO probe and is reduced from the true affinity. Recall that, as discussed above, a ligand binding constant of $>10^5$ M^{-1} is required to register TO displacement in the FID assay.

Even with these considerations, the lack of correlation between the thermal shift and FID assays remains puzzling. Additional explanations for the differences are needed. First, it is possible that the ligand may bind to a different site than TO and instead of displacement, a ternary (ligand-G4-TO) complex may form. For example, if TO is bound by an 'end pasting' mode and stacked on a terminal G-quartet, the ligand might bind within a groove and not perturb the bound TO. Second, it is possible that the FRET labels in the thermal shift assay might somehow form a binding pocket to facilitate ligand binding, and since they are not present in the FID assay, no ligand binding site would be available.

We conclude that a positive FID confirms and validates ligand binding registered by the primary thermal shift assay, but that a valid initial hit might be invisible in the FID assay. Since the intent of our FID assay is as a secondary screen
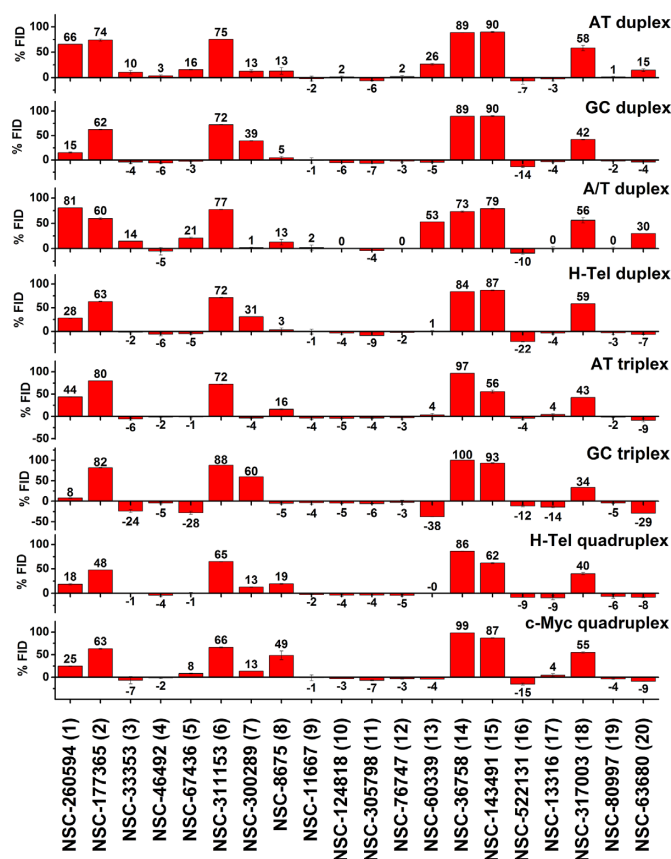

Figure 4. Results of the FID assay for the 20 compounds with the highest T_m shift from the NCI diversity set III library and eight oligonucleotide structures. Each bar represents the %FID of a given compound for an oligonucleotide sequence. The error bars correspond to standard deviation of the %FID values from two different plate measurements.

to assess binding selectivity, we focused further analysis on those compounds that showed appreciable FID ($>10\%$) in their interaction with the human telomere hybrid G4. Five of these eight compounds with an FID response had $\Delta T_m > 25°C$ in the primary screen.

