Published in final edited form as:
J Am Chem Soc. 2007 November 7; 129(44): . doi:10.1021/ja074560a.

Design of DNA Minor Groove Binding Diamidines that Recognize GC base pair Sequences: A Dimeric-Hinge Interaction Motif

Manoj Munde[§], **Mohamed A. Ismail**[§], **Reem Arafa**[§], **Paul Peixoto**[†], **Catharine J. Collar**[§], **Yang Liu**[§], **Laixing Hu**[§], **Marie-Hélène David-Cordonnier**[†], **Amélie Lansiaux**[†,‡], **Christian Bailly**[†,#], **David W. Boykin**[§], and **W. David Wilson**[§,*]

[§]Department of Chemistry, Georgia State University, P.O. Box 4098, Atlanta, Georgia 30302-4098, USA

[†]INSERM U-837, JPARC, Equipe N°4, IRCL, Lille 59045, France

[‡]Laboratoire de Pharmacologie Antitumorale du Centre Oscar Lambret, IRCL, Lille, France

Abstract

The classical model of DNA minor groove binding compounds is that they should have a crescent shape that closely fits the helical twist of the groove. Several compounds with relatively linear shape and large dihedral twist, however, have been found recently to bind strongly to the minor groove. These observations raise the question of how far the curvature requirement could be relaxed. As an initial step in experimental analysis of this question, a linear triphenyl diamidine, DB1111 and a series of nitrogen tricyclic analogues were prepared. The goal with the heterocycles is to design GC binding selectivity into heterocyclic compounds that can get into cells and exert biological effects. The compounds have a zero radius of curvature from amidine carbon to amidine carbon but a significant dihedral twist across the tricyclic and amidine-ring junctions. They would not be expected to bind well to the DNA minor groove by shape-matching criteria. Detailed DNaseI footprinting studies of the sequence specificity of this set of diamidines indicated that a pyrimidine heterocyclic derivative, DB1242, has remarkable binding specificity for a GC rich sequence, -GCTCG-. It binds to the GC sequence more strongly than to the usual AT recognition sequences for curved minor groove agents. Other similar derivatives did not exhibit the GC specificity. Biosensor-surface plasmon resonance and isothermal titration calorimetry experiments indicate that DB1242 binds to the GC sequence as a highly cooperative stacked dimer. Circular dichroism results indicate that the compound binds in the minor groove. Molecular modeling studies support a minor groove complex and provide an inter-compound and compound-DNA hydrogen bonding rational for the unusual GC binding specificity and the requirement for a pyrimidine heterocycle. This compound represents a new direction in development of DNA sequence specific agents and it is the first non-polyamide, synthetic compound to specifically recognize a DNA sequence with a majority of GC base pairs.

Introduction

The design of molecules that can recognize specific sequences and structures of nucleic acids is a research goal that is important both for understanding nucleic acid molecular recognition as well as for development of new therapeutics and reagents for biotechnology. It has been estimated, for example, that only a small percentage of cellular proteins are both

[*]Correspondence may be addressed to: WDW: Telephone: (404) 651-3903. Fax: (404) 651-2751. wdw@gsu.edu.
[#]Present address: Institut de Recherche Pierre Fabre, Toulouse, France
Contact information: W. David Wilson, Department of Chemistry, Georgia State University, P.O. Box 4098, Atlanta, GA 30302-4098, wdw@gsu.edu, Tel: 404-651-3903, Fax: 404-651-1416

"druggable" and disease modifying.[1, 2] This suggests that in order to develop novel and improved therapeutics, it will be useful to identify new, highly promising cellular receptors and design drugs which are selective for those receptors. Defining new nucleic acid targets is thus a very promising route for expanding useful drug design approaches. Diamidines have excellent transport properties into a variety of cells[3–6] and an orally available prodrug of the diamidine, DB75, furamidine (Figure 1), is currently in phase III clinical trials against trypanosomes, which cause sleeping sickness, as well as other microbial parasites.[4, 7–11] DB75 binds strongly in the minor groove of DNA and recognizes sequences of at least four AT base pairs.[7–9] For design of new agents that target additional disease organisms/cells as well as evading any possible resistance that could develop, we have focused on modifications of the structure, heterocycles, and properties of the basic units of the DB75 molecule. A key aim of the design considerations was to discover motifs that could expand the sequence recognition properties of diamidines by including a variety of nitrogen heterocycles into molecules of different shape.

