



Published in final edited form as:

J Am Chem Soc. 2018 November 07; 140(44): 14761–14769. doi:10.1021/jacs.8b08152.

Compound Shape Effects in Minor Groove Binding Affinity and Specificity for Mixed Sequence DNA

Pu Guo, Abdelbasset A. Farahat, Ananya Paul, Narinder K. Harika, David W. Boykin, and W. David Wilson

Department of Chemistry and Center for Diagnostics and Therapeutics Georgia State University, 50 Decatur St Se, Atlanta, GA 30303, USA.

Abstract

AT specific heterocyclic cations that bind in the DNA duplex minor groove have had major successes as cell and nuclear stains and as therapeutic agents which can effectively enter human cells. Expanding the DNA sequence recognition capability of the minor groove compounds could also expand their therapeutic targets and have an impact in many areas, such as modulation of transcription factor biological activity. Success in the design of mixed sequence binding compounds has been achieved with *N*-methylbenzimidazole (*N*-MeBI) thiophenes which are preorganized to fit the shape of the DNA minor groove and H-bond to the –NH of G·C base pairs that projects into the minor groove. Initial compounds bind strongly to a single G·C base pair in an AT context with a specificity ratio of 50 (K_D AT-GC/ K_D AT) or less and this is somewhat low for biological use. We felt that modifications of compound shape could be used to probe local DNA microstructure in target mixed base pair sequences of DNA and potentially improve the compound binding selectivity. Modifications were made by increasing the size of the benzimidazole *N*-substituent, for example, by using *N*-isobutyl instead of *N*-Me, and by changing the molecular twist by introducing substitutions at specific positions on the aromatic core of the compounds. In both cases, we have been able to achieve a dramatic increase in binding specificity, including no detectable binding to pure AT sequences, without a significant loss in affinity to mixed base pair target sequences.

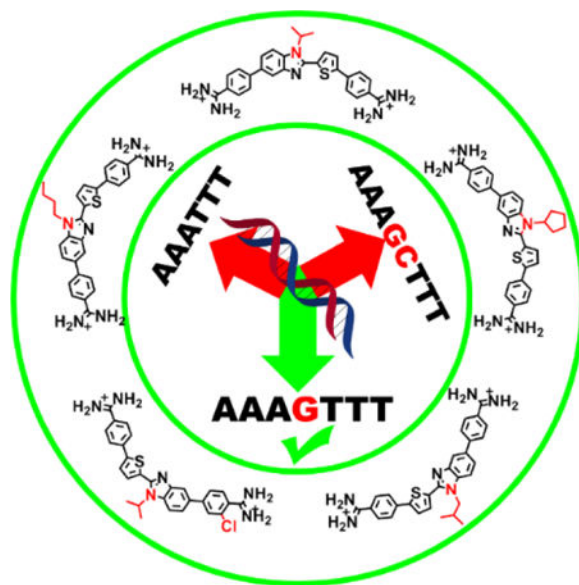
Graphical Abstract

Corresponding Author: W. David Wilson, wdw@gsu.edu.

Associated Content

Supporting Information. Experimental details, DNA melting experimental data table, CD spectrums, Molecular structure with specific atom types of DB2708, Frmod file of the DB2708, ¹H NMR spectra of key Intermediates and final products

Notes: The authors declare no competing financial interest.



Introduction

Design and preparation of agents that can recognize mixed base pair (bp) sequences of DNA containing combinations of A·T and G·C bps have been of long term interest. Starting in the mid-1980s, this took a step forward with work in the Dickerson and Lown laboratories to redesign the minor groove binding polyamides netropsin (Nt.) and distamycin (Dst.), which specifically recognizing A·T bp, into synthetic compounds that could also bind to a G·C bp.^{1,2} They proposed doing this by replacing one or more pyrrole groups in Nt and Dst with an imidazole group. The extra N in imidazole relative to pyrrole allows the imidazole group to form an H-bond with the G-NH₂ in the minor groove of a G·C bp.¹⁻³ The proposal attracted attention in several additional laboratories and success in GC recognition was achieved by connecting two recognition units to make a hairpin polyamide.⁴⁻⁶ Although several groups have worked with these synthetic polyamides, solution difficulties, aggregation, and poor cell uptake have limited their applications.⁷⁻⁹

In contrast to netropsin (Nt.) and distamycin (Dst.), other types of AT specific heterocyclic cations, such as, Hoechst dyes, furamidine, DAPI, pentamidine, bisamidinium, as well as numerous compounds that covalently attach to the minor groove, have had major successes as cell / nuclear stains and therapeutic agents.¹⁰⁻¹⁸ Interestingly, however, no systematic effort to redesign these agents into mixed bps recognition compounds, as with the polyamides, has been made. We felt that these promising and successful agents could potentially be redesigned to have significantly expanded use with broadened sequence recognition capability. We have now had considerable success in the design of agents that can recognize G·C bps as well as AT.¹⁹⁻²¹ Such new agents would have broad applications, for example, in targeting transcription factors to modulate gene expression.^{22,23}

Mutations, protein fusions or other changes that modify the activity of transcription actors (TF), which help control critical cell functions, can lead to a number of different kinds of

cancer.^{23–25} Given the variety of functions and diseases controlled by TFs, a strong interest in targeting and controlling this activity with small molecules has developed.²³ The problem is that TFs have evolved to bind high molecular weight nucleic acids, and not, in general, small molecules.^{26,27} This has made finding binding sites on TFs to interact strongly and specifically with small molecules, a difficult task. As a result, TFs are often defined as “undruggable”.²⁷ Our concept is to target TF-DNA complexes by binding of designed agents to specific DNA sequences. To do this, we must design new types of cell-permeable DNA binding agents with broader sequence recognition capabilities than the classical AT-specific agents.

Incorporating the *N*-MeBI thiophene module into the heterocyclic cations, as in Figure 1, initially gave DB2429 which specifically binds to a target G·C bp with flanking A·T bps.²¹ DB2429 was expanded to DB2457 to better cover the A·T bps that flank the central GC. These compounds incorporate the σ -hole concept and are a significant step forward in our molecular design and synthesis project for recognition of mixed bp DNA sequences.^{19,20,28} The σ -hole interaction preorganizes the thiophene-*N*-MeBI unit for GC interaction in the minor groove. Thiophene C-S single bonds present a relatively positive electrostatic potential that can form an interaction with electron donating atoms such as the unsubstituted *N*-MeBI N, a 1,4 N...S interaction. The interaction is based on the presence of low lying thiophene C-S σ^* orbitals on S that give rise to the positive electrostatic potential or a σ -hole.²⁹

Analysis of the importance of the *N*-MeBI-thiophene module to binding specificity and selectivity clearly shows that both the *N*-MeBI and thiophene are necessary for a strong, specific interaction.²¹ Replacing the thiophene with a furan or pyridine, for example, reduced both binding strength and specificity. The same was true for replacing the *N*-MeBI with a benzoxazole group which in principle has the same H-bonding capability as *N*-MeBI. Compounds with the *N*-Me substitution replaced by an NH group bind much more strongly to pure AT DNA sequences than to any sequence with a G·C bp.^{21,30} The remarkable reversal of binding specificity with a relatively small change in the overall compound structure is an important observation in new ideas for the design of agents to recognize specific DNA sequences. The –NH of benzimidazole is a strong A·T bp recognizing element while the unsubstituted N of *N*-MeBI is a strong G·C bp interacting element.

The specificity ratio for binding to the single G·C bp sequence versus the pure AT sequence is only about a factor of 10 for DB2429 while it increases to about 50 for the larger DB2457 which binds more strongly to all sequences.²¹ These selectivity ratios are too low to push these two initial compounds forward into biological testing and additional development of the compounds is required. As described above, the strong decrease of binding affinity and selectivity for most structural changes with DB2429 and DB2457 raised the question of whether the selectivity ratio of these structures could be enhanced by different types of substitutions. Because of the critical importance of high selectivity in targeting DNA, we have investigated new modifications of these initial compounds that have the potential to increase the selectivity. Initial synthetic modifications and DNA binding studies have been conducted with derivatives of the larger, stronger binding compound, DB2457.