The most striking result to emerge from inspection of Figure 4 is that only one of the eight G4 hybrid binders exhibited any specificity. Compounds 2, 6, 14, 15 and 18 (Figure 4, Supplementary Figure S32) bind to all structures in the FID assay. Compound 1 has a selectivity for AT-oligonucleotides and a shape similar to the classical minor groove binders. Compound 7 is interesting because it shows a clear preference for GC-oligonucleotides, with significant binding to the GC triplex structure and GC-rich duplexes in addition to binding to the two quadruplex structures. Compound 8 emerges as the most selective compound for G4 structures. Interestingly, although it was selected for binding to an antiparallel hybrid G4 structure, it clearly binds preferentially to the c-myc parallel G4 structure. It shows some binding to AT-rich triplex and duplex structures, but not to GC-rich duplex or triplex structures. The key point is that the FID assay fulfilled its intended purpose as a secondary screen for binding specificity. While it is disappointing that most hits from the primary thermal shift assay seem
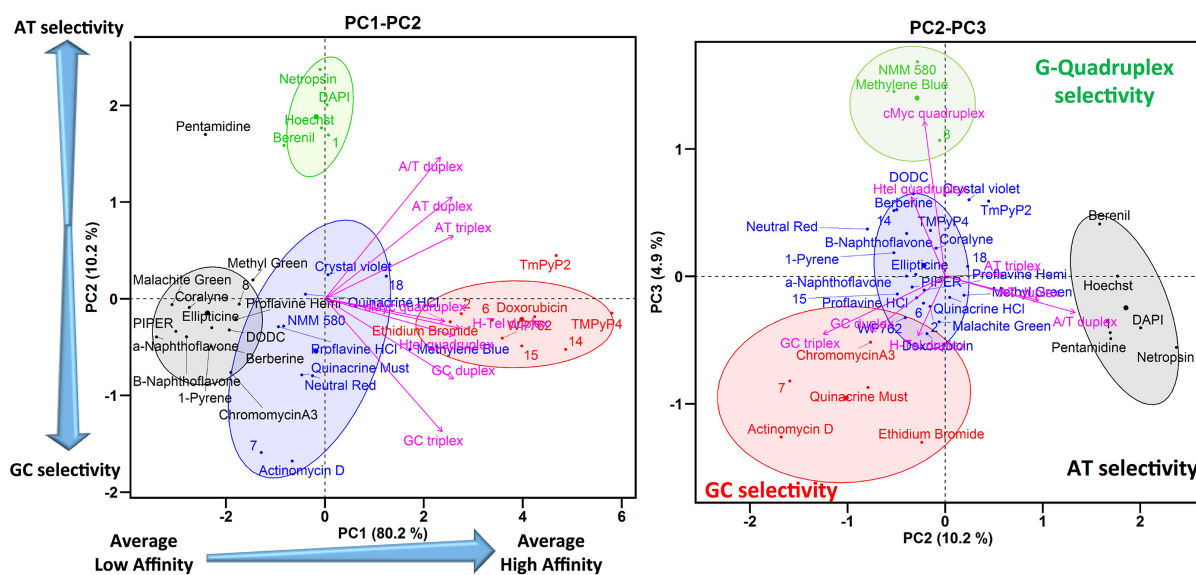
Figure 5. PCA and HCA of FID results for the combination of 30 nucleic acids binders and the eight compounds with the highest thermal shift (ΔT_m) and % FID for H-Tel quadruplex from the NCI diversity set III library with 8 oligonucleotide structures. *Left*: biplot of PC1 and PC2. *Right*: biplot of PC2 and PC3. The center of each cluster is represented by larger point size.

to lack specificity, it is essential to find that out sooner rather than later to avoid further development of less than ideal compounds for targeting G4 structures. Perhaps such lack of specificity is a contributing reason for the unsuccessful discovery of G4 targeted drugs so far.