DB921 (Figure 1) is a successful example of the modified-shape-design strategy.[12, 13] The compound, as well as CGP40215A[14](Figure 1), is more linear than DB75 and does not have the shape to match the curvature of the minor groove to allow both amidines to form hydrogen bonds with base pairs as DB75 does. Both compounds, however, incorporate a water molecule into the recognition complex with DNA. The specifically bound water completes the curvature of the compound in the complex and forms linking hydrogen bonds between the compounds and DNA base pair edges at the floor of the minor groove. DB921 is particularly successful in its interactions with the DNA minor groove in AT sequences and has a binding constant of greater than 10^8 M^{-1} under physiological conditions,[12] one of the highest binding constants observed for a molecule of this size. Incorporation of water into the complex would generally be expected to be unfavorable due to entropy cost. As Cooper and coworkers[15, 16] have shown, however, a water molecule in optimal arrangement in a DNA complex, can add an enthalpy component to the binding energy that is greater than the entropy loss. In terms of optimizing the compound-DNA interaction, the flexibility of the water unit to change position and orientation is another very favorable feature.

These results emphasize two very important design considerations: the shapes of compounds do not necessarily have to closely match the curvature of the DNA minor groove for very strong sequence-specific binding and nitrogen heterocycles are very useful recognition units when properly positioned in minor groove binding diamidines.[17] Previous design criteria focused strongly on compound curvature and ability to match the curvature of the minor groove.[18–20] The new design concepts have relaxed this curvature requirement and have resulted in new series of diamidines that radically depart from classical compounds such as netropsin, Hoechst 33258 and DB75. As part of this new design effort, DB1111 and the series of nitrogen tricyclic analogues shown in Figure 1 were prepared.

As can be seen, the compounds in Figure 1 have a zero radius of curvature for amidine carbon to amidine carbon, and would not be expected to significantly bind to the DNA minor groove by classical shape-matching criteria. The compounds are also triphenyl analogues and generally have significant dihedral twist across the tricyclic six member ring junctions as well as at the amidine ring connection.[21–23] Unlike the relatively linear DB921 and CGP, however, Tm studies suggest that DB1111 binds to AT sequences of DNA more weakly than DB75, and the nitrogen derivatives of DB1111 bind even more weakly.[24] For the design of GC specific recognition molecule, however, weak bound to AT is the desired result. We have now conducted detailed studies of the sequence specificity of this new set of designed diamidines with the exciting discovery that DB1242 (Figure 1) has remarkable binding specificity for a -GCTCG-sequence. It actually binds to the GC sequence more strongly than the usual AT recognition sequences for heterocyclic diamidines such as DB75.

This compound thus represents a new direction in development of DNA sequence specific agents, particularly for cell permeable compounds with therapeutic promise. It is the first non-polyamide synthetic compound to specifically recognize GC rich DNA segment. In addition to DNase I footprinting a variety of powerful methods to characterize the recognition of AT and GC sequences by the linear tricyclic derivatives of Figure 1 were used and the results are reported here.