From the first crystal structures of a minor groove binder to extensive structures of heterocyclic diamidines by the Neidle group,³¹ and recent ideas about minor groove from the microstructural analysis,^{32–36} the wider DNA minor groove in GC versus AT sequences has been recognized. As a reference for the three binding sites of interest in this report (Figure 1), we have used the DNA Shape algorithm of Rohs and coworkers to estimate the minor groove widths.³² The program takes DNA sequences and predicts a number of double helical properties, including minor groove width. The predictions are based on an extensive compilation of DNA duplex properties from experimental structure (PDB) as well as computational analysis by the Rohs and other groups.^{32–36} For AAATTT the predicted minor groove width in the center is 2.9 Å and widens to almost 5 Å at the flanking G-C bps. At the AAAGTTT binding site, the predicted minor groove width is between 3.3–3.4 Å, considerably wider on a molecular scale than the AAATTT pure AT sequence. For AAAGCTTT the predicted minor groove width is over 4 Å. All sequences widen to near 5 Å at the flanking, terminal GC sequences. These results agree with the general concept of widening of AT sequence minor grooves with the incorporation of G-C bps.^{31–33}

Based on these observations, the concept for the compound modifications described here is to target DNA microstructural variations, such as the wide minor groove in G-C bp containing regions. We reasoned that compounds with increased bulk might favor the wider, mixed bp sequences over pure AT DNA. To test this idea the following general changes in the compound structure and functional groups were made with the goal of increasing specificity: (i) by increasing compound bulk primarily by replacing the *N*-Me group of *N*-MeBI with groups of increased size and complexity, and (ii) by modifying the overall twist in the compound linked aromatic core structure with appropriate substituents (Figure 1).

As we move to design new compounds to inhibit specific transcription factors, two things are very important: the selectivity ratio and the ability of the compounds to enter different cell types. At least some substituent changes at the *N*-BI position should have positive effects on both of these critical factors. We were able to successfully synthesize compounds with a wide variety of substituents at the *N*-BI position and three compounds with substituents on the aromatic groups of DB2457 that modified the effective steric bulk of the compounds. With these modifications, we were able to make significant improvements in compound specificity without a decrease in binding affinity.

Compound Design

The first set of compounds that were prepared had modifications at the *N*-BI position of DB2457 (Table 1). We wanted to increase the length and complexity of the alkyl group, for example, methyl – ethyl – isopropyl – isobutyl – neopentyl substituents. There are also linear groups with more polar substituents as in DB2728. To significantly change the shape of the alkyl group, compounds with cyclobutyl (DB2726), cyclopentyl (DB2714) and cyclohexyl (DB2727) substituents were prepared. Finally, the chemistry of the substituent was changed dramatically with different phenyl (Ph) substituents on the BI group, *N*-Ph (DB2740) and *N*-ortho-Me phenyl (DB2747). A control compound without an *N*-BI substituent, a benzimidazole, was also prepared. Two compounds, with an *N*-MeBI and an *N*-isopropyl-BI were prepared with the thiophene S replaced by Se to evaluate the effect of

size and changes in polarizability at that position.^{37,38} Four additional compounds were prepared with substitutions designed to increase the twist of the aromatic rings of the core structure. DB2754 has a Me on the thiophene, DB2753 has a Me on the Ph attached to the BI while DB2759 has a chloro and DB2762 has a trifluoromethyl group adjacent to the amidine on the same phenyl ring. We were unable to synthesize a compound with a Me adjacent to the amidine. Brief synthetic details are given here for the synthesis of the new compounds along with Scheme 1. Full synthetic details for all compounds are given in the Supporting Information.

Results and Discussion

Chemistry

The general synthetic approach to the target diamidine compounds **7** is outlined in Scheme 1. Nucleophilic aromatic substitution reactions between 4-bromo-1-fluoro-2-nitrobenzene **1** and various aliphatic and aromatic amines yields the 4-bromo-*N*-alkyl and *N*-aryl-2-nitrobenzenes **2** in reasonable yields (41–79%), which ultimately leads to the *N*-alkyl and *N*-aryl benzimidazoles. The nucleophilic aromatic substitution reaction with aliphatic amines was achieved at room temperature in ethanol whereas the less basic aryl amines required heating in dimethylacetamide in the presence of Cs₂CO₃ at 160 °C. Suzuki-Miyaura coupling of 4-cyanophenylboronic acids with the various 4-bromo-*N*-alkyl and *N*-aryl-2-nitrobenzenes under standard conditions provides the nitro biphenyl analogues **3** in good yields (51–79%). The arylphenylene diamines **4** required for formation of the *N*-substituted benzimidazoles were obtained by stannous chloride reduction of the nitro groups of **3** and were subsequently allowed to react directly without characterization. Sodium metabisulfite mediated the oxidative condensation and cyclization of the phenylene diamines **4** with the previously reported²¹ 5-(cyanophenyl)-2-formylthiophenes **5** to yield the bis-nitriles **6** in acceptable yields (39–76%). Lithium bis(trimethylsilyl)amide in THF reacts with the bis-nitriles **6** at room temperature to yield, after acidic work-up, the diamidines **7** generally in acceptable yields (30–79%).

DNA thermal melting: Screening for mixed DNA sequence binding

Changes in DNA thermal melting temperature (T_m) provide a rapid way for initial ranking of compounds for binding affinity with different DNA sequences and are shown in Table S1. We have used the three selected sequences shown in Figure 1 for comparative experiments. Most heterocyclic cations prepared to this time have been pure AT sequence binders and a DNA with the binding site -AAATTT- was used to test for pure AT sequence binding affinity. The target DNA of primary interest has a single G·C bp with the binding sequence -AAAGTTT-. The third sequence with two G·C bps was also used to test for binding affinity and selectivity versus AAAGTTT.

The reference compound, DB2457, with an *N*-Me substituent was previously reported²¹ and has preferential binding to the single G·C bp sequence and weaker binding to the pure AT and two G·C bps sequences (Table S1). The compounds with the *N*-ethyl and *N*-isopropyl substituents had similar binding to AAAGTTT and weaker binding to the pure AT sequence for improved selectivity as desired. At the level of evaluation of T_m , most of the other

derivatives at the *N*-X position had very weak binding to the pure AT sequence but variable binding to the single and double G-C bps sequences. A -Cl substituent ortho to an amidine group, DB2759, had relatively strong binding to AAAGTTT and very low binding to the pure AT and two G-C bps sequences. It provides a potential new lead for biological studies. Two compounds were prepared with S in thiophene replaced with Se, but they showed no improvement in T_m for AAAGTTT or in selectivity. The compounds will be of a long term interest, however, in cell uptake studies.

Biosensor-SPR: Methods for quantitative binding

A biacore sensorchip (CM5) was functionalized with streptavidin and used to immobilize 5'-biotin labeled DNA sequences (Figure 1) in flow cells 2–4 and flow cell 1 was left as a blank, for background subtraction. With different compounds in the flow solution, we were able to determine comparative binding constants for all of the derivatives (Table 1). Sensorgrams were obtained for all compounds and are shown for representative compounds with the different DNAs in Figure 2. With the AAATTT and AAAGCTTT sequences, all compounds showed significantly weaker binding to AAATTT than with AAAGTTT with generally fast on and off rates. No kinetics and limited K_D values could be accurately determined for those DNA sequences (Table 1). With the AAAGTTT sequence, however, excellent sensorgrams were obtained at low concentrations for most derivatives (Figure 2). Kinetics fits with a one site model were excellent and allowed a determination of K_D values (Table 1). Kinetics fits were generally required to determine K_D values since many sensorgrams did not reach a steady-state level, especially at the lower concentrations.

DB2429, the original compound (Figure 1), has a K_D for AAAGTTT of 50 nM with 10 fold higher K_D value for AAATTT and 20 fold higher K_D value for AAAGCTTT.²¹ The parent *N*-Me compound in this study, DB2457, has a 10 fold lower K_D value for AAAGTTT, 4 nM, with significantly better, 50 fold higher, K_D values for AAATTT and AAAGCTTT relative to DB2429. The *N*-ethyl and *N*-isopropyl compound have K_D values similar to the *N*-Me for single G-C bp sequence but higher values for AAATTT and impressive improvement in selectivity ratio of K_D s of near 250. Unfortunately, they have similar selectivity for G-C bp to the *N*-Me compound and it was necessary to synthesize additional compounds for a full increase in specificity. The *N*-isobutyl (DB2711) has slightly weaker binding to the single G-C bp sequence ($K_D = 13$ nM) but the desired high selectivity with no measurable binding to both the AAATTT and AAAGCTTT sequences. This compound, with its broad, high selectivity in binding validates the idea of *N*-substitution to enhance selectivity, a key feature for designing any new potential drug molecules in this series. Most of the remaining substituents at the *N*-position in Table 1 are longer than isopropyl (DB2708) and these compounds have no detectible binding to the narrowest full AT sequence, AAATTT, binding site. They generally have satisfactory binding to the AAAGTTT single G sequence with only one above 21 nM K_D . Most of these compounds, however, do not have satisfactory selectivity ratios with the two GC sequence, only two with selectivity ratio >100. It seems likely that these compounds are able to induce a narrowing of the AAAGCTTT minor groove to provide a minor groove that is a reasonable, but not as good as AAAGTTT, binding site.

