The Thiophene "Sigma-Hole" as a Concept for Preorganized, Specific Recognition of G•C Base Pairs in the DNA Minor Groove

Dr. Pu Guo[a], **Dr. Ananya Paul**[a], **Dr. Arvind Kumar**[a], **Dr. Abdelbasset A. Farahat**[a,b], **Dr. Dhiraj Kumar**[a], **Dr. Siming Wang**[a], **Prof. Dr. David W. Boykin**[a], and **Prof. Dr. W. David Wilson**[a]

[a]Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30303-3083, USA.

[b]Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt.

Abstract

In spite of its importance in cell function, targeting DNA is under-represented in the design of small molecules. A barrier to progress in this area is the lack of a variety of modules that recognize G•C base pairs (bp) in DNA sequences. To overcome this barrier an entirely new design concept for modules that can bind to mixed GC and AT sequences of DNA is reported. Because of their successes in biological applications, minor-groove-binding heterocyclic cations were selected as the platform for design. Binding to AT sequences requires H-bond donors while recognition of the G-NH$_2$ requires an acceptor. The concept that we report here uses preorganized *N*-methylbenzimidazole (*N*-MeBI) thiophene modules for selective binding with mixed bp DNA sequences. The interaction between the thiophene sigma-hole (positive electrostatic potential) and electron donor nitrogen of *N*-MeBI preorganizes the conformation for accepting an H-bond from G-NH$_2$. The compound-DNA interactions were evaluated with a powerful array of biophysical methods and the results show that *N*-MeBI-thiophene monomer compounds can strongly and selectively recognize single G•C bp sequences. Replacing the thiophene with other moieties significantly reduces binding affinity and specificity, as predicted by the design concept. These results show that the use of molecular features, such as sigma-hole, can lead to new approaches for small molecules in biomolecular interactions.

Graphical abstract

A new concept of Sigma-hole is applied in design of N-methylbenzimidazole-thiophene compounds to specifically recognize G•C base pair in the DNA minor groove.
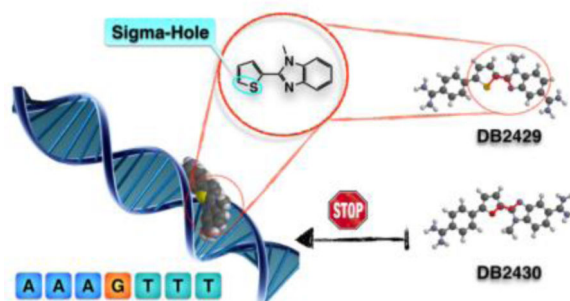
Correspondence to: W. David Wilson.

Introduction

Design and preparation of new types of compounds that recognize mixed base pairs (bps) nucleic acid sequences is a major goal for the use of DNA as a cellular receptor for therapeutics and biotechnology applications as well as for development of gene specific probes.[1–8] Such compounds, for example, are being designed as transcription factor inhibitors by binding to a DNA promoter sequence rather than by targeting the transcription factor part of the complex.[9–14] There are almost unlimited numbers of applications of gene specific probes of this type. The current lack of a variety of small molecules that can bind strongly and sequence specifically to mixed A•T and G•C bp sequences of DNA is, however, a barrier to progress in this field. With the current increasing availability of both gene sequences and knowledge of their functions from bacteria to humans, there is clearly an unmet need for new sequence specific, small molecule DNA probes.[7,9,15] Significant variations in the cell uptake potential of molecular structures and pharmacokinetic differences will also require discovery and development of a diversity of different molecular systems for DNA recognition that do not now exist.[16–19]

Compounds: design concepts

We report here a new concept for the modular design of novel mixed sequence DNA-binding compounds that focuses on both the DNA minor groove structure and chemistry as well as on key chemical and structural features of the designed compounds.[3,20–22] The compounds, for example, should have appropriate curvature to fit the convex shape of the minor groove.[21–25] Although not essential, cationic groups have been attached at one or both ends of the new compounds to help solubility and increase the electrostatic interaction with negatively charged DNA backbone.[13,21,25–27] Amidine cationic groups that have been quite successful for targeting DNA, provide a H-bond donating, planar surface for optimum interactions with A•T bps as flanking sites of G•C bps.[24,28,29] An amidine attached to a six member ring, Figure 1, will also have a slight twist that helps the compound track the helical curvature of the minor groove. For the design of G•C bp binding capability, H-bond acceptor functions are needed to form a specific interaction with the G-NH$_2$ groups that project into

the minor groove.[15,30,31] Linked heterocyclic systems have been used in our design schemes because of their excellent cell uptake properties,[9,27] flexibility, and strong molecular interaction potential, particularly with the DNA minor groove.[26,30] The goal is to design modules that can be linked in a variety of ways to strongly and specifically recognize a range of target DNA sequences. Many well characterized A•T recognition units have been reported, such as benzimidazole/indole,[32] phenyl-amidine with numerous types of linking groups, such as -O-, -NH-, furan and others,[26,28,33,34] pyrrole-amide,[6,35–37] and alkylamines,[38] however, very few types of molecules and modules that specifically recognize one or more G•C bps have been reported.[15,34,37,39] Pyridine-based H-bond acceptor units (such as DB2120, Figure 1A),[30,34] and azabenzimidazole (azaBI) H-bond acceptors (DB2277, Figure 1A),[31,40] have recently been designed to recognize G•C bp and complement the polyamide imidazole group[4]. Dervan's group and others have expanded the five-member hydrocycle used in polyamide. The amide-hydrocycle unit has also been replaced with BI or azaBI to recognize A·T or G·C bps.[4,41–44] Our understanding of G•C bp recognition by small molecule minor groove binders remains very limited, however, and additional molecular structures are needed to not only expand the use of the DNA as a cellular drug receptor but also the compound uptake potential in different organisms and cell types

The new design concept that we report here obeys the "rules" described above but is based on the hypothesis that an *N*-methylbenzimidazole (*N*-MeBI) module with the methyl pointed away from the minor groove floor would make an excellent new G-NH$_2$ recognition unit. Unfortunately, none of our original designs, such as DB1454 in Figure 1A, gave significant strong and specific binding to G•C bps. Analysis of *ab initio* molecular models suggested that an *N*-MeBI-thiophene module could fit the minor groove shape[24] to successfully recognize G•C bps based on interactions between low-lying σ* orbitals (the "sigma-hole") with positive electrostatic potential and electron donors such as the nitrogen atom of *N*-MeBI.[45–47] Such intramolecular interactions can favorably modulate the conformational preferences of a molecule.[48] The conformation that results from these intramolecular 1,4 N-S interactions of the thiophene and adjacent nitrogen heterocyclic ring pre-organizes the module to point the unsubstituted N of *N*-MeBI into the minor groove floor for accepting a H-bond from the minor groove G-NH$_2$.

From extended compound designs (Figure 1B), synthetic studies, and biophysical experiments described here, we show that the *N*-MeBI-thiophene motif can provide excellent G•C recognition ability in an appropriate motif. Unlike classical heterocyclic thiophene derivatives such as DB818 (Figure 1A),[23] the new *N*-MeBI-thiophenes bind significantly more strongly to sequences with G•C bps than to pure AT sequences.

Extensions and tests of the Sigma-Hole design concept for DNA recognition

To test the key role of the thiophene in DB2429, furan (DB2430, DB2501), pyridine (DB2465) and phenyl (DB1454) analogs were prepared. The importance of the diamidines was investigated with the mono-amidines, DB2432 and DB2454. The significance of molecular shape and substituents was probed with DB2464, DB2325, DB1300 and DB2457. The role of the *N*-MeBI in DB2429 has been evaluated with benzoxazole DB847. Extensive

biophysical analysis of the DNA interactions of these compounds show very clearly that the *N*-MeBI-thiophene module provides an essential basic recognition unit for mixed bps DNA sequence recognition in this compound set. The results validate the use of the sigma-hole concept in the design of DNA targeted agents. This type of compound preorganization to recognize target sites can significantly increase affinity and specificity independent of compound-receptor contacts and is a valuable new concept in DNA recognition modules. (Figure 1B)

Results and Discussion

Compound synthesis

The syntheses of the key *N*-methylbenzimidazole heterocycles are outlined in Scheme 1. Reaction of the readily available bromoheteroaryl aldehydes (1) with 4-cyanophenylboronic acid under standard Suzuki coupling[49] conditions conveniently provides the 5-(4-cyanophenyl)-2-formyl 5-ring heterocycles (2) which can be used to prepare both the *tri*- and *tetra*-heteroaryl target molecules 4 and 6. Oxidative condensation and cyclization, mediated by sodium metabisulfite, of the aldehydes 2 with 3-amino-4-(methylamino) benzonitrile gives the *bis*-nitriles 3 in acceptable yields.[50] The *bis*-nitriles 3 are readily converted into the desired diamidines 4 by the action of LiN(TMS)$_2$ in THF.[51] In a similar sequence the aldehydes 2 are allowed to react with 3'-amino-4'-(methylamino) [1,1'-biphenyl]- 4-carbonitrile, again mediated by sodium metabisulfite, to produce the *bis*-nitriles 5 which are converted to the diamidines 6 as previously described. The supplemental information contains the experimental details for the compounds described in Scheme 1 as well as the other novel compounds in Figure 1B used in this study.