A PCA analysis provides further insight into the selectivity of the eight compounds, with the results shown in Figure 5 (Supplementary Figures S34–S35). In this analysis, the FID data of Figure 4 for the eight compounds showing TO displacement for the hTel G4 was combined with the FID data for the 30 reference compounds. The interesting results that emerge are that compound 8 clusters with known quadruplex binders, compound 1 clusters with known groove binders, and compound 7 clusters with known GC-specific intercalators. The remaining five compounds have little or no selectivity. These results illustrate the utility of the FID assay and PCA analysis for defining specificity and possible binding modes. Inspection of the structure of compound 8 shows that it is similar to crystal violet that is in our set of reference compounds, but seems to have improved selectivity for G4 binding.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Neidle,S. (2007) *Principles of Nucleic Acid Structure*. Academic Press, Inc., San Diego.
2. Sinden,R.R. (1994) *DNA Structure and Function*. Academic Press Inc., San Diego.
3. Dai,J., Hatzakis,E., Hurley,L.H. and Yang,D. (2010) I-Motif structures formed in the human c-MYC promoter are highly dynamic–insights into sequence redundancy and i-motif stability. *PLoS ONE*, **5**, e11647.
4. Frank-Kamenetskii,M.D. and Mirkin,S.M. (1995) Triplex DNA Structures. *Annu. Rev. Biochem.*, **64**, 65–95.
5. Felsenfeld,G., Davies,D.R. and Rich,A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.*, **79**, 2023–2024.
6. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
7. Guéron,M. and Leroy,J.-L. (2000) The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.*, **10**, 326–331.
8. Völker,J., Klump,H.H. and Breslauer,K.J. (2007) The energetics of i-DNA tetraplex structures formed intermolecularly by d(TC5) and intramolecularly by d[(C5T3)3C5]. *Biopolymers*, **86**, 136–147.
9. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
10. Mergny,J.-L. (2012) Alternative DNA structures: G4 DNA in cells: itae missa est? *Nat. Chem. Biol.*, **8**, 225–226.
11. Buske,F.A., Mattick,J.S. and Bailey,T.L. (2011) Potential in vivo roles of nucleic acid triple-helices. *RNA Biol.*, **8**, 427–439.
12. Han,H. and Hurley,L.H. (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.*, **21**, 136–142.
13. Miller,K.M. and Rodriguez,R. (2011) G-quadruplexes: selective DNA targeting for cancer therapeutics? *Expert Rev. Clin. Pharmacol.*, **4**, 139–142.
14. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
15. Belotserkovskii,B.P., Mirkin,S.M. and Hanawalt,P.C. (2013) DNA sequences that interfere with transcription: implications for genome function and stability. *Chem. Rev.*, **113**, 8620–8637.
16. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
17. Kouzine,F., Wojtowicz,D., Baranello,L., Yamane,A., Nelson,S., Resch,W., Kieffer-Kwon,K.-R., Benham,C.J., Casellas,R., Przytycka,T.M. *et al.* (2017) Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.*, **4**, 344–356.

18. Hurley,L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer*, **2**, 188–200.

19. Chaires,J.B. (2015) A small molecule—DNA binding landscape. *Biopolymers*, **103**, 473–479.

20. Chaires,J.B. and Mergny,J.-L. (2008) Targeting DNA. *Biochimie*, **90**, 973–975.

21. Ragazzon,P. and Chaires,J.B. (2007) Use of competition dialysis in the discovery of G-quadruplex selective ligands. *Methods*, **43**, 313–323.

22. Chaires,J.B. (2003) A Competition Dialysis Assay for the Study of Structure-Selective Ligand Binding to Nucleic Acids. *Curr. Protoc. Nucleic Acid Chem.*, **11**, doi:10.1002/0471142700.nc0803s11.

23. Ragazzon,P.A., Garbett,N.C. and Chaires,J.B. (2007) Competition dialysis: a method for the study of structural selective nucleic acid binding. *Methods*, **42**, 173–182.

24. Holt,P.A., Buscaglia,R., Trent,J.O. and Chaires,J.B. (2011) A discovery funnel for nucleic acid binding drug candidates. *Drug Dev. Res.*, **72**, 178–186.

25. Ren,J. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.

26. Boger,D.L., Fink,B.E., Brunette,S.R., Tse,W.C. and Hedrick,M.P. (2001) A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc.*, **123**, 5878–5891.

27. Largy,E., Hamon,F. and Teulade-Fichou,M.-P. (2011) Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.*, **400**, 3419–3427.

28. Largy,E., Saettel,N., Hamon,F., Dubruille,S. and Teulade-Fichou,M.P. (2012) Screening of a chemical library by HT-G4-FID for discovery of selective G-quadruplex binders. *Curr. Pharm. Des.*, **18**, 1992–2001.

29. Alcaro,S., Musetti,C., Distinto,S., Casatti,M., Zagotto,G., Artese,A., Parrotta,L., Moraca,F., Costa,G., Ortuso,F. *et al.* (2013) Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J. Med. Chem.*, **56**, 843–855.

30. Kumar,S., Xue,L. and Arya,D.P. (2011) Neomycin−neomycin dimer: an all-carbohydrate scaffold with high affinity for AT-rich DNA duplexes. *J. Am. Chem. Soc.*, **133**, 7361–7375.

31. Glass,L.S., Bapat,A., Kelley,M.R., Georgiadis,M.M. and Long,E.C. (2010) Semi-automated high-throughput fluorescent intercalator displacement-based discovery of cytotoxic DNA binding agents from a large compound library. *Bioorg. Med. Chem. Lett.*, **20**, 1685–1688.

32. Holt,P.A., Ragazzon,P., Strekowski,L., Chaires,J.B. and Trent,J.O. (2009) Discovery of novel triple helical DNA intercalators by an integrated virtual and actual screening platform. *Nucleic Acids Res.*, **37**, 1280–1287.