The isobutyl compound, as with all of the larger *N*-R substituted derivatives does not fit well into the narrowest AAATTT groove and give no detectible binding to this sequence under our conditions. The *N*-isobutyl compound, DB2711, is not able to reshape the AAAGCTTT minor groove into a conformation that supports significant minor groove binding. Relative to the *N*-Me, -ethyl, and -isopropyl derivatives DB2711 has only about a 3 fold increase in K_D and still binds very strongly to AAAGTTT.

The second type of major substitution changes were made on the core aromatic system. Methyl substitution on a phenyl (DB2753) and the thiophene (DB2754) apparently causes too much twist in the compound structure with the *N*-isopropyl group and the K_D values for AAAGTTT increase are 20 fold relative to the unsubstituted *N*-isopropyl compound (DB2708). We also tried to increase the twist at a phenyl amidine position (DB2759 –Cl and DB2762 –CF₃) with mixed results. The –CF₃ derivative apparently has too much twist and binds relatively weakly to all DNAs. The derivative with the smaller –Cl substitution, however, apparently has near perfect twist on the amidine to fit the DNA microstructure at the end of the AAAGTTT sequence where there are G-C bps and some expansion of the groove. This compound has negligible binding to AAATTT and AAAGCTTT but only a 3 fold increase in K_D relative to unsubstituted *N*-isopropyl derivative (DB2708). These excellent binding results with AAAGTTT coupled with effectively infinite selectivity over the classical recognition sequence AAATTT as well as the two GC sequence make the –Cl (DB2759) and *N*-isobutyl derivatives our current lead compounds for biological development. This entire compound set and especially the two lead compounds provide excellent ideas for future design efforts of minor groove binding compounds.

The two derivatives with Se in place of S do not have results that are more attractive than the equivalent thiophenes. The *N*-isopropyl derivative has about a 2 fold increase in K_D for the single GC sequence. It has weaker binding to both the AT and two GC sequence. In other minor groove binders series, we have found that Se has some attractive cell and nuclear uptake properties and the *N*-isopropyl Se compound will be an important one for biological testing.²²

Each of the compounds in Table 1 can also induce changes in the minor groove structure of DNA and at this time an experimental analysis is the best way to find the optimum compound for affinity and selectivity to a specific sequence. The most important finding, however, is that modification of DB2457 can dramatically increase binding selectivity.

Biosensor-SPR Kinetics

A selection of compounds with a range of K_D values and a variety of chemical substitutions and structures was used to evaluate the kinetics of association, k_a , and dissociation, k_d , for the AAAGTTT sequence. As can be seen in Figure 2, the binding kinetics were generally too rapid to fit accurately with the AAATTT and AAAGCTTT sequences. To help compare the kinetics results for AAAGTTT, $\log k_a$ is plotted versus $\log k_d$ in Figure 3 and results are summarized in Table 2. The green dashed lines in Figure 3 plot the binding affinities (K_D) in logarithmic jumps. The parent compound DB2457 and the strongest binding compounds, DB2740 and DB2708, are located on a K_D line with a slope of 3~4 nM. Below these, there are six compounds with different, but similar sized *N*-alkyl substitutions that are closely

distributed around the 10 nM K_D green dashed line. The on-rates and off-rates for this group change in a correlated manner to give similar K_D values.

For the *N*-substituted compounds with an unsubstituted aromatic system, the K_D values with AAAGTTT in Table 2 cover a narrow range, from around 4 to 13 nM with only two compounds above 10 nM. The kinetics values for these compounds also cover a relatively small range for k_a , from around 2 to $5 \times 10^6 \text{ m}^{-1}\text{s}^{-1}$. The k_d values cover a somewhat larger range, from 0.008 to 0.05 s^{-1} but as noted above, a general correlation of k_a and k_d values accounts for the small variation in K_D . The *N*-Me (DB2457) and *N*-Ph (DB2740) compounds have the lowest K_D values in this set and this arises from their similar low k_d values. DB2715 with *n*-butyl and DB2711 with isobutyl have the fastest disassociation rates with the weakest binding (12.6 and 9.6 nM) to AAAGTTT among the compounds with an unsubstituted aromatic system. These results suggest that the very dynamic structure of the butyl substituents interfere with binding in the minor groove giving them fast association and rapid dissociation with the highest K_D values in this set. These observations indicate that the fast association results from a lack of optimum penetration into the minor groove and, as a result, the compounds also do not make strong contact with the groove and dissociate rapidly. DB2764 with 2-methoxyethyl substituent, which is the approximately same length and flexibility as *n*-butyl (DB2715), also has a relatively high k_d and K_D . The Se *N*-isopropyl derivative DB2712 has a k_a value similar to the equivalent thiophene but a higher k_d and as a result, a K_D that is around 2 fold weaker. Clearly, the larger, more polarizable Se is a slight disadvantage in DNA complexes in this set. A methyl-substituted phenyl, *N*-isopropyl derivative, DB2753, has much weaker binding, $K_D > 100 \text{ nM}$, that is due to a very low k_a of $0.073 \times 10^6 \text{ m}^{-1}\text{s}^{-1}$. The k_d for this compound, however, is in the range of other compounds in Table 2. The core aromatic system of this compound is significantly twisted due to repulsion between the -Me substituent and the adjacent phenyl ring (Figure 4). To bind into the minor groove, the twist must be reduced at an energy cost that slows association with lowering of k_a and an increase in K_D . Once bound it appears that the compound can form interactions similar to the compounds with an unsubstituted aromatic system to give a k_d in the same range.

CD: Binding mode

CD titration experiments are an effective and convenient method of evaluating the binding mode and the saturation limit for compounds binding with DNA sequences. CD spectra monitor the asymmetric environment of the compounds binding to DNA and therefore can be used to obtain information on the binding mode.³⁹ There are no CD signals for the free compounds but on the addition of the compounds into DNA, substantial positive induced CD signals (ICD) arose in the absorption region between 300 and 450 nm. These positive ICD signals indicate a minor groove binding mode by these ligands, as expected from their structures. As can be seen from Figure S1, all tested compounds form complexes in the minor groove of the AAGTTT sequences with a 1:1 stoichiometry, in agreement with SPR results. In summary, the CD titration results confirm a minor groove binding mode for the compounds in Figure 1.

Molecular Structure: Calculations

To help to understand the large differences in binding affinities between the *N*-isopropyl compound DB2708, with no substitutions on the core aromatic system, and the aromatic substituted *N*-isopropyl compounds DB2753, DB2754, DB 2759 and DB2762, torsional angle maps were constructed in the *SPARTAN 16* software package⁴⁰ (Figure 4). The relative E in these plots is the energy required to convert the compound torsional angle from the lowest energy angle to a planar conformation at 0°. This gives an approximate idea of the relative energy input into the compound for a structural change to bind to the minor groove. This analysis assumes that the torsional angle must be between about 0–10° for optimum binding to the minor groove. The substituted compounds all have larger K_D values than DB2708 and both their lowest relative E torsional angles and the energies required to reach 0° increase in a correlated manner with the K_D values (Figure 4H). These results clearly show why the –Cl compound (DB2759) has a much lower K_D than the –CF₃ compound (Figure 4). In the same manner, the methyl substituted compounds have large K_D values and correlated large low energy torsional angles. It seems highly likely that these conformational energy costs are responsible for at least part of large binding constant differences for DB2708, DB2759, DB2762, DB2753, and DB2754. It should be noted that choosing a specific minimum torsional angle, such as 5°, leads to the same conclusions.