Thermal melting (T_m): relative binding affinity

DNA thermal melting experiments provide a rapid qualitative evaluation of the relative binding affinities of compounds with selected DNA sequences.[52] Small molecules binding with DNA generally enhance the duplex stability and an increase in the DNA melting temperature upon the addition of ligands which can be used as a screen for relative binding affinity.[11] Hairpin DNA oligomers, which have monomolecular melting transition (Figure 1) were chosen as model sequences to study relative binding affinities. The benzimidazole-thiophene DB818 (Figure 1A) has very strong binding with pure A-tract sequences, as previously reported.[24] The T_m results of DB2429 and its analogues are shown in Table 1. DB2429, the parent *N*-MeBI-thiophene after replacing the benzimidazole of DB818 with the *N*-MeBI moiety, has a high T_m (11 °C) with the target single G•C bp containing AAA**G**TTT sequence but weaker interactions with AAATTT (T_m=5 °C) and AAA**GC**TTT (T_m = 5 °C). The insertion of a phenyl ring between the *N*-MeBI and an amidine group of DB2429 yields DB2457 with a slightly increased T_m value with AAA**G**TTT (T_m=12 °C). Interestingly, the addition of a 3-methyl group into DB2429 increased the sequence selectivity due to a much lower affinity with the AAATTT and AAA**GC**TTT sequences. Modification of the thiophene group to furan (DB2430) and pyridine (DB2465) resulted in compounds with much weaker binding to single G sequences. To evaluate the effects of the *N*-MeBI without an adjacent thiophene ring, DB1454 was compared with DB2429 and found to have very low binding with the three test sequences. From these results it is clear

that the special molecular properties of the *N*-MeBI-thiophene moiety help to enhance DNA minor groove binding ability and specificity. The benzoxazole, DB847, was considered to be a possible additional way to use the thiophene sigma-hole-N interaction to provide single G•C bp binding but the compound gave relatively moduate T_m increase (AAA**G**TTT, $T_m =$ 7 °C). Molecular modeling, described below, provides a possible explanation for these results. In order to potentially expand the application of the *N*-MeBI-thiophene motif, DB1300 and DB2325 with two *N*-MeBI-thiophene modules, were studied. The binding activities of the compounds with AAA**G**TTT are relatively low and there is no significant interaction with the two G•C bps containing sequences. It seems likely that the curvature and placement of H-bond acceptors in these compounds does not properly index with donors on bp at the floor of the minor groove.

Based on the excellent results for AAA**G**TTT binding with *N*-MeBI-thiophenes, DB2429 and DB2457 were evaluated with sequences AAT**G**AAT, ATA**G**TAT and AATT**G**AATT (Supporting Information Table S1). For both compounds the binding with all three DNAs was weaker than with AAA**G**TTT. In summary, with the compounds in Figure 1B, only the mono-*N*-MeBI-thiophene sigma-hole diamidines give strong and specific binding to the AAA**G**TTT mixed bp DNA sequence.

Biosensor-surface plasmon resonance (SPR): binding affinity and stoichiometry

Biosensor-SPR methods provide an excellent way to quantitatively evaluate the interaction of small organic molecules with immobilized biomolecules.[53] SPR provides sensitive real-time progress of the binding reaction as well as the equilibrium binding affinity, kinetics, and stoichiometry of complex formation.[21,26] Based on the T_m results, the interactions of DB2429 with AAATTT, AAA**G**TTT and AAA**I**TTT were evaluated by SPR (Figure 2). As can be seen in Figure 2A, DB2429 binds strongly with the single G•C bp containing sequence. Global kinetics fitting yielded a single binding site and an approximate K_D of 50 nM for DB2429 with an AAA**G**TTT sequence. The reactions are relatively fast and the RU response in the plateau region was also plotted *vs.* the free compound concentration to determine the K_D values for other sequences. The maximum value (RU$_{max}$) allows determination of stoichiometry.[53] The AAA**I**TTT sequence was used to evaluate, in a more direct fashion, the influence of the G-NH$_2$ group on binding in the minor groove. Based on the RU$_{max}$, both the single "G" and "I" sequences had only one binding site. The AAATTT sequence appeared to have a quite weaker second binding site that could not be saturated under our conditions. As with the T_m results, DB2429 has much stronger binding with AAA**G**TTT than AAATTT sequences. The K_D value of AAA**G**TTT binding with DB2429 is quite strong (50 nM) compared to the K_D values of AAATTT (524 nM) and AAA**I**TTT (322 nM) (Figure 2C). This difference shows that DB2429 binds to AAA**G**TTT with excellent selectivity and that the G-NH$_2$ is an essential component of the strong interaction.

Based on the results with DB2429, several additional analogues were evaluated with AAATTT, AAA**G**TTT, and AAA**GC**TTT binding sites by using biosensor-SPR (Table 2). DB2457, the extended *N*-MeBI-phenyl analogue, shows 10 times stronger binding affinity than DB2429 with AAA**G**TTT (Table 2). The differences in binding affinity are most clearly seen in the global kinetic fitting of the dissociation phase of the sensorgrams (Figure 2).

With minor groove binders that are structurally related, differences in affinity can be related to changes in either the on or off rate or some combination. With DB2429 and DB2457 very similar on-rates ($k_a = 1.3 \pm 0.25 \times 10^6$ M^{-1} s^{-1}) are observed while the off-rate k_d for DB2457 is 10-fold lower ($0.51 \pm 1.3 \times 10^{-2}$ S^{-1}) than for DB2429 ($k_d = 6.5 \pm 1.8 \times 10^{-2}$ S^{-1}). As a result of this slow off-rate, DB2457 is the strongest binder for single G•C bp containing sequences in this set of molecules. The extra phenyl group not only facilitated the higher binding ability for AAA**G**TTT but also increased the sequence selectivity of 50 times over the pure AT sequence (K_D = 222 nM for AAATTT). DB2464, the 3-methylthiophene compound, however, has lower binding affinity (K_D = 70 nM) with AAA**G**TTT in agreement with the T_m results. The three related *N*-MeBI compounds showed much lower affinity for the "I" containing sequence (Table 2). All of these results indicate that the *N*-MeBI moiety is playing a crucial role in G-NH$_2$ recognition and DNA sequence selectivity.

The binding of DB2429, DB2464, and DB2457 were also determined with a single G•C bp and other AT flanking sequences (ATA**G**TAT, AATT**G**AATT). Interestingly, relative to AAA**G**TTT, the compounds show weaker binding towards all other sequences (Table 2). Other compounds were also tested by SPR experiments with AAATTT, AAA**G**TTT, and AAA**G**CTTT. In agreement with the T_m results, DB2432, DB847 and DB1300 all have weak to no binding with AAA**G**TTT and other sequences. Surprisingly, DB2430 with an *N*-MeBI-furan motif shows higher binding affinity for AAATTT (K_D = 74 nM) over AAA**G**TTT (K_D = 180 nM). DB2465 with an *N*-MeBI-pyridine motif has quite weak binding to all test sequences (Table 2). The SPR results are all correlated with the T_m experiments.

In summary, an *N*-MeBI adjacent to a phenyl, a pyridine or a furan ring has very poor binding with the target AAA**G**TTT sequence. Our optimum thiophene sigma-hole compound, DB2457, however, has a K_D with AAA**G**TTT of 4 nM and a K_D with AAATTT of >200 nM. Clearly the *N*-MeBI-thiophene is an effective new module for strong, single G•C bp specific recognition.