Experimental Section

Compounds, Buffers and Solutions

The compounds of Figure1 were synthesized as previously described.[24] Their purity was verified by NMR and elemental analysis. Concentrated stock solutions, 1–2 mM, were prepared in water. Solutions of the compounds for biosensor-surface plasmon resonance (SPR), calorimetric and spectroscopic studies were prepared by dilution with 0.01 M cacodylic buffer, pH 6.25 with 0.001 M EDTA and 0.1 M NaCl. SPR binding, studies were conducted with 5′-biotinated DNAs, while calorimetric and spectroscopic studies were performed with non-biotinated DNAs (Figure 1). The concentration of the DNA solutions was determined spectrophotometrically at 260 nm using extinction coefficients per nucleotide of 9082 and 8996 M^{-1} cm^{-1} for -GCTCG-(5′-CAGCTCGAGTTTTCTCGAGCTG-3′),-AATT-(5-CGCAATTGGCTTTTGCCAATTGCG-3′) hairpin DNA respectively. The extinction coefficients were calculated on a per strand basis by the nearest-neighbor method and divided by the number of nucleotides per strand.[25]

Purification and Radiolabeling of DNA Restriction Fragments and DNase I Footprinting

The pBS plasmid was isolated and purified from E. coli using Qiagen columns. The 265 bp DNA fragment was prepared by 3′-[^{32}P]-end labeling of the EcoRI-PvuII double digest of the pBS plasmid (Stratagene) using α-[^{32}P]-dATP and AMV reverse transcriptase. The products were separated on a 6% polyacrylamide gel under non-denaturing conditions in TBE buffer (89 mM Tris-borate pH 8.3, 1 mM EDTA). After autoradiography, the requisite band of DNA was excised, crushed and soaked in water overnight at 37°C. This suspension was filtered through a Millipore 0.22 mm filter and the DNA was precipitated with ethanol. Following washing with 70% ethanol and vacuum drying of the precipitate, the labeled DNA was resuspended in 10 mM Tris adjusted to pH 7.0 containing 10 mM NaCl. DNase I footprinting experiments were performed essentially as previously described.[26–28] Briefly, reactions were conducted in a total volume of 10 μl. Samples (3 μl) of the labeled DNA fragments were incubated with 5 μl of compound solution for 30 min of incubation. Digestion was initiated by the addition of 2 μl of a DNase I solution whose concentration was adjusted to yield a final enzyme concentration of ~0.01 U/ml in the reaction mixture. After 3 min, the reaction was stopped by freeze-drying. Samples were lyophilized and resuspended in 5 ul of an 80% formamide solution containing tracking dyes. The DNA samples were then heated at 90°C for 4 min and chilled in ice for 4 min prior to electrophoresis under denaturing conditions on a 0.3 mm thick, 8% polyacrylamide gel containing 8 M urea at 60 W in TBE buffer, BRL sequencer model S2). Gels were then soaked in 10% acetic acid, transferred to Whatman 3MM paper and dried under vacuum at 80 °C, to then be exposed on a phosphorimager screen. Gels were analyzed with a Molecular Dynamics 425E PhosphorImager and densitometric measurements were made by using ImageQuant™ software. The densitometric plots represent the differential cleavage at each bond relative to that in the control, express as an *ln* function. The position of each bases is deduced from the guanine lane (G-track). Logarithmic positive values indicate enhanced cleavage whereas negative values indicate blockage.

SPR-Biosensor Binding Determinations

SPR measurements were performed with a four-channel BIAcore 2000 optical biosensor system (BIAcore Inc.). 5′-biotin labeled DNA samples (Figure 1) were immobilized onto streptavidin-coated sensor chips (BIAcore SA) as previously described.[29] Three flow cells were used to immobilize the DNA oligomer samples, while a fourth cell was left blank as a control. The SPR experiments were performed at 25°C in filtered, degassed, 10 mM cacodylic acid buffer (pH 6.25) containing 100 mM NaCl, 1mM EDTA. Steady state binding analysis was performed with multiple injections of different compound concentrations over the immobilized DNA surface at a flow rate of 25ul/min and 25°C. Solutions of known ligand concentration were injected through the flow cells until a constant steady-state response was obtained. Compound solution flow was then replaced by buffer flow resulting in dissociation of the complex. The reference response from the blank cell was subtracted from the response in each cell containing DNA to give a signal (RU, response units) that is directly proportional to the amount of bound compound. The predicted maximum response per bound compound in the steady-state region (RU_{max}) was determined from the DNA molecular weight, the amount of DNA on the flow cell, the compound molecular weight, and the refractive index gradient ratio of the compound and DNA, as previously described.[30] The number of binding sites and the equilibrium constant were obtained from fitting plots of RU versus C_{free}. Binding results from the SPR experiments were fit with either a single site model ($K_2 = 0$) or with a two site model:

$$r=(K_1 * C_{free}+2 * K_1 * K_2 * {C_{free}}^2)/(1+K_1 * C_{free}+ * K_1 * K_2 * {C_{free}}^2) \quad (1)$$

where r represents the moles of bound compound per mole of DNA hairpin duplex, K_1 and K_2 are macroscopic binding constants, and C_{free} is the free compound concentration in equilibrium with the complex.

Isothermal Titration Calorimetry

Calorimetric titrations were performed with a VP-ITC (Microcal, Inc., Northampton, MA). Software provided with the calorimeters is used for control and data collection. ITC experiments were conducted by injecting 10 μl of the ligand in cacodylate buffer every 300 s for a total of 29 injections into a DNA hairpin solution in the same buffer. The compound concentration was 0.2 mM for all experiments and DNA concentrations were 0.012 mM for the -GCTCG- and -AATT- hairpins. Similar experiments were performed to determine the heats of dilution of the ligand with buffer. The heat produced for each injection of compound into DNA or buffer was obtained by integration of the area under each peak of the titration plots with respect to time. The heats of reaction were obtained by subtraction of the integrated heats of dilution of the compounds from the heats corresponding to the injection of compound into DNA.

Averaged subtraction was applied to ITC titration data. Data corresponding to the first injection were discarded. The binding enthalpy (ΔH) for each titration was obtained by fitting the results of heat per mole as a function of total molar ratio (ligand/DNA) as described below.

CD Spectroscopy

A 1 cm path length cell was used and all experiments were done at 25°C. Specific aliquots of the (5′-CAGCTCGAGTTTTCTCGAGCTG-3′) and (5′CGCAATTGGCTTTTGCCAATTGCG-3′) hairpin duplexes (3×10^{-6} M per hairpin duplex), were titrated with increasing concentrations of compound. The resulting ratios were between 0.05 and 3.5 (mol compound to mol DNA duplex). The experiments were performed in cacodylic acid buffer. The sensitivity was set at 1 mdeg and the scan speed was

set at 50 nm/min. Four scans were accumulated and averaged by the computer for each titration point.

Molecular Calculations, Modeling and Docking Studies

Geometry optimized structures for DB1242 were calculated at the Hartree-Fock 631G** level with the Spartan '04 software package.[31] Different stacking arrangements of two optimized DB1242 molecules were then visually evaluated for intermolecular amidine-pyrimidine nitrogen hydrogen bonds. An optimum arrangement was found and used in DNA docking studies.

Docking studies were performed with the SYBYL 7.2 software package[32] on a Fedora Core 5 Linux Workstation. A DNA duplex, d(CCAAGCTCGAAGC)•d(GCTTCGAGCTTGG) with the recognition site -GCTCG- was constructed in the Biopolymer module. After visual evaluation of DNA-stacked dimer crystal structures, protein data bank (PDB) crystal structure 1CYZ, a d(GAACTGGTTC)•d(GAACCAGTTC) tri-imidazole polyamide complex, was selected for DNA and DB1242 ligand alignment based on the location of common bases in the binding site and the size of the tri-imidazole dimer. For preliminary docking studies, the rigid DNA designed through the implementation of the Biopolymer module was minimized for 100 iterations using the Tripos force field. This process altered the rigid DNA by slightly enlarging the minor groove. The designed DNA was then aligned to the DNA of PDB crystal structure 1CYZ. DB1242 was constructed and individual atoms were assigned Gasteiger-Marsili charges. [33] The molecule was then minimized using the Tripos force field until a terminating conjugate gradient of 0.01 kcal/mol Å was reached.[34, 35]