Molecular Structure: Molecular Dynamics

To help better understand the structural basis of molecular recognition of DNA sequences with a single G·C bp in an AT context, molecular dynamics (MD) simulations for the representative *N*-isopropyl compound, DB2708, were conducted⁴¹. Force constants for DB2708 were determined as described previously and in the methods section, and added to the force field for the simulations. After 500 ns of MD simulation, 25 PDB structures for the complex of DB2708 in the AAAGTTT binding site (Figure 5) were collected. Structures were evaluated in detail every 20 ns (25 structures for 500 ns) to determine what major features of the DB2708-DNA complex are primarily responsible for the excellent stability (K_D) and specificity ratios (K_D of AAAGTTT/ K_D of AAATTT). There are three optimum H-bonds in the complex. Both amidine groups form –N-H to T=O H-bonds (Figure 5B) that are an average of 2.8 Å in length (SD = 0.1 Å). The amidines also form numerous highly dynamic H-bonds to water molecules that move in and out of the minor groove.⁴² These water molecules frequently also form H-bonds with A·T bps at the floor of the minor groove and help link the compound to the specific binding site in the groove and stabilize the complex. The third strong H-bond is from the central G-NH that projects into the minor groove to the unsubstituted imidazole N in the BI group of DB2708, Figure 5C, to account for much of the binding selectivity of DB2708. Additional selectivity in binding is provided by the –CH group of the six-member ring of BI that points into the minor groove. This -CH forms a dynamic close interaction with the –C=O of the dC base of the central G·C bp as well as to a T=O on an adjacent A·T bp (Figure 5C).

Additional direct stabilizing interactions are formed by –CH groups that point to the floor of the minor groove from the two phenyls of DB2708. These –CH groups form significant dynamic interactions with A-N3 groups on the bases at the floor of the groove (see Figure 5B for an example). DB2708 tracks optimally along the minor groove with appropriate twist

to match the minor groove curvature. Extensive interactions are formed by the conjugated aromatic-system of DB2708 with the sugar-phosphate walls of the minor groove. In addition to H-bonds with the A·T bps, the amidines of DB2708 also form electrostatic interactions with the backbone phosphates that result in Na⁺ release and an entropy increase that is another stabilizing component, along with water release from the minor groove, on complex formation. The sum of these numerous, relatively weak, stabilizing interactions results in strong binding of DB2708, a K_D of 4 nM (Table 1) with a ΔG of binding of 11.5 kcal/mol. The isopropyl group of DB2708 forms a tight interaction with the DNA backbone in the G·C bp region. As noted above the AAATTT sequence has a narrower minor groove in this central region of the binding site than AAAGTTT and, the isopropyl group would have increased steric hindrance to increase the K_D for AAATTT binding. The K_D for binding to the AAATTT sequence, for example, increases from around 200 nM with DB2457, the *N*-Me derivative, to around 1000 nM with the isopropyl compound DB2708 (Table 1) to provide a significant increase in selectivity. Clearly, many of the *N*-X derivatives in Table 1 can match the minor groove microstructure in AAAGTTT better than with AAATTT. Simulation of additional *N*-X derivatives and more detailed analysis will be presented in a future paper.

Conclusion

Here we have synthesized a series of DNA sequence-specific recognition ligands based on the structure of DB2457 with *N*-MeBI-thiophenes which are preorganized by a σ -hole interaction to fit the shape of the DNA minor groove and H-bond to the –NH of G·C bps that project into the minor groove. Over a long period of time it has been observed that adding a G·C bp into an entire AT DNA sequence widens the minor groove at the G·C bp. We show some examples of the widened groove width from AAATTT to AAAGTTT and AAAGCTTT in this report. Based on these observations, we prepared a series of compounds with different sized substituents to determine if we could take advantage of the difference in groove width to increase the sequence specificity of minor groove binder. The compounds were localized to the G·C bp in the AT sequence by an H-bond acceptor group on an *N*-X substituted BI-thiophene module in all of the compounds. Varying the X substituent on the *N*-X BI produced a remarkable enhancement of selectivity for the single G·C bp sequence (Table 1). Molecular dynamics simulation showed that the strong binding of the compound in the minor groove was the result of numerous relatively weak interaction that can sum to give an excellent ΔG of binding. These interactions include H-bonds, including a critical H-bond from the G-NH₂ to the unsubstituted N of the *N*-X BI group that accounted for a major part of the sequence selectivity of the compounds in this series. Other important interactions are listed in the MD section.

A second major change in the structure of some pouds in Table 1 was caused by adding substituents to the core aromatic system of the compound. The goal of this modification was to increase the aromatic system twist to match the sequence-dependent microstructural change in the minor groove. Most of the substituents apparently caused too much twist and caused major decrease in the binding affinity. Computational studies indicated that the minor groove of the AAAGTTT sequence increased in width at the ends of the sequence (Figure 1). This suggested that an increase in twist at a terminal amidine might again select for

minor groove width difference for the different sequences. A –Cl and a –CF₃ substituent were placed next to one phenyl-amidine to investigate this possibility. The –CF₃ caused too much twist and decreased binding for all sequences. The –Cl substituent, however, slightly decrease binding to the single GC sequence but essentially abolished binding to the full AT and two G-C bps sequences under our conditions (Table 1) and is now one of our lead compounds for biological studies.

In summary, the compound design concept described and illustrated in this report provide now the idea for the field. The compounds in Table 1 show that high specific binding to DNA can be obtained for minor groove binders. It seems likely that more complex sequence recognition can be obtained by combining recognition modules together such that many different target sequence can be selectively targeted. Similar idea for targeting RNA could greatly enhance therapeutic targeting of RNA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

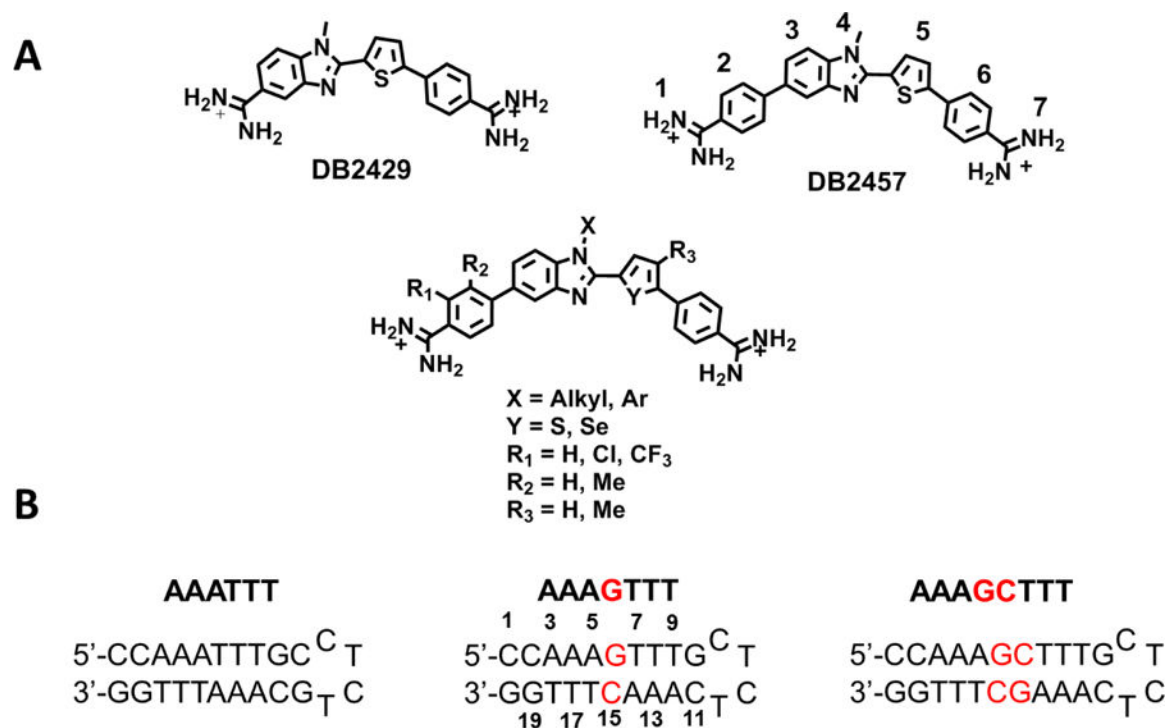
Acknowledgment

We thank the National Institutes of Health Grant GM111749 (W.D.W. and D.W.B.) for financial support. We also thank the Leukemia & Lymphoma Society (LLS) award: Leukemia & Lymphoma Society 6504-17 for support.