Fluorescence emission spectroscopy: sequence dependent fluorescence differences

Fluorescence emission spectroscopy is an effective and sensitive method to study the binding between DNA and fluorescent small molecules[54] and to monitor the cellular location of compounds in culture.[10,28,55] The fluorescence titration results of DB2429 and DB2430 with sequences AAATTT, AAA**G**TTT and AAA**G**CTTT are shown in supporting information (Figure S3). The fluorescence intensity (F.I.) of DB2429 was quenched by all three test sequences but to different extents. The AAA**G**TTT sequence showed the highest level of F.I. decrease and the spectra reached the saturation state near a 1:1 [ligand/DNA] ratio, indicating monomer complex formation in agreement with SPR results. The intensity of the furan, DB2430, was quenched by AAA**G**TTT and AAA**G**CTTT to a smaller extent than DB2429. In summary, adding a G•C bp to the AAATTT sequence causes a marked decrease in fluorescence for all test compounds.

Circular dichroism (CD): probing the binding mode and ratio

CD titration experiments as a function of compound concentration were evaluated to monitor the binding mode and the saturation limit for compound binding with DNA sequences (Figure 3). CD spectra monitor the asymmetric environment of the compounds binding to DNA and therefore can be used to obtain information on the binding mode.[56,57] There are no CD signals for the free compounds but on the addition of the compounds into DNA, substantial positive induced CD signals (ICD) arise in the compound absorption region between 300 and 450 nm wavelength. These positive ICD signals indicate a minor groove binding mode by these molecules as expected from their structures. A monotonic increase of ICD signals at the range of 320 to 440 nm wavelength was observed by incremental addition of DB2429, DB847, and DB2432 to the AAA**G**TTT sequence. As can be seen from Figure 3A and 3B, DB2429 and DB847 are binding in the minor groove of the AAA**G**TTT sequences with 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.

Competition electrospray ionization mass spectrometry (ESI–MS): binding stoichiometry and relative binding affinity

Mass spectrometry is an excellent method for the evaluation of relative binding affinity,[26,30,40] stoichiometry and the validation of experimental results obtained from other methods, such as SPR, where macromolecule-ligand stoichiometry is obtained by fitting the signal at different concentrations.[50,58] In this case the *N*-MeBI- thiophene compounds are far from ideal and a considerable amount of the compounds are adsorbed on the surface of the injector tube. Satisfactory results were obtained with the DB2429 complex to evaluate stoichiometry with the AAA**G**TTT sequence (Supporting Information Figure S4). The AAA**G**TTT sequence has a molecular weight of 8539 g/mol and this is observed as a single peak in the mass spectrometry results. With the addition of the compound DB2429 (mass of 374.5 g/mol) a new peak was observed at $m/z = 8913$ (difference of 374) indicating a 1:1 [ligand/DNA] complex formation which agrees with SPR and other results. For understanding the sequence specificity and binding stoichiometry of DB2429 with additional sequences, a competition ESI-MS experiment was performed (Figure 4).[58] Figure 4A represents the signal of three free DNA sequences, AAATTT, AAA**G**TTT, and AAA**G**CTTT at their molecular mass positions. With the addition of DB2429 the peak of AAA**G**TTT disappeared with the simultaneous appearance of a new peak at $m/z = 8914$ as a 1:1 DB2429-AAA**G**TTT complex. There was no signal of other ligand-DNA complex peaks observed which indicates excellent specificity of DB2429 with the AAA**G**TTT sequence. The signal to noise in these experiments is less than usual due to the problem of thiophene compound adsorption to the injector tube. It is clear, however, in agreement with the T_m and SPR results, that ESI-MS also shows that DB2429 can selectively recognize single G•C bp sequences with 1:1 stoichiometry.

Structural calculations and molecular docking: explanations for binding modes and mechanism

A specific question for the computational methods is: what is the low energy conformation of the test compounds and how well does this conformation match the minor groove? To help answer these, torsional angle maps and molecular conformations of DB2429, DB2430, DB847, and DB2465 were calculated (Figure 5). DB2429 has the lowest energy conformation at a torsional angle of 9° and this most stable structure of DB2429 has an appropriate shape and the G-NH$_2$-N (H to N) H-bonding ability for the DNA minor groove (Figure 5A). Interestingly, DB2430, the furan derivative, reached the low energy state around a 180° torsional angle, the reverse orientation with the N-methyl group directed in the same orientation as the –O– of furan (Figure 5B). The N-methyl group is a steric block in this orientation and will not allow a hydrogen bond between the N-MeBI with G-NH$_2$ to form. Results from SPR and T_m experiments show that DB2429 has a much stronger affinity with the AAA**G**TTT sequence than DB2430 in agreement with the observed structures. These results for the thiophene and furan compounds are also in complete agreement with the sigma-hole concept.[45–48] The torsional angle map of DB847 (Figure 5C), with the benzoxazole group, revealed that an orientation at 180° has a similar energy to 0°. As the SPR and T_m results show, DB847 exhibits intermediate binding affinity with the AAA**G**TTT sequence in agreement with the mixed orientational results. As with the furan derivative, the same side orientation of the N-methyl group and the "N" of pyridine in DB2465 represents the most stable structure (Figure 5D). The energy of the opposite orientation is quite high, due to the repulsive van der Waals interaction between the N-methyl group and a pyridine C-H. In agreement with the conformational predictions, DB2465 has the weakest binding ability with the AAAGTTT sequence among all the test compounds in this set (Table 2).

To evaluate ideas for how DB2429 and DB2457 are able to effectively bind to the AAA**G**TTT sequence, a molecular docking study was conducted with ds-[(5'-CCAAA**G**TTTG-3') (5'-CAAACTTTGG-3')] sequence using the Autodock software 4.2 package.[59] Low energy binding complexes were obtained for both compounds with the unsubstituted "N" of N-MeBI and "S" of thiophene forming bifurcated H-bonds to the exocyclic minor groove G-NH$_2$ group (Figure 6B, S6B). The top amidine in Figure 6 and S6 participates in an H-bond with a thymine (T) carbonyl group at the complementary strand of DNA (the C-strand of the G•C bp) with 2.4 Å bonding distances for both DB2429 and DB2457. Interestingly, given this strong binding it is impossible to obtain simultaneous H-bonds by both amidine groups with A•T bps at the floor of the groove. The strong binding affinity and sequence selectivity of these two compounds can be explained by two possibilities: (i) a conformational change of the local DNA structure to allow the amidine and the N-MeBI to form an H-bond, (ii) if the DNA cannot undergo such a change due to the energy cost, a water molecule could be incorporated into the complex. In this case the non-H-bonded amidine is linked to base pair acceptor groups at the floor of the groove by the water molecule.

We are unable to model the duplex DNA conformation in model (i) above, but we were able to construct a stable complex by including a water molecule at the floor of the minor groove. A bound water has been observed for linking the biphenyl benzimidazole diamidine, DB921,

to the floor of the minor groove to form a strong complex.[32] One amidine group can form a direct H-bond to the DNA while the other is too far from the minor groove. An X-ray crystal structure clearly showed a water molecule completing the link between that amidine and N3 of Adenine (A) at the floor of the minor groove. Given this observation, we suggest that complexes with DB2429 and DB2457 involve a bridging water molecule between the compound and the floor of the minor groove (Figure 6, S6). Future work on NMR and molecular dynamics (MD) studies will be applied to address this issue.

The thiophene and furan *ab initio* models in Figure 6C, 6D illustrate the reversed equilibrium conformation of these systems as in Figure 5. The electrostatic potential maps provide an explanation for this observation. The thiophene C-S bond results in two electron deficient σ* orbitals on S. The *N*-MeBI unsubstituted N is in the close proximity with S and increases electron density from the lone pair electron of N to S. This can be seen in the diagram by the asymmetric electron density potential map on S. The *N*-MeBI side is significantly more electronegative than the phenyl side. The same interaction is not possible in the furan compound. In this case, the dominant inter-ring interaction is the negative potential on both the unsubstituted N and the furan O. These are rotated 180° apart to minimize the repulsive interaction. For the thiophene, the S-N σ-hole interaction provide the appropriate conformation for binding in the minor groove of DNA, while for the furan the O-N repulsion give an unfavorable conformation.