33. Nasiri,H.R., Bell,N.M., McLuckie,K.I.E., Husby,J., Abell,C., Neidle,S. and Balasubramanian,S. (2014) Targeting a c-MYC G-quadruplex DNA with a fragment library. *Chem. Commun.*, **50**, 1704–1707.

34. Watkins,D., Ranjan,N., Kumar,S., Gong,C. and Arya,D.P. (2013) An assay for human telomeric G-quadruplex DNA binding drugs. *Bioorg. Med. Chem. Lett.*, **23**, 6695–6699.

35. Ranjan,N. and Arya,D. (2013) Targeting C-myc G-quadruplex: dual recognition by aminosugar-bisbenzimidazoles with varying linker lengths. *Molecules*, **18**, 14228–14240.

36. Schoonover,M. and Kerwin,S.M. (2012) G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, **20**, 6904–6918.

37. Yang,D. and Okamoto,K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.*, **2**, 619–646.

38. Neidle,S. (2016) Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.*, **59**, 5987–6011.

39. Maizels,N. (2015) G4-associated human diseases. *EMBO Rep.*, **16**, 910–922.

40. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.

41. Neidle,S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118–1125.

42. Hansel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.

43. Islam,M.K., Jackson,P.J., Rahman,K.M. and Thurston,D.E (2016) Recent advances in targeting the telomeric G-quadruplex DNA sequence with small molecules as a strategy for anticancer therapies. *Future Med. Chem.*, **8**, 1259–1290.

44. Lytton-Jean,A.K.R., Han,M.S. and Mirkin,C.A. (2007) Microarray detection of duplex and triplex DNA binders with DNA-modified gold nanoparticles. *Anal. Chem.*, **79**, 6037–6041.

45. Xu,N., Yang,H., Cui,M., Wan,C. and Liu,S. (2012) High-performance liquid chromatography–electrospray ionization-mass spectrometry ligand fishing assay: a method for screening triplex DNA binders from natural plant extracts. *Anal. Chem.*, **84**, 2562–2568.

46. De Cian,A., Guittat,L., Shin-ya,K., Riou,J.-F. and Mergny,J.-L. (2005) Affinity and selectivity of G4 ligands measured by FRET. *Nucleic Acids Symp. Ser.*, **49**, 235–236.

47. Boger,D.L. and Tse,W.C. (2001) Thiazole orange as the fluorescent intercalator in a high resolution fid assay for determining DNA binding affinity and sequence selectivity of small molecules. *Bioorg. Med. Chem.*, **9**, 2511–2518.

48. Lewis,M.A. and Long,E.C. (2006) Fluorescent intercalator displacement analyses of DNA binding by the peptide-derived natural products netropsin, actinomycin, and bleomycin. *Bioorg. Med. Chem.*, **14**, 3481–3490.

49. Monchaud,D., Allain,C. and Teulade-Fichou,M.-P. (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorg. Med. Chem. Lett.*, **16**, 4842–4845.

50. Monchaud,D., Allain,C. and Teulade-Fichou,M.P. (2007) Thiazole orange: a useful probe for fluorescence sensing of G-quadruplex-ligand interactions. *Nucleosides Nucleotides Nucleic Acids*, **26**, 1585–1588.

51. Tse,W.C. and Boger,D.L. (2004) A fluorescent intercalator displacement assay for establishing DNA binding selectivity and affinity. *Acc. Chem. Res.*, **37**, 61–69.

52. Tse,W.C. and Boger,D.L. (2001) A fluorescent intercalator displacement assay for establishing DNA binding selectivity and affinity. *Curr. Protoc. Nucleic Acid Chem.*, **20**, 8.5.1–8.5.11.

53. Jamroskovic,J., Livendahl,M., Eriksson,J., Chorell,E. and Sabouri,N. (2016) Identification of compounds that selectively stabilize specific G-quadruplex structures by using a thioflavin T-displacement assay as a tool. *Chemistry*, **22**, 18932–18943.

54. Tran,P.L.T., Largy,E., Hamon,F., Teulade-Fichou,M.-P. and Mergny,J.-L. (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie*, **93**, 1288–1296.