References

- (1). Kopka ML; Yoon C; Goodsell D; Pjura P; Dickerson RE J. Mol. Biol 1985, 183 (4), 553. [PubMed: 2991536]
- (2). Lown JW; Krowicki K; Bhat UG; Skorobogaty A; Ward B; Dabrowiak JC Biochemistry 1986, 25 (23), 7408. [PubMed: 3026455]
- (3). Kielkopf CL; Baird EE; Dervan PB; Rees DC Nat. Struct. Mol. Biol 1998, 5 (2), 104.
- (4). Dervan PB; Bürlri RW Curr. Opin. Chem. Biol 1999, 3 (6), 688. [PubMed: 10600731]
- (5). Bando T; Sugiyama H Acc. Chem. Res 2006, 39 (12), 935. [PubMed: 17176032]
- (6). Wang S; Nanjunda R; Aston K; Bashkin JK; Wilson WD Biochemistry 2012, 51 (49), 9796. [PubMed: 23167504]
- (7). Nozeret K; Loll F; Cardoso GM; Escudé C; Boutorine AS Biochimie 2018, 149, 122. [PubMed: 29626498]
- (8). Hargrove AE; Raskatov JA; Meier JL; Montgomery DC; Dervan PB J. Med. Chem 2012, 55 (11), 5425. [PubMed: 22607187]
- (9). Nishijima S; Shinohara K-I; Bando T; Minoshima M; Kashiwazaki G; Sugiyama H Bioorg. Med. Chem 2010, 18 (2), 978. [PubMed: 20022508]
- (10). Chazotte B Cold Spring Harb. Protoc 2011, 2011 (1), pdb.prot5557.
- (11). Giordani F; Munde M; Wilson WD; Ismail MA; Kumar A; Boykin DW; Barrett MP Antimicrob. Agents Chemother 2014, 58 (3), 1793. [PubMed: 24366732]
- (12). Wilson WD; Nguyen B; Tanious FA; Mathis A; Hall JE; Stephens CE; Boykin DW Curr. Med. Chem. Anticancer Agents 2005, 5 (4), 389. [PubMed: 16101490]
- (13). Paine MF; Wang MZ; Generaux CN; Boykin DW; Wilson WD; De Koning HP; Olson CA; Pohlig G; Burri C; Brun R; Murilla GA; Thuita JK; Barrett MP; Tidwell RR Curr. Opin. Investig. Drugs 2010, 11 (8), 876.
- (14). Soeiro MNC; Werbovetz K; Boykin DW; Wilson WD; Wang MZ; Hemphill A Parasitology 2013, 140 (8), 929. [PubMed: 23561006]

- (15). Ming X; Ju W; Wu H; Tidwell RR; Hall JE; Thakker DR *Drug Metab. Dispos* 2009, 37 (2), 424. [PubMed: 18971316]
- (16). Wong C-H; Nguyen L; Peh J; Luu LM; Sanchez JS; Richardson SL; Tuccinardi T; Tsoi H; Chan WY; Chan HYE; Baranger AM; Hergenrother PJ; Zimmerman SC *J. Am. Chem. Soc* 2014, 136 (17), 6355. [PubMed: 24702247]
- (17). Nguyen L; Luu LM; Peng S; Serrano JF; Chan HYE; Zimmerman SC *J. Am. Chem. Soc* 2015, 137 (44), 14180. [PubMed: 26473464]
- (18). Siboni RB; Bodner MJ; Khalifa MM; Docter AG; Choi JY; Nakamori M; Haley MM; Berglund JA *J. Med. Chem*, 2015, 58 (15), 5770. [PubMed: 26103061]
- (19). Chai Y; Paul A; Rettig M; Wilson WD; Boykin DW *J. Org. Chem* 2014, 79 (3), 852. [PubMed: 24422528]
- (20). Paul A; Nanjunda R; Kumar A; Laughlin S; Nhili R; Depauw S; Deuser SS; Chai Y; Chaudhary AS; David-Cordonnier M-H; Boykin DW; Wilson WD *Bioorg. Med. Chem. Lett* 2015, 25 (21), 4927. [PubMed: 26051649]
- (21). Guo P; Paul A; Kumar A; Farahat AA; Kumar D; Wang S; Boykin DW; Wilson WD *Chem. Eur. J* 2016, 22 (43), 15404. [PubMed: 27624927]
- (22). Munde M; Wang S; Kumar A; Stephens CE; Farahat AA; Boykin DW; Wilson WD; Poon GM K. *Nucleic Acids Res* 2014, 42 (2), 1379.
- (23). Antony-Debré I; Paul A; Leite J; Mitchell K; Kim HM; Carvajal LA; Todorova TI; Huang K; Kumar A; Farahat AA; Bartholdy B; Narayanagari S-R; Chen J; Ambesi-Impiombato A; Ferrando AA; Mantzaris I; Gavathiotis E; Verma A; Will B; Boykin DW; Wilson WD; Poon GMK; Steidl UJ *Clin. Invest* 2017, 127 (12), 4297.
- (24). Jiang X; Yang Z *OncoTargets Ther* 2018, 11, 3533.
- (25). Lambert M; Jambon S; Depauw S; David-Cordonnier M-H *Molecules* 2018, 23 (6), 1479.
- (26). Darnell JE *Nat. Rev. Cancer* 2002, 2 (10), 740. [PubMed: 12360277]
- (27). Koehler AN *Curr. Opin. Chem. Biol* 2010, 14 (3), 331. [PubMed: 20395165]
- (28). Harika NK; Germann MW; Wilson WD *Chem. Eur. J* 2017, 23 (69), 17612. [PubMed: 29044822]
- (29). Beno BR; Yeung K-S; Bartberger MD; Pennington LD; Meanwell NA *J. Med. Chem* 2015, 58(11), 4383 [PubMed: 25734370]
- (30). Mallena S; Lee MPH; Bailly C; Neidle S; Kumar A; Boykin DW; Wilson WD *J. Am. Chem. Soc* 2004, 126 (42), 13659. [PubMed: 15493923]
- (31). Neidle S *Nat. Prod. Rep* 2001, 18 (3), 291. [PubMed: 11476483]
- (32). Zhou T; Yang L; Lu Y; Dror I; Dantas Machado AC; Ghane T; Di Felice R; Rohs R *Nucleic Acids Res* 2013, 41(W1), W56. [PubMed: 23703209]
- (33). Azad RN; Zafiroopoulos D; Ober D; Jiang Y; Chiu TP; Sagendorf JM; Rohs R; Tullius TD *Nucleic Acids Res* 2018, 46(5), 2636. [PubMed: 29390080]
- (34). Rohs R; West SM; Sosinsky A; Liu P; Mann RS; Honig B *Nature* 2009, 461 (7268), 1248. [PubMed: 19865164]
- (35). Laughlin-Toth S; Carter EK; Ivanov I; Wilson WD *Nucleic Acids Res* 2017, 45 (3), 1297. [PubMed: 28180310]
- (36). Bishop EP; Rohs R; Parker SCJ; West SM; Liu P; Mann RS; Honig B; Tullius TD *ACS Chem. Biol* 2011, 6 (12), 1314. [PubMed: 21967305]
- (37). Steinmann D; Nauser T; Koppenol WH *J. Org. Chem* 2010, 75 (19), 6696. [PubMed: 20806911]
- (38). Reich HJ; Hondal RJ *ACS Chem. Biol* 2016, 11 (4), 821. [PubMed: 26949981]
- (39). Fornander LH; Wu L; Billeter M; Lincoln P; Nordén BJ *Phys. Chem. B* 2013, 117 (19), 5820.
- (40). Hehre W; Ohlinger S *Spartan™16 Tutorial and User'S Guide* Wavefunction, Inc., 2016
- (41). Case DA; Babin V; Berryman J; Betz RM; Cai Q *Amber 14*, NYU Press, 2014
- (42). Harika NK; Wilson WD *Biochemistry* 2018, 57 (33), 5050. [PubMed: 30048590]

**Figure 1.**

A) Chemical structures of reported single G-C bp binders²¹ and the modified compounds from DB2457 used in this study. B) The DNA sequences used in this study; DNA sequences used for SPR studies were labeled with 5'-biotin.