Conclusions

New design concept for molecular structures with quite different properties that can recognize mixed AT and GC DNA sequences will help to deal with the problem of a lack of a variety of compounds that can specifically target DNA. In this report we have used an entirely new concept to design and prepare compounds that can recognize G•C bps. By using the orienting interaction of a "S" of thiophene sigma-hole with the unsubstituted "N" of *N*-MeBI, a pre-organized module for the DNA minor groove is obtained that can accept an H-bond from the G–NH$_2$. SPR, ESI-MS, fluorescence emission spectrometry and CD experiments all indicate that these *N*-MeBI-thiophene compounds bind specifically to a G•C bp with flanking A•T bps in a strong 1:1 minor groove complex. When the thiophene is replaced by a furan or pyridine moiety, the reverse orientation of the "O" in furan or "N" in pyridine and unsubstituted "N" in *N*-MeBI is favored and there is an energy barrier to convert the module to the H-bonding conformation with G-NH$_2$. In agreement with SPR results, the furan and pyridine compounds bind more weakly to single G•C bp containing sequences than the thiophenes.

The rationally designed molecules, DB2429 and DB2457, are an important discovery in the design process for sequence specific DNA recognition. The large decrease in sequence specificity and affinity by either replacement of the thiophene or *N*-MeBI moiety demonstrates the importance of the *N*-MeBI-thiophene module in G•C bp recognition. The decreases in binding affinity for DB2429 and DB2457, after replacement of "G" by "I" in the minor groove of the DNA sequence are clear indications of a strong H-bond interaction between G-NH$_2$ and the central *N*-MeBI-thiophene moiety. This new type of G•C bp specific recognition provides ideas on expansion of modules for sequence specific DNA

recognition and in due course new therapeutics and biotechnology reagents. With these available units, we can now combine them with different linkers for recognition of expanded sequences with additional G•C base pairs.

Experimental Section

Biophysical experimental details, compound synthesis and characterisation data of the new compounds used in this article can be found in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB. Nature. 1997; 387:202–205. [PubMed: 9144294]

2. Dervan, PB.; Poulin-Kerstien, AT.; Fechter, EJ. DNA Binders and Related Subjects. Waring, MJ.; Chaires, JB., editors. Vol. 253. Berlin Heidelberg: Springer; 2005. p. 1-31.

3. Neidle S. Nat. Prod. Rep. 2001; 18:291–309. [PubMed: 11476483]

4. Satam V, Babu B, Patil P, Brien KA, Olson K, Savagian M, Lee M, Mepham A, Jobe LB, Bingham JP, Pett L, Wang S, Ferrara M, Bruce CD, Wilson WD, Lee M, Hartley JA, Kiakos K. Bioorg. Med. Chem. Lett. 2015; 25:3681–3685. [PubMed: 26122210]

5. Edwards TG, Vidmar TJ, Koeller K, Bashkin JK, Fisher C. PLoS ONE. 2013; 8:e75406. [PubMed: 24098381]

6. Blackledge MS, Melander C. Bioorg. Med. Chem. 2013; 21:6101–6114. [PubMed: 23665141]

7. Tidwell, RR.; Boykin, DW. Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes. Demeunynck, M.; Bailly, C.; Wilson, WD., editors. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA; 2002. p. 414-460.

8. Pazos E, Mosquera J, Vázquez ME, Mascareñas JL. Chembiochem. 2011; 12:1958–1973. [PubMed: 21805550]

9. Koehler AN. Curr. Opin. Chem. Biol. 2010; 14:331–340. [PubMed: 20395165]

10. Munde M, Wang S, Kumar A, Stephens CE, Farahat AA, Boykin DW, Wilson WD, Poon GMK. Nucleic Acids Res. 2014; 42:1379–1390. [PubMed: 24157839]

11. Munde M, Kumar A, Peixoto P, Depauw S, Ismail MA, Farahat AA, Paul A, Say MV, David-Cordonnier M-H, Boykin DW, Wilson WD. Biochemistry. 2014; 53:1218–1227. [PubMed: 24495039]

12. Rodríguez J, Mosquera J, García-Fandiño R, Vázquez ME, Mascareñas JL. Chem. Sci. 2016; 7:3298–3303. [PubMed: 27252825]

13. Rodríguez J, Mosquera J, Couceiro JR, Vázquez ME, Mascareñas JL. Chem. Sci. 2015; 6:4767–4771. [PubMed: 26290687]

14. Asamitsu S, Kawamoto Y, Hashiya F, Hashiya K, Yamamoto M, Kizaki S, Bando T, Sugiyama H. Bioorg. Med. Chem. 2014; 22:4646–4657. [PubMed: 25127467]

15. Kielkopf CL, Baird EE, Dervan PB, Rees DC. Nat. Struct. Biol. 1998; 5:104–109. [PubMed: 9461074]

16. Cominetti MMD, Goffin SA, Raffel E, Turner KD, Ramoutar JC, O'Connell MA, Howell LA, Searcey M. Bioorg. Med. Chem. Lett. 2015; 25:4878–4880. [PubMed: 26115576]

17. Snyder RD, Holt PA, Maguire JM, Trent JO. Environ. Mol. Mutagen. 2013; 54:668–681. [PubMed: 23893771]

18. Kumar S, Spano MN, Arya DP. Bioorg. Med. Chem. 2015; 23:3105–3109. [PubMed: 26025072]

19. Xie X, Choi B, Largy E, Guillot R, Granzhan A, Teulade-Fichou M-P. Chem. Eur. J. 2013; 19:1214–1226. [PubMed: 23292703]

20. Bishop EP, Rohs R, Parker SCJ, West SM, Liu P, Mann RS, Honig B, Tullius TD. ACS Chem. Biol. 2011; 6:1314–1320. [PubMed: 21967305]

21. Munde M, Lee M, Neidle S, Arafa R, Boykin DW, Liu Y, Bailly C, Wilson WD. J. Am. Chem. Soc. 2007; 129:5688–5698. [PubMed: 17425312]

22. White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB. Nature. 1998; 391:468–471. [PubMed: 9461213]

23. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. Nature. 2009; 461:1248–1253. [PubMed: 19865164]

24. Mallena S, Lee MPH, Bailly C, Neidle S, Kumar A, Boykin DW, Wilson WD. J. Am. Chem. Soc. 2004; 126:13659–13669. [PubMed: 15493923]

25. Vázquez O, Vázquez ME, Blanco JB, Castedo L, Mascareñas JL. Angew. Chem. Int. Ed. Engl. 2007; 46:6886–6890. [PubMed: 17674388]

26. Liu Y, Chai Y, Kumar A, Tidwell RR, Boykin DW, Wilson WD. J. Am. Chem. Soc. 2012; 134:5290–5299. [PubMed: 22369366]

27. Hahn L, Buurma NJ, Gade LH. Chem. Eur. J. 2016; 22:6314–6322. [PubMed: 26997208]

28. Wilson WD, Tanious FA, Mathis A, Tevis D, Hall JE, Boykin DW. Biochimie. 2008; 90:999–1014. [PubMed: 18343228]

29. Nanjunda R, Wilson WD. Current Protocols in Nucleic Acid Chemistry. 2012; Chapter 8(Unit8.8)

30. 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; 25:4927–4932. [PubMed: 26051649]

31. Harika NK, Paul A, Stroeva E, Chai Y, Boykin DW, Germann MW, Wilson WD. Nucleic Acids Res. 2016; 44:4519–4527. [PubMed: 27131382]

32. Liu Y, Kumar A, Depauw S, Nhili R, David-Cordonnier M-H, Lee MP, Ismail MA, Farahat AA, Say M, Chackal-Catoen S, Batista-Parra A, Neidle S, Boykin DW, Wilson WD. J. Am. Chem. Soc. 2011; 133:10171–10183. [PubMed: 21627167]

33. Vázquez O, Sánchez MI, Martínez-Costas J, Vázquez ME, Mascareñas JL. Org. Lett. 2010; 12:216–219. [PubMed: 20000489]

34. Sánchez MI, Vázquez O, Martínez-Costas J, Vázquez ME, Mascareñas JL. Chem. Sci. 2012; 3:2383–2387.

35. Kielkopf CL, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC. Science. 1998; 282:111–115. [PubMed: 9756473]

36. Saha A, Hashiya F, Kizaki S, Asamitsu S, Hashiya K, Bando T, Sugiyama H. Chem. Commun. 2015; 51:14485–14488.

37. Dervan PB, Edelson BS. Curr. Opin. Struct. Biol. 2003; 13:284–299. [PubMed: 12831879]

38. Koeller KJ, Harris GD, Aston K, He G, Castaneda CH, Thornton MA, Edwards TG, Wang S, Nanjunda R, Wilson WD, Fisher C, Bashkin JK. Med. Chem. (Los Angeles). 2014; 4:338–344. [PubMed: 24839583]