By using the optimum stacking arrangement found for the stacked dimer in Spartan, DB1242 molecules were aligned to the tri-imidazole dimer of 1CYZ. After all alignment was complete 1CYZ was deleted leaving the designed DNA in complex with a DB1242 dimer. The DB1242 dimer was moved into a second memory location, so that the dimer could move independently of the DNA. For each docking, the genetic algorithm of the Flexidock module was employed implementing 5 different random numbers and large number of generations. The suggested minimum amount of generations to be used in Flexidock studies should be the number of rotatable bonds plus six times 500; this started the study out with 98 000 generations and still converging low energy compounds.[32] To insure that the best low energy complexes were obtained 570 000 generations, 3000×190 rotatable bonds, were used for each docking. This insured that low energy compounds would be obtained. Both the DNA and DB1242 were permitted torsional flexibility in the docking process. Atomic charges were calculated using the Kollman All-Atom protocol for the DNA and all of the hydrogen bond sites were marked for the DNA and DB1242. Each docking generated 20 low energy structures; thus, a total of 100 structures, 5×20, were obtained and observed.

DB1242 was then used as a template for DB1111 and DB1164. Minimization and Flexidock were implemented as before, using the same DNA strand with binding site -GCTCG-. Therefore, 100 structures were acquired for DB1111 and 400 structures were gained for DB1164, since DB1164 dimer can be bound in four different orientations. All of the obtained dimer to DNA complexes were observed. This Flexidock module docking process generated similar structures for all of the random numbers used; thus, the main differences observed are found within the interactions, including the hydrogen bonds, of the dimer to DNA complexes as a result of the ligands bound.[36, 37]

Results

DNaseI Footprinting: Identify the Binding Sites

DNaseI footprinting is the method of choice for combinatorial evaluation of DNA binding specificity and binding site size in long DNA sequences that contains a large number of different sequence and site length possibilities.[27] Results for linear compounds from Figure 1 are shown in Figure 2 with an experimental gel and densitometer scans. The results are compared to the well characterized minor groove binding diamidine, furamidine, DB75, at 1μM concentration. It is clear that all compounds give several strong footprints in similar positions with this DNA sequence, for example, between base positions 70–80, but other footprints with greatly different intensity are also observed. A densitometer scan of the segments provides a comparison of footprinting results for DB75 and DB1242 (Figure 2B). Moving from the 5′ position (left side of the top scan) in the 3′ direction, a strong footprint is seen for DB75 but not for DB1242 at the -AATT- site at position 140. This is the expected result for groove binding compounds such as DB75,[27, 38] a strong footprint at AT sequences of four or more base pairs (white boxes). As expected, the linear compound, DB1242, does not bind to the minor groove in the usual manner. At 5 μM (but not at 1μM), DB1242 slightly perturbs the extent of cleavage by DNase I at the two AATT sites located near positions 125 and 140 (and this is in agreement with SPR binding data, see below) but these are not true footprints *per se* and the extent of DNase I cleavage inhibition is considerably higher at the GC-containing site at position 85. Continuing to scan the sequence in the 3′ direction, footprints are seen for DB75 at the -AATTT- (position 121–125) and -ATTA- (position 92–95) sites while DB1242 does not footprint particularly well at either sequence. Very surprisingly, DB1242 exhibits a very strong footprint at the -GCTCG- sequence between positions 80–90, while DB75 shows no footprint at this site. The same -GCTCG- sequence is also protected from DNase I digestion using DB1242 but not DB75 at position 155–159, but the resolution at that position was not good enough to properly quantify it. Both compounds have footprints at the long, -AAATTAA- sequence between positions 70–80.