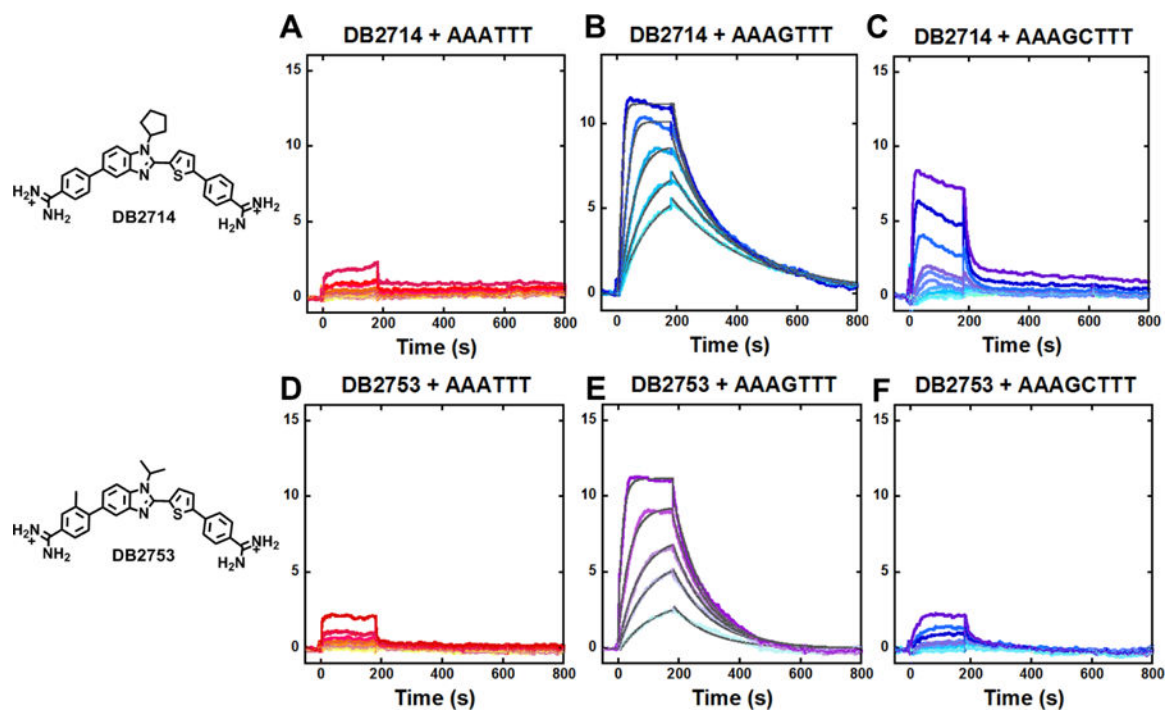


Figure 2.

Representative SPR sensorgrams for A-C, DB2714, D-F, DB2753 in the presence of AAATTT, AAAGTTT and AAAGCTTT hairpin DNAs. In A, C, D, F, the concentrations of sensorgrams is 2–500 nM of each compound from bottom to top. In B, the concentrations of DB2714 from bottom to top are 15, 20, 30, 50, and 100 nM; In E, the concentrations of DB2753 from bottom to top are 30, 70, 100, 200, and 500 nM. In B and E, the solid black lines are best-fit values for the global kinetic fitting of the results with a single site function.

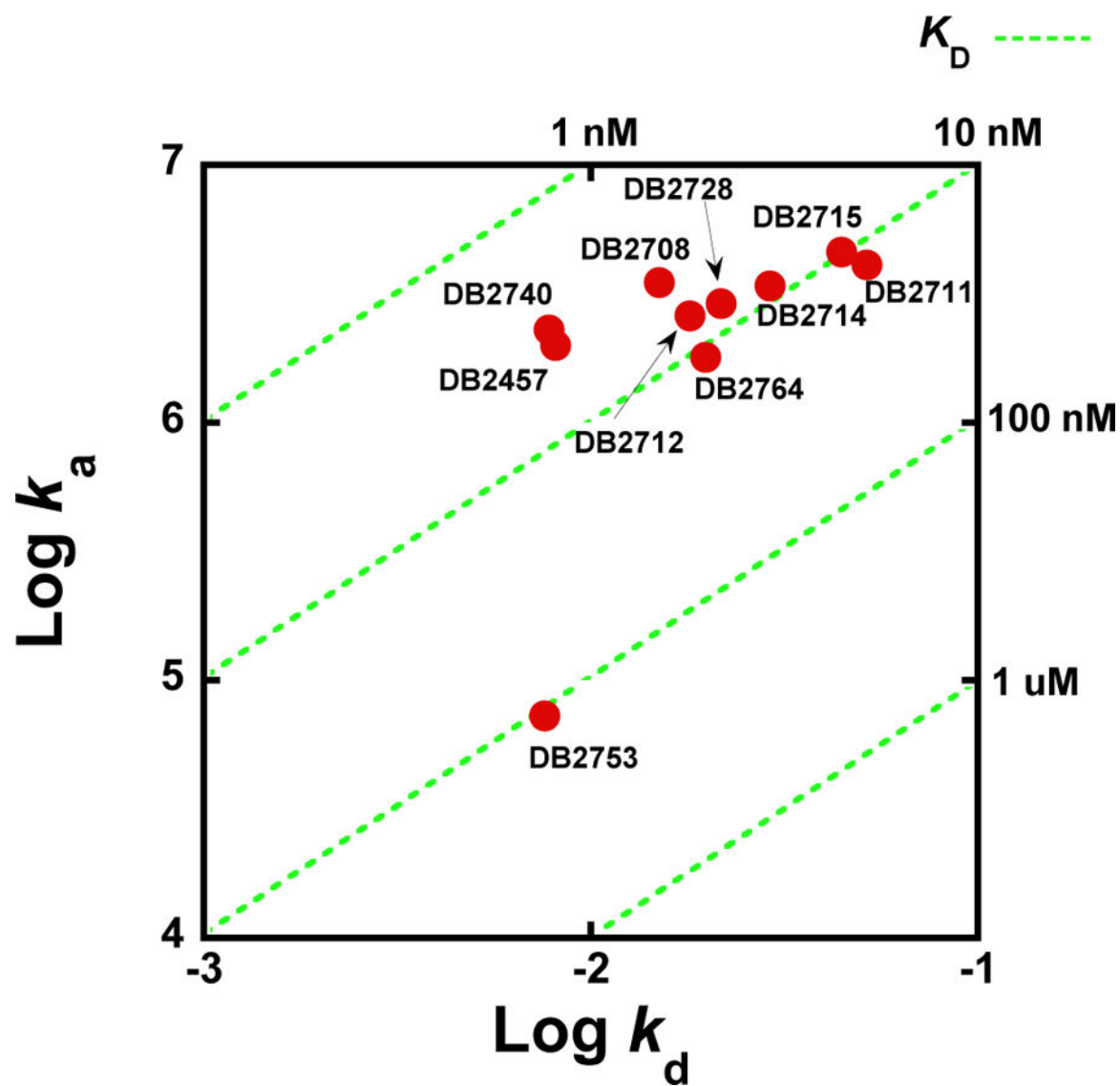


Figure 3. Kinetic plot showing the relationship between on-rates (k_a), off-rates (k_d) and binding affinities (K_D) (Green dash line) represented on the diagonal axis for compounds in Table 2, and determined by SPR. On-rates and off-rates are plotted as logarithmic values.