39. Buchmueller KL, Staples AM, Howard CM, Horick SM, Uthe PB, Le NM, Cox KK, Nguyen B, Pacheco KAO, Wilson WD, Lee M. J. Am. Chem. Soc. 2005; 127:742–750. [PubMed: 15643900]

40. Paul A, Chai Y, Boykin DW, Wilson WD. Biochemistry. 2015; 54:577–587. [PubMed: 25495885]

41. Chenoweth DM, Poposki JA, Marques MA, Dervan PB. Bioorg. Med. Chem. 2007; 15:759–770. [PubMed: 17095230]

42. Chenoweth DM, Viger A, Dervan PB. J. Am. Chem. Soc. 2007; 129:2216–2217. [PubMed: 17279754]

43. Brucoli F, Guzman JD, Maitra A, James CH, Fox KR, Bhakta S. Bioorg. Med. Chem. 2015; 23:3705–3711. [PubMed: 25921267]

44. Chavda S, Liu Y, Babu B, Davis R, Sielaff A, Ruprich J, Westrate L, Tronrud C, Ferguson A, Franks A, Tzou S, Adkins C, Rice T, Mackay H, Kluza J, Tahir SA, Lin S, Kiakos K, Bruce CD, Wilson WD, Hartley JA, Lee M. Biochemistry. 2011; 50:3127–3136. [PubMed: 21388229]

45. Clark TT. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2012; 3:13–20.

46. Murray JS, Lane P, Politzer P. Int. J. Quantum Chem. 2008; 108:2770–2781.

47. Tilly D, Chevallier F, Mongin F. Synthesis. 2015; 48:184–199.

48. Beno BR, Yeung K-S, Bartberger MD, Pennington LD, Meanwell NA. J. Med. Chem. 2015; 58:4383–4438. [PubMed: 25734370]

49. Farahat AA, Boykin DW. J. Heterocycl. Chem. 2013; 50:585–589.

50. Laughlin S, Wang S, Kumar A, Farahat AA, Boykin DW, Wilson WD. Chem. Eur. J. 2015; 21:5528–5539. [PubMed: 25703690]

51. Farahat AA, Kumar A, Say M, Barghash AE-DM, Goda FE, Eisa HM, Wenzler T, Brun R, Liu Y, Mickelson L, Wilson WD, Boykin DW. Bioorg. Med. Chem. 2010; 18:557–566. [PubMed: 20031421]

52. Shi X, Chaires JB. Nucleic Acids Res. 2006; 34:e14. [PubMed: 16432258]

53. Nanjunda, R.; Munde, M.; Liu, Y.; Wilson, W. Methods for Studying Nucleic Acid/Drug Interactions. Wanunu, Y., editor. CRC Press; 2011. p. 91-119.

54. Neubauer H, Gaiko N, Berger S, Schaffer J, Eggeling C, Tuma J, Verdier L, Seidel CAM, Griesinger C, Volkmer A. J. Am. Chem. Soc. 2007; 129:12746–12755. [PubMed: 17900110]

55. Wang M, Mao Z, Kang TS, Wong CY, Mergny JL, Leung, leung C-H, Ma D-L. Chem. Sci. 2016; 7:2516–2523.

56. Eriksson, M.; Nordén, B. Drug-Nucleic Acid Interactions vol. Vol. 340. Elsevier; 2001. p. 68-98.

57. Fornander LH, Wu L, Billeter M, Lincoln P, Nordén B. J. Phys. Chem. B. 2013; 117:5820–5830. [PubMed: 23607615]

58. Laughlin S, Wilson WD. Int. J. Mol. Sci. 2015; 16:24506–24531. [PubMed: 26501262]

59. Trott O, Olson AJ. J. Comput. Chem. 2010; 31:455–461. [PubMed: 19499576]
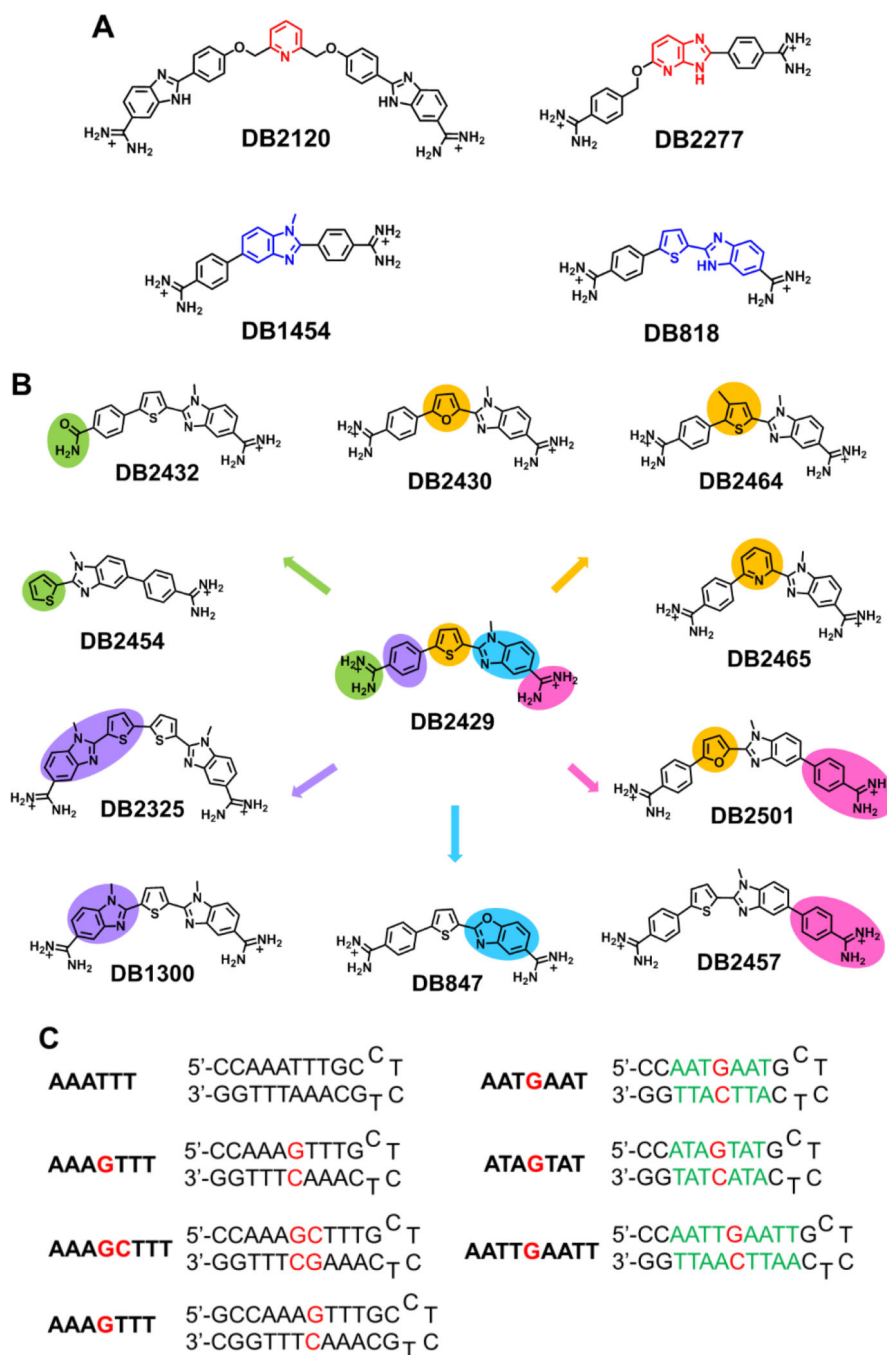
Figure 1.