Footprinting results for three other linear compounds from Figure 1, at a 5μM concentration, are compared to the 1μM results for DB75 in the bottom densitometer trace in Figure 2. As can be seen, all of these compounds behave much more like DB75 and other classical minor groove binding compounds with footprints at AT sites at the 5μM concentration. None of these compounds have a footprint at the -GCTCG- site, however, as observed with DB1242. Interestingly, these compounds appear to footprint better, relative to DB75, at sites with a -TA- base pair step in the sequence. For example, stronger footprints are seen for the linear compounds at the -ATAA- site near position 130, at the ATTA site between 90–100 and at the TAAA site between 60–70 (Figure 2C). DB1228, DB1164 and DB111 are clearly AT-selective but do not recognize GC-containing sites as does DB1242. It has been proposed that -TA- steps widen the groove in AT base pair sequences and generally result in weaker binding of classical binding agents such as DB75, netropsin and Hoechst33258.[39–41] The linear compounds give a modest but unexpected footprint at the -AAAC- sequence near position 110. This site does not have the generally required four AT base pairs for minor groove interactions, but it should also have a wider groove and, with three AT base pairs, can apparently bind the linear diamidines reasonably well. In conclusion, DB1242 gives a unique site of protection from DNaseI digestion of DNA at a -GCTCG- sequence that is not seen with other minor groove binding agents, even those with very similar structure.

Biosensor - SPR: Affinity, Stoichiometry, Cooperativity

The biosensor-SPR method can provide essential information on the kinetics, affinity, stoichiometry, and cooperativity of DNA interactions,[29] even in the complex binding

reactions observed for the linear compounds (Figure 1) with -GCTCG-. The method enables high resolution analysis of interesting binding sites discovered by the DNase I footprinting method. The sequences of the DNAs used in the SPR experiments are based on the results from footprinting (Figure 2) and are shown in Figure 1. As can be seen from the sensorgrams for the interaction of DB1242 with the -GCTCG- hairpin duplex (Figure 3A), the kinetics for association and dissociation of the compound with the DNA site are too fast for analysis at the concentrations where DB1242 binds to DNA at 25 °C and 0.1M NaCl. From the steady-state plateau over the same concentration range, much less binding is observed with the -AATT- hairpin sequence (Figure 3B) in agreement with the lack of an observed DNaseI footprint by DB1242 at -AATT-.

To evaluate the affinity, stoichiometry and cooperativity for the interaction of DB1242 with the -GCTCG- site, the RU values at each concentration were determined in the steady-state region, where the on and off rates are equal, and are plotted versus the C_f values for DB1242, the concentration of the compound in each flow solution (Figure 4A). The results of three separate experiments are included in Figure 4A to illustrate the reproducibility of the method. The fitting results in the figure provide several key pieces of information about the interaction of DB1242 with DNA. First, the predicted RU value per bound compound, based on the amount of DNA on the chip is 33 RU and the observed value in the two site model used to fit the results in Figure 3B is 34 (68 RU total predicted at saturation) in excellent agreement with a two site interaction model. Second, the equilibrium constant for binding of the first molecule of DB1242 to the DNA is over a factor of 500 less than for binding of the second molecule: $K_1 = 2.0 \times 10^4\ M^{-1}$, and $K_2 = 9.1 \times 10^6\ M^{-1}$ (Table 1). This is indicative of an interaction with very strong positive cooperativity, as can be seen by the characteristic curvature for positive cooperativity in the plot (Figure 4A).[28, 41, 42] The strong, cooperative binding of DB1242 creates a dimer structure that can block DNaseI digestion and can account for the unexpected interaction and strong footprint of DB1242 in the -GCTCG- sequence. It should be noted that results for all of the compounds in Figure 1 could only be obtained in SPR experiments up to approximately 4 μM due to erratic results for subtraction of the blank flow cell signal above the 4 μM concentration range. Such results are observed for many organic cations at high concentration and are probably due to non-uniform, kinetically-controlled association of such compounds with flow cell surfaces at high concentration.[29] More limited studies were done with DB1242 binding to -GCTCG- at lower salt concentration, 50 instead of 100 mM, and the K values increased as expected for a cation-DNA interaction: $K1 = 6.4 \times 10^4\ M^{-1}$, and $K2 = 10.6 \times 10^6\ M^{-1}$ (Table 1).