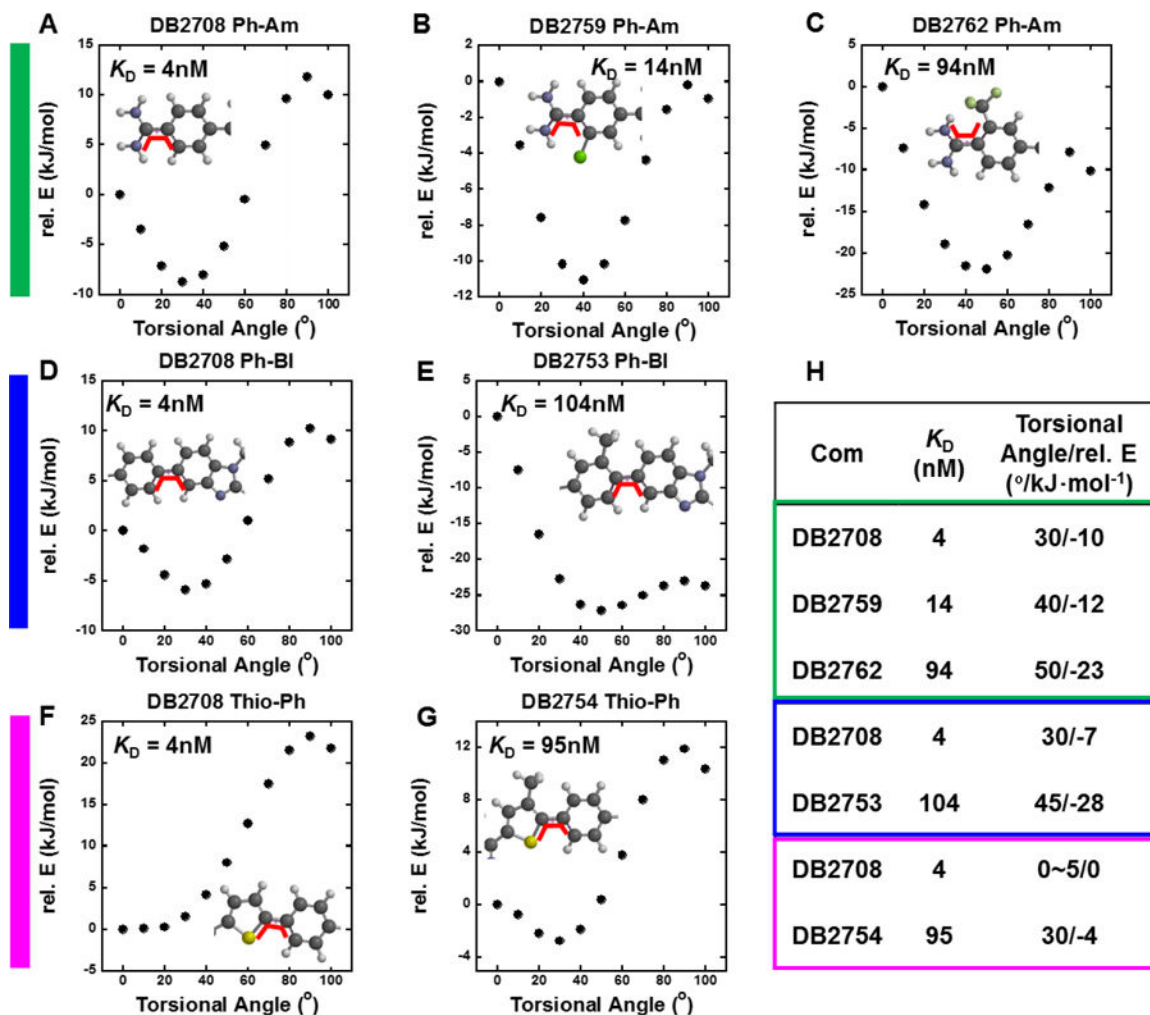


Figure 4.

(A, B, C, green label) Torsional angle maps of a Phenyl (Ph)-amidine (Am) bond for DB2708, DB2759, and DB2762; (D, E, blue label) a Ph-BI bond for DB2708 and DB2753; and (F, G, purple label) a Thiophene (Thio)-Ph bond for DB2708 and DB2754. A comparison of the K_D values and torsional angles/rel. Es for the three aromatic substituted compounds versus the unsubstituted control compound, DB2708 is shown in H. The torsional angles at minimum relative energies and the energies are also shown in H. The energy required to move the compounds from the structurally most stable state to a 0° torsional angle is the minimum relative E in H and the plots. All calculations are performed at the B3LYP/6-31G* level of theory. The range of dihedral is from 0° – 100° . The scanned dihedral angle calculations are shown as the bold red line at 0° in each plot.

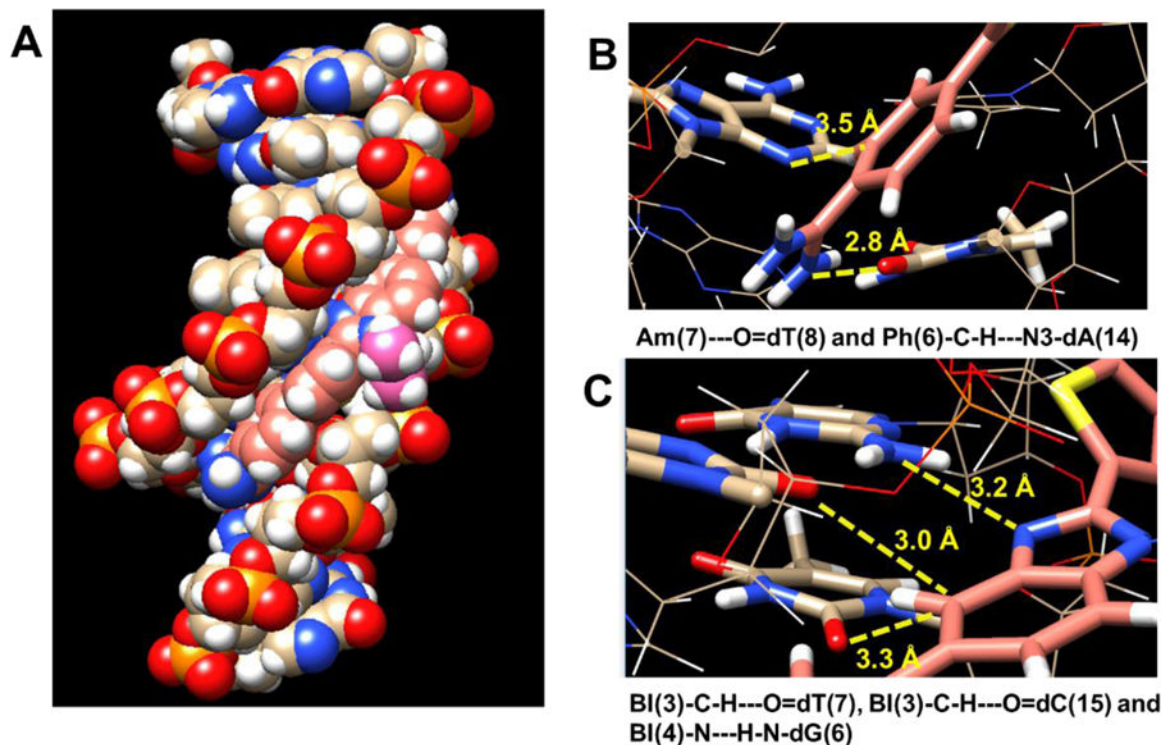
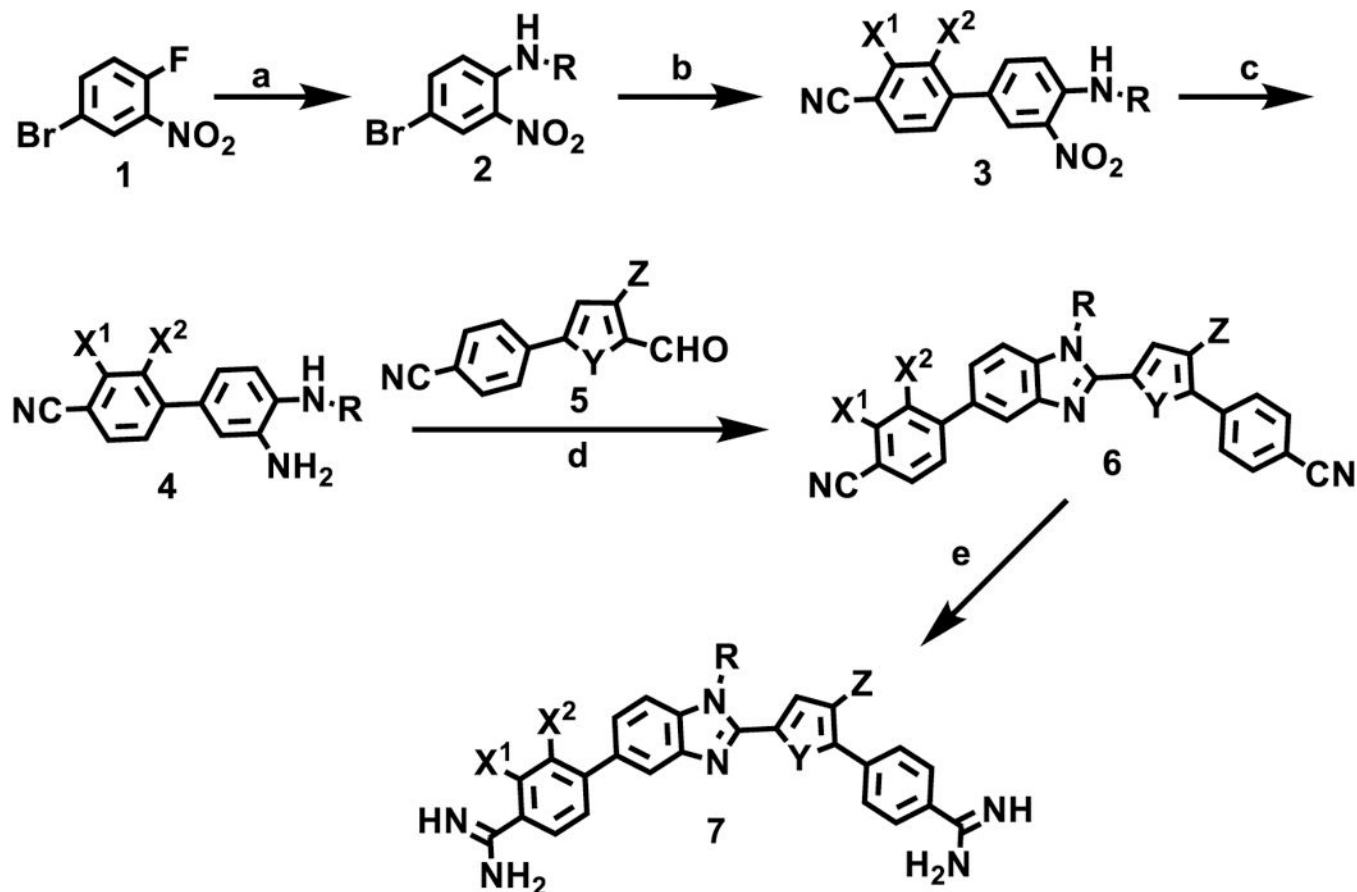