(A) Chemical structure of reported single G•C bp binding molecules (top) and some stand molecules used in this study (below). (B) Chemical Structure of DB2429 and its analogues used in this study. Green: Mono-amidine; Orange: Furan, Pyridine, 3-methylthiophene; Pink: Extended Phenyl-amidine; Blue: Benzoxazole-thiophene module; Purple: *bis-N-* MeBI-thiophene modules. (C) The DNA sequences used in this study; DNA sequences with 5'-biotin labels were used for SPR studies.
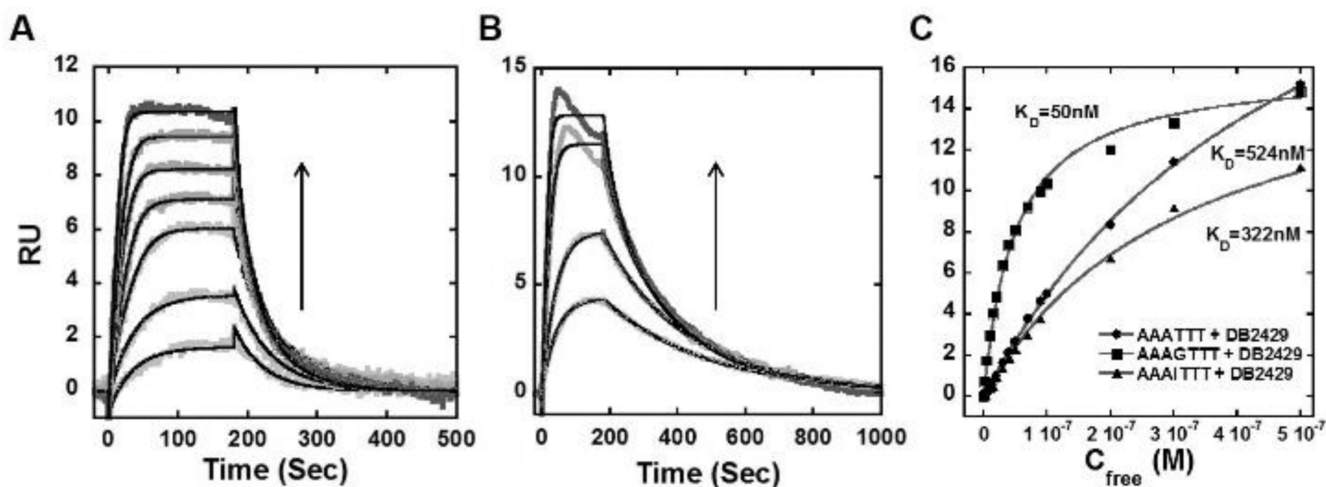
Figure 2.
Representative SPR sensorgrams for (A) DB2429 and (B) DB2457 in the presence of
AAA**G**TTT hairpin DNA, concentrations of DB2429 from bottom to top are 5, 10, 20, 30,
40, 50, and 70 nM, and for DB2457 from bottom to top are 5, 10, 20 and 30 nM; (C)
Comparison of steady-state binding plots for DB2429 with AAA**G**TTT, AAATTT and
AAA**I**TTT sequences. The data are fitted to a steady state binding function using a 1:1
model to determine equilibrium binding constants. In (A) and (B) the solid black lines are
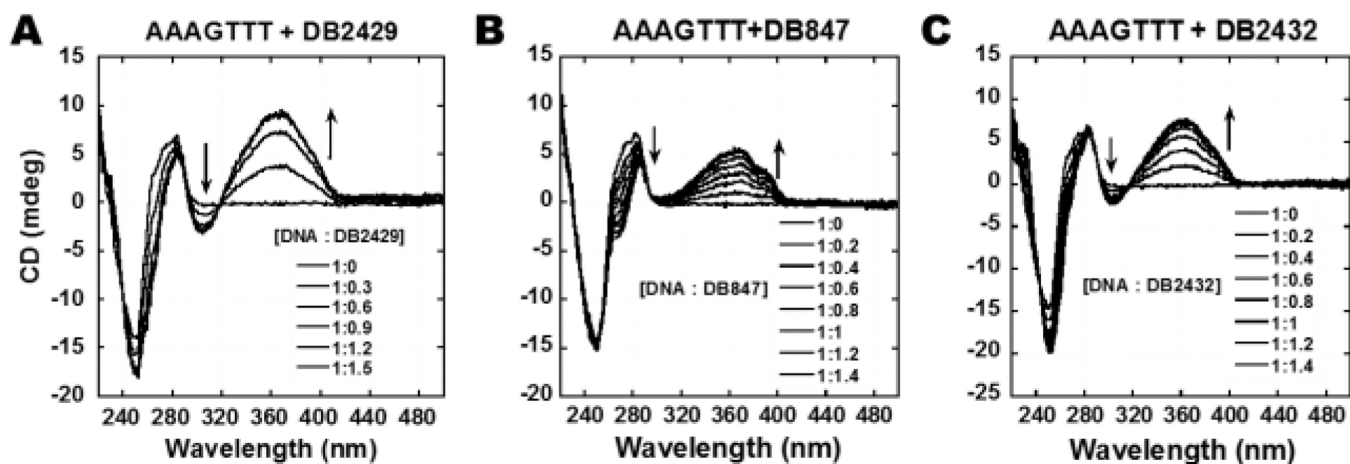best fit values for global kinetic fitting of the results with a single site function.