55. Chan,D.S.-H., Yang,H., Kwan,M.H.-T., Cheng,Z., Lee,P., Bai,L.-P., Jiang,Z.-H., Wong,C.-Y., Fong,W.-F., Leung,C.-H. *et al.* (2011) Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer. *Biochimie*, **93**, 1055–1064.

56. Monchaud,D., Allain,C., Bertrand,H., Smargiasso,N., Rosu,F., Gabelica,V., De Cian,A., Mergny,J.L. and Teulade-Fichou,M.P. (2008) Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie*, **90**, 1207–1223.

57. Riechert-Krause,F., Autenrieth,K., Eick,A. and Weisz,K. (2012) Spectroscopic and calorimetric studies on the binding of an indoloquinoline drug to parallel and antiparallel DNA triplexes. *Biochemistry*, **52**, 41–52.

58. Le,S., Josse,J. and Husson,F. (2008) FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.*, **25**, 1–18.

59. Husson,F., Josse,J. and Pagès,J. (2010) Principal component methods—hierarchical clustering—partitional clustering: why would we need to choose for visualizing data? Technical Report, https://francoishusson.files.wordpress.com/2017/02/hcpc_husson_josse.pdf.

60. Praseuth,D., Guieysse,A.L. and Hélène,C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta (BBA) - Gene Struct. Expression*, **1489**, 181–206.

61. Phan,A.T. and Leroy,J.-L. (2000) Intramolecular i-motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.*, **17**, 245–251.

62. Phan,A.T., Guéron,M. and Leroy,J.-L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.

63. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.

64. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.

65. Dai,J., Carver,M. and Yang,D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.

66. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.

67. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.A. and Yang,D. (2006) 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.

68. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.

69. Vorlíčková,M., Kejnovská,I., Bednářová,K., Renčiuk,D. and Kypr,J. (2012) Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, **24**, 691–698.

70. Bishop,G.R. and Chaires,J.B. (2002) Characterization of DNA structures by circular dichroism. *Curr. Protoc. Nucleic Acid Chem.*, **11**, 7.11.1–7.11.8.

71. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

72. Jaumot,J., Eritja,R., Navea,S. and Gargallo,R. (2009) Classification of nucleic acids structures by means of the chemometric analysis of circular dichroism spectra. *Anal. Chim. Acta*, **642**, 117–126.

73. Kypr,J., Kejnovská,I., Renčiuk,D. and Vorlíčková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.

74. del Villar-Guerra,R., Gray,R.D. and Chaires,J.B. (2017) Characterization of quadruplex DNA structure by circular dichroism. *Curr. Protoc. Nucleic Acid Chem.*, **68**, 17.8.1–17.8.16.

75. Mergny,J.-L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.

76. Qu,X. and Chaires,J.B. (2000) Analysis of drug-DNA binding data. *Methods Enzymol.*, **321**, 353–369.

77. Krafčíková,P., Demkovičová,E. and Víglaský,V. (2017) Ebola virus derived G-quadruplexes: Thiazole orange interaction. *Biochim. Biophys. Acta (BBA) - Gen. Subj.*, **1861**, 1321–1328.

78. Jolliffe,I.T. (2002) *Principal Component Analysis*. Springer, NY.

79. Wold,S., Esbensen,K. and Geladi,P. (1987) Principal component analysis. *Chemom. Intell. Lab. Syst.*, **2**, 37–52.

80. Ringner,M. (2008) What is principal component analysis? *Nat Biotech*, **26**, 303–304.

81. Lever,J., Krzywinski,M. and Altman,N. (2017) Points of significance: principal component analysis. *Nat. Methods*,**14**, 641–642.

82. Ivosev,G., Burton,L. and Bonner,R. (2008) Dimensionality reduction and visualization in principal component analysis. *Anal. Chem.*, **80**, 4933–4944.

83. Gbriel,K.R. (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453–467.

84. Ding,C. and He,X.F. (2004) In: Dai,H, Srikant,R and Zhang,C. (eds). *Advances in Knowledge Discovery and Data Mining, Proceedings*. Springer-Verlag Berlin, Berlin, Vol. **3056**, pp. 414–418.

85. Barceló,F., Ortiz-Lombardía,M., Martorell,M., Oliver,M., Méndez,C., Salas,J.A. and Portugal,J. (2010) DNA binding characteristics of mithramycin and chromomycin analogues obtained by combinatorial biosynthesis. *Biochemistry*, **49**, 10543–10552.

86. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

87. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.