Figure 5.

Molecular Dynamics (MD) model of DB2708 bound to an AAAGTTT site: (A) A space-filling model viewed into the minor groove of the AAAGTTT binding site with bound DB2708. The DNA bases are represented in tan-white-red-blue-yellow(C-H-O-N-P) color scheme and DB2708 is light pink-white-blue-yellow (C-H-N-S) color scheme with the *N*-isopropyl group facing out of the minor groove in a brighter-pink. The important interactions between different sections of the DB2708-DNA complex are illustrated in (B) and (C). The terminal amidine group forms a strong hydrogen bond with the carbonyl group of a dT (yellow dashed lines), and another direct stabilizing interaction is observed from a Ph-CH group of DB2708 which points to N3 of a dA at the floor of the minor groove (B). In (C), the imidazole-*N* makes a strong hydrogen bond interaction with the exocyclic NH of dG. The BI-C-H that points to the floor of the minor groove forms strong interactions with the carbonyl group of a dT in the minor groove and this unit also points towards the carbonyl group of a dC with close distances.



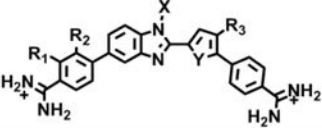
Reagents and conditions: a) RNH₂/EtOH or Cs₂CO₃/DMA; b) 4-cyanophenylboronic acid, Pd(PPh₃)₄, Na₂CO₃/H₂O, dioxane, reflux; c) SnCl₂/EtOH, reflux; d) Na₂S₂O₅/ DMSO, 130 °C; e) i- LiN(TMS)₂/THF, ii- HCl/EtOH.












Scheme 1.

The general synthetic approach to the target mixed bp binding diamidine compounds

Table 1.

Summary of Binding Affinity (K_D , nM) for the Interaction of All Test Compounds with Biotin-labeled DNA Sequences using Biosensor-SPR Method^a



Compounds	R ₁	R ₂	R ₃	X	Y	K_D (nM) /Selectivity Ratio AAATTT	K_D (nM) AAAGTTT	K_D (nM) /Selectivity Ratio AAAGCTTT
DB2457 (7a)	H	H	H	Me	S	222 / 56	4	192 / 48
DB2737 (7b)	H	H	H	Et	S	750 / 250	3	145 / 48
DB2708 (7c)	H	H	H	<i>i</i> -Pr	S	1020 / 255	4	223 / 56
DB2711 (7d)	H	H	H		S	NB / --	13	NB / --
DB2718 (7e)	H	H	H		S	NB / --	31	690 / 22
DB2715 (7f)	H	H	H		S	NB / --	10	1080 / 108
DB2728 (7g)	H	H	H		S	NB / --	7	237 / 34
DB2764 (7h)	H	H	H		S	NB / --	11	347 / 32
DB2726 (7i)	H	H	H		S	NB / --	5	180 / 36
DB2714 (7j)	H	H	H		S	NB / --	9	1330 / 148
DB2727 (7k)	H	H	H		S	NB / --	14	388 / 28
DB2738 (7l)	H	H	H		S	NB / --	21	NB / --
DB2740(7m)	H	H	H		S	NB / --	3	107 / 36
DB2747 (7n)	H	H	H		S	NB / --	3	130 / 43
DB2759 (7o)	Cl	H	H	<i>i</i> -Pr	S	NB / --	14	NB / --
DB2762 (7p)	CF ₃	H	H	<i>i</i> -Pr	S	NB / --	94	NB / --
DB2753 (7q)	H	Me	H	<i>i</i> -Pr	S	NB / --	104	NB / --
DB2754 (7r)	H	H	Me	<i>i</i> -Pr	S	NB / --	95	NB / --
DB2673 (7s)	H	H	H	Me	Se	650 / 77	9	300 / 33
DB2712 (7t)	H	H	H	<i>i</i> -Pr	Se	NB / --	7	290 / 41

^a. All the results in this table were investigated in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.05% P20, pH 7.4) at a 100 $\mu\text{L min}^{-1}$ flow rate. NB means no detectible binding. The listed binding affinities are an average of two independent experiments carried out with two different sensor chips and the values are reproducible within 10% experimental errors.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.Comparison of SPR Binding Kinetics Results for Selected Compounds^a

Compounds	$k_a(\text{M}^{-1} \text{s}^{-1})$	$k_d(\text{s}^{-1})$	K_D (nM)
DB2457 (7a)	$2.0 \pm 0.2 \times 10^6$	0.0081 ± 0.001	4.1 ± 0.4
DB2708 (7c)	$3.5 \pm 0.3 \times 10^6$	0.0150 ± 0.002	4.3 ± 0.3
DB2711 (7d)	$4.1 \pm 0.3 \times 10^6$	0.0516 ± 0.004	12.6 ± 0.1
DB2715 (7f)	$4.6 \pm 0.4 \times 10^6$	0.0444 ± 0.004	9.6 ± 0.4
DB2728 (7g)	$2.9 \pm 0.1 \times 10^6$	0.0217 ± 0.001	7.4 ± 0.2
DB2764 (7h)	$1.8 \pm 0.2 \times 10^6$	0.0198 ± 0.002	11.0 ± 0.2
DB2714 (7j)	$3.4 \pm 0.2 \times 10^6$	0.0290 ± 0.002	8.5 ± 0.1
DB2740 (7m)	$2.3 \pm 0.3 \times 10^6$	0.0078 ± 0.001	3.4 ± 0.3
DB2753 (7q)	$7.3 \pm 0.2 \times 10^4$	0.0076 ± 0.001	104 ± 2
DB2712 (7t)	$2.6 \pm 0.3 \times 10^6$	0.0180 ± 0.002	7.0 ± 0.2

^aKinetics analysis was performed by global fitting with a 1:1 binding model. k_a is the association rate constant, while k_d is the dissociation rate constant. K_D was calculated through global fitting of the kinetic data obtained for various concentrations of the compounds; K_D is given by k_d / k_a . The kinetics errors are based on two independent experiments and the data fitting of different concentrations of compounds. The data of DB2457, DB2711, DB2714 and DB2740 are based on three independent experiments.