Figure 3.
Circular dichroism spectra of the titration of representative compounds, (A) DB2429, (B) DB847, and (C) DB2432 with a 5 μM AAA**G**TTT sequence in the T_m buffer at 25 °C. Arrows indicate the changes.
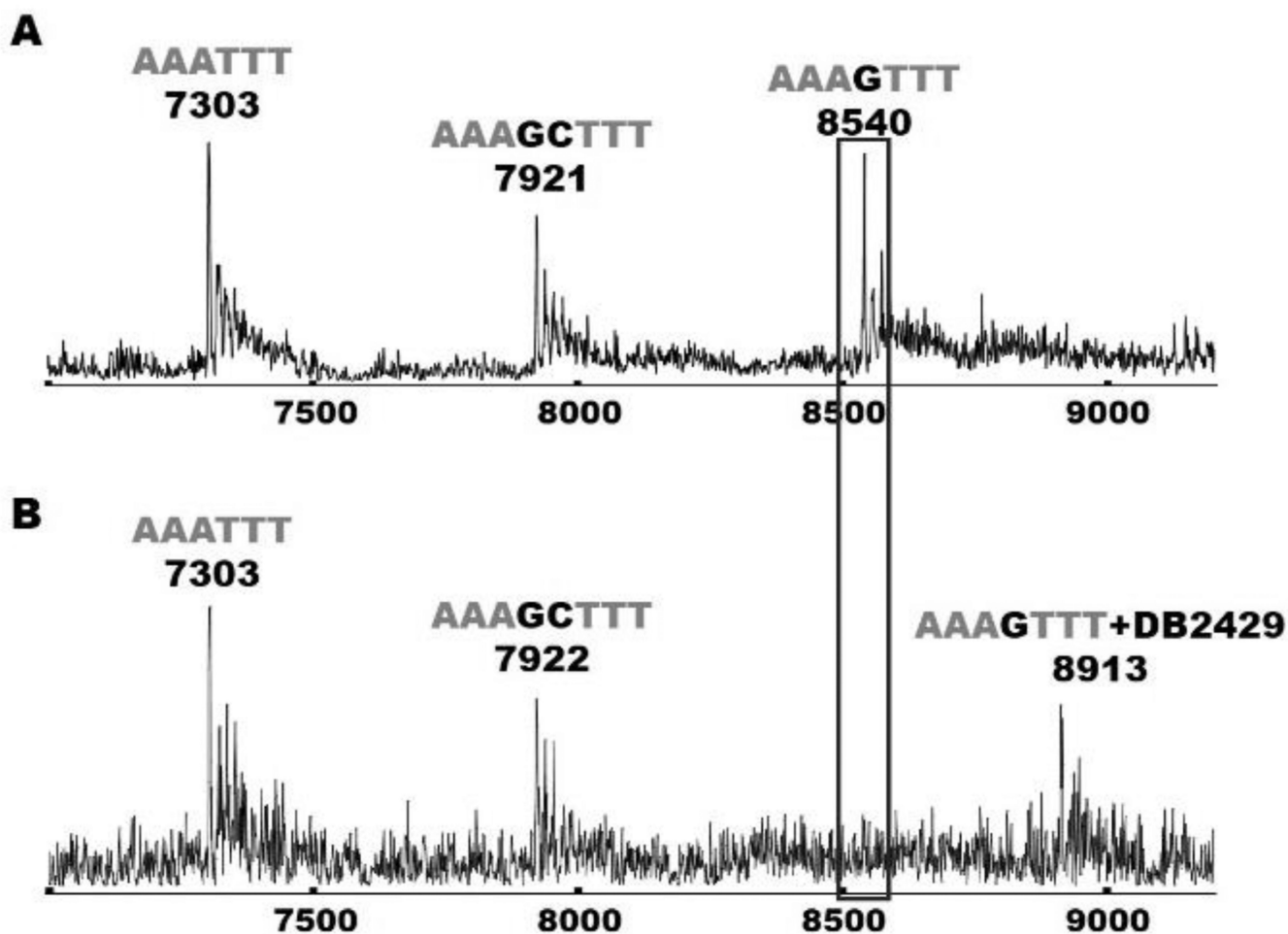
Figure 4.
ESI-MS negative mode spectra of the competition binding of sequences AAATTT, AAA**G**TTT and AAA**GC**TTT (10 μM each) with 40 μM DB2429 in ammonium acetate buffer (150 mM ammonium acetate with 5% methanol (v/v), pH 6.8). (A) The ESI-MS spectra of free DNA mixture. (B) The ESI-MS spectra of DNA mixture with DB2429. The ESI-MS results shown here are deconvoluted spectra and molecular weights are shown with each peak.
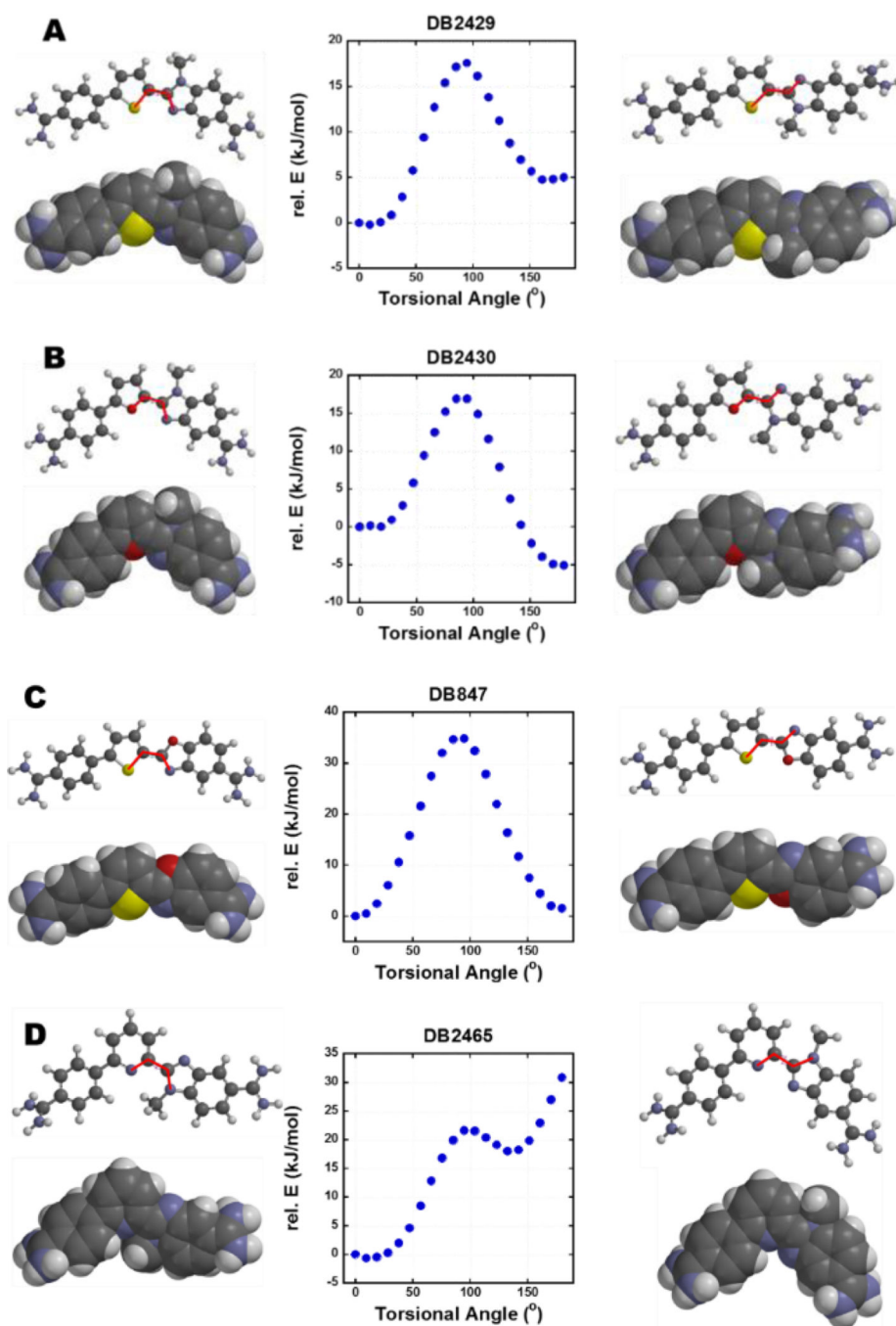
Figure 5.
Torsional angle maps for DB2429, DB2430, DB847, and DB2465; dihedral plots for (A) S-C of thiophene and C=N of *N*-MeBI. (B) O-C of furan and C=N of *N*-MeBI. (C) S-C of thiophene and C=N of benzoxazole. (D) N=C of pyridine and C=N of *N*-MeBI. All calculations are performed at the B3LYP/6–31G* level of theory. The scanned dihedral is depicted in red bold line.
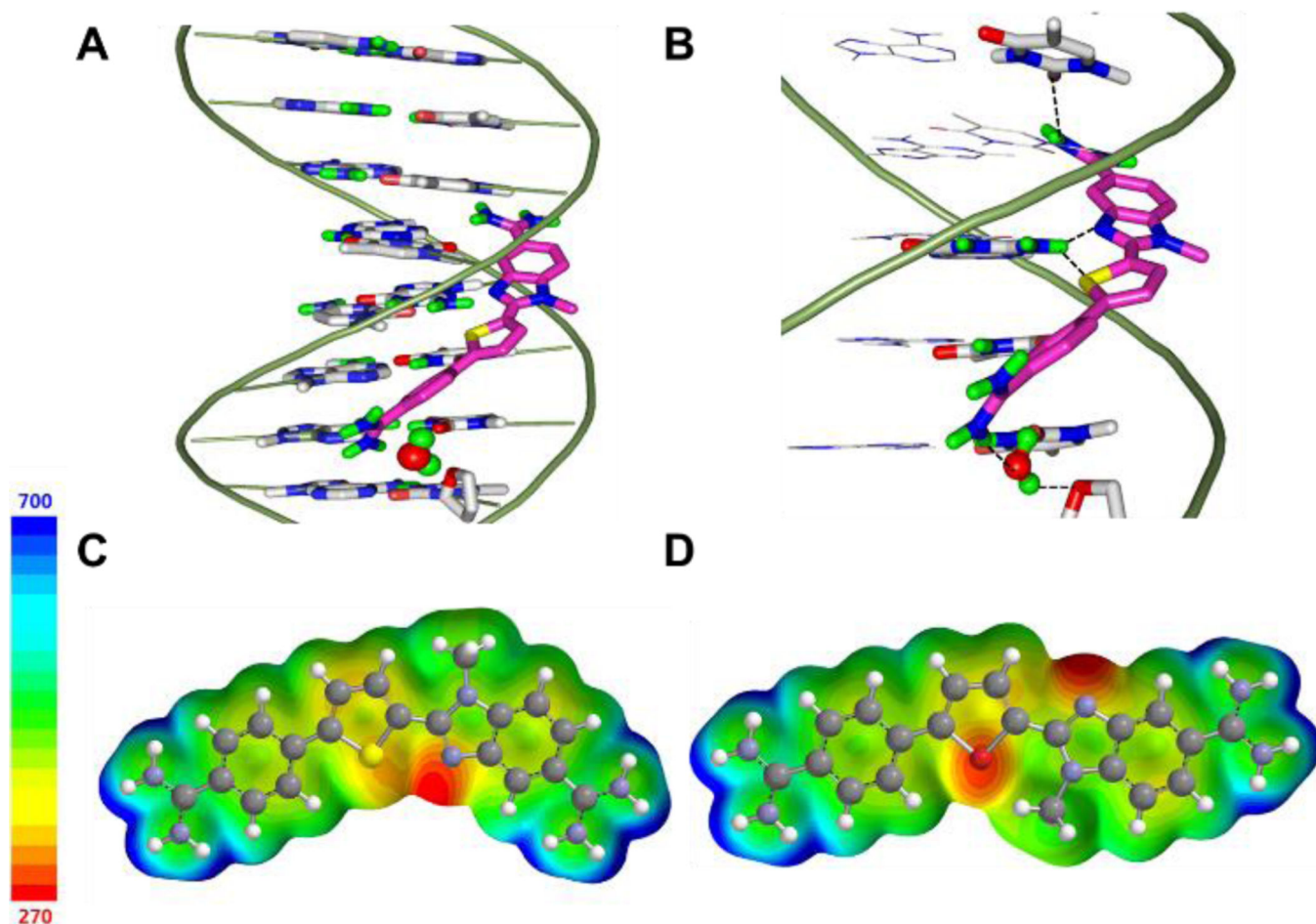
Figure 6.
Minor groove views of docked conformations of (A) DB2429 with the AAA**G**TTT sequence. The compounds are shown as stick model and colored by atom type (magenta for carbon, blue for nitrogen, yellow for thiophene sulfur, and green for amidine hydrogen). The DNA backbone is represented as a tube form in light green, DNA nucleobases are represented by sticks colored by atom type (gray for carbon, blue for nitrogen, red for oxygen, green for polar hydrogen). A water molecule (red sphere) is shown at the bottom amidine of each compound. (B) Important H-bond interactions between compounds and DNA nucleobases (shown with a black dashed line), the thiophene S and *N*-MeBI N form bifurcated hydrogen bonds with the exocyclic G6-NH in the minor groove, the bottom amidine group forms H-bonds with the thymine (T) carbonyl group through the water molecule, and the top amidine group forms a direct hydrogen bond with the thymine (T) carbonyl group. (C) (D) Electrostatic potential map of DB2429 and DB2430 with calculation performed at the B3LYP 6–31+G* level of theory (the energy range is negative to positive from red to blue color shown in a ladder).
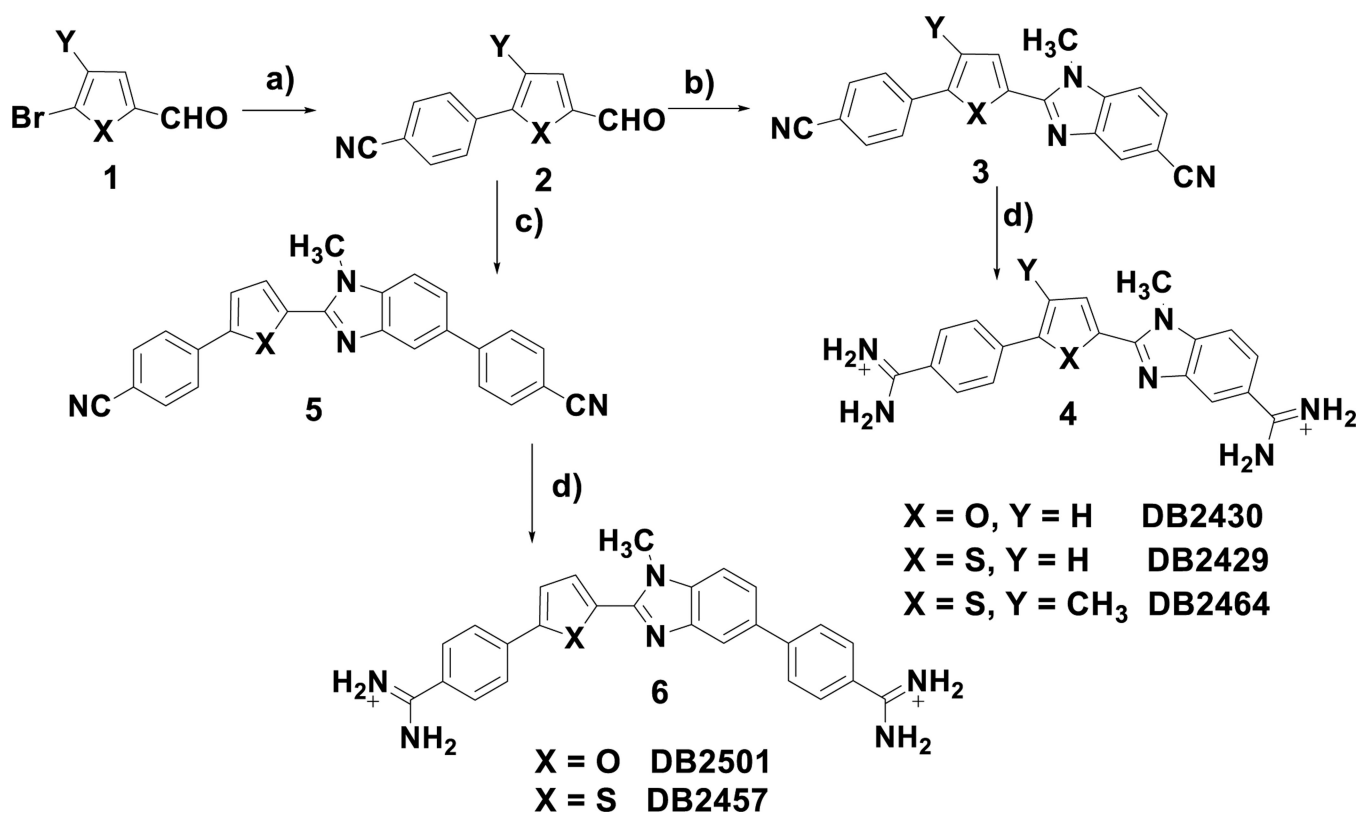
Scheme 1.
Reagents and conditions: a) 4-cyanophenylboronic acid, Pd(PPh$_3$)$_4$, Na$_2$CO$_3$/H$_2$O, dioxane, reflux b) 3-amino-4-methylaminobenzonitrile, sodium metabisulfite, DMF, reflux c) 3'-amino-4'-methylamino[1,1'-biphenyl]-4-carbonitrile, sodium metabisulfite, DMF,110 °C d) i. LiN(TMS)$_2$, THF, rt ii. HCl, ethanol.

Table 1

Thermal melting studies (ΔT_m, °C) of DB2429 and analogues with pure AT and mixed DNA sequences.[a]

DNA Ligand	AAA TTT	AAA G TTT	AAA GC TTT	DNA Ligand	AAA TTT	AAA G TTT	AAA GC TTT
DB2429	5	11	5	DB1300	4	8	3
DB2464	3	10	2	DB2325	2	8	<1
DB847	5	7	2	DB2457	6	12	5
DB2501	3	3	<1	DB2430	8	6	3
DB2432	2	3	<1	DB2465	5	4	3
DB1454	2	2	<1	DB2454	<1	<1	<1

[a] $\Delta T_m = T_m$ (the complex) − T_m (the free DNA). 3 µM DNA sequences were studied in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) with the ratio of 2:1 [ligand/DNA]. An average of two independent experiments with a reproducibility of ±0.5 °C. Full DNA sequences: AAATTT-5'-CCAAATTTGCCTCTGCAAATTTGG-3'; AAA**G**TTT-5'-CCAAAG**TTT**GCTCTCAAACTTTGG-3'; AAA**G**CTTT-5'-CCAAAG**C**TTTGCTCTCAAAGCTTTGG-3'.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Summary of binding affinity (K_D, nM) for the interaction of all test compounds with biotin labeled DNA sequences using biosensor-SPR method [a]

	DB 2429	DB 2464	DB 2432	DB 847	DB 1300	DB 2457	DB 2430	DB 2465
AAA TTT	524	1160	NB	731	277	222	74	330
AAA G TTT	50	70	303	559	196	4	180	690
AAA GC TTT	1086	1060	NB	NB	708	192	360	1400
AAA I TTT	322	883	--	--	--	322	--	--
ATA G TAT	432	590	--	--	--	297	--	--
AATT G AATT	983	1493	--	--	--	226	--	--

[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 µL/min flow rate. "--" experiment not done. "NB" no 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.