The interaction of DB1242 with the -AATT- sequence (Figure 1) is quite different than observed with -GCTCG-. The affinity for DB1242 binding to -AATT- is much weaker than for binding to -GCTCG- and the kinetics of association and dissociation are also too fast for analysis. The best fit to the results is with a one site model and gives an equilibrium constant of $K = 3.1 \times 10^5\ M^{-1}$ (Figure 4B; Table 1). The lower value of K is typical of weak and non-specific interactions. This weak binding to the -AATT- site is in agreement with the lack of a footprint by DB1242 at this sequence (Figure 2), but is unlike results from any previously observed diamidine footprinting experiment. In contrast to DB1242, DB1111 does give a footprint at the -AATT- sequence (Figure 2) and it displays stronger binding to the AT sequence in SPR results (Figure 4B). The association and dissociation kinetics for DB1111 at the -AATT- site are slow enough for fitting at low concentration. The best fit results for the steady-state RU values versus Cf are for a single site with an equilibrium constant of $K = 2.1 \times 10^6\ M^{-1}$ (Table 1). DB1164 with a single ring nitrogen has binding results very similar to DB1111. With the -GCTCG- sequence both DB1111 and 1164 bind much more weakly than DB1242 (Figure 4). Thus, both DB1111 and DB1164 bind better to the -AATT- sequence than DB1242. Both, however, have significantly weaker binding to the -GCTCG- sequence than DB1242 and their RU values increase almost linearly with time. The best fit

values are for K less than 2.0×10^5 M^{-1} (Table 1) and such results are in agreement with weak, nonspecific binding of DB1111 and DB1164 to the -GCTCG- sequence.

Netropsin is a well-known AT specific minor groove binding compound that gives very strong footprints at AT sequences of four or more AT base pairs and it was used as a control in the SPR studies. It binds very strongly to the -AATT- sequence in a 1:1 complex with K $=3.6\times10^7$ M^{-1}. At the same netropsin concentrations no significant binding to the -GCTCG- sequence is observed as expected for an AT specific binding agent (data not shown). In summary, the biosensor-SPR results for three separate experiments on all linear compounds of Figure 1 are in agreement with the footprinting experiment: strong, cooperative binding in a dimer complex for DB1242 at -GCTCG- but weak nonspecific binding with -AATT-. Opposite results are obtained in both SPR and footprinting experiments for DB1111 and DB1164 as well as for netropsin.

Isothermal Titration Calorimetry: Detailed Thermodynamics of Binding

ITC experiments with the -GCTCG- and -AATT- DNA hairpin DNAs (figure 1) were conducted to obtain a full thermodynamic comparison of the very different interactions of DB1111, DB1164 and DB1242 with the -GCTCG- and -AATT- sequences. Figure 5A shows the calorimetric results for titrating DB1242 into the -GCTCG- hairpin. A plot of heat versus molar ratio, after subtracting the heat of dilution for addition of the compound into buffer, is also shown in Figure 5A. As expected from the complex binding curve obtained in SPR experiments, the ITC titration of -GCTCG- is also complex. The heat/injection (Figure 5A) first decreases and then levels out between ratios of 0.5 – 1.0. Above 1.0, the heat/injection approaches zero as saturation of binding sites occurs. To fit these results we have used a sequential binding model with the K_1 and K_2 values fixed at the SPR values. Fixing the K values greatly increases the reliability of fitting complex ITC curves since only the binding enthalpy values must be determined by the fitting routine. The line in Figure 5A shows the best fit obtained by varying ΔH_1 and ΔH_2 and the results for all ITC experiments are in Table 1. The fit is quite acceptable given the constraints on and complexity of the model. Both ΔH values are exothermic and favorable for binding but $\Delta H_2 = -15.7$ kcal/mole is more favorable than $\Delta H_1 = -3.3$ kcal/mole in agreement with the higher binding constant for the second binding molecule. The ΔH values along with $\Delta G°$ values determined from SPR binding constants allow calculation of the binding entropies from $\Delta G = \Delta H - T\bullet\Delta S$ at 298 K (Table 1).

Unlike DB1242, both DB1111 and DB1164 (Figure 5C and 5E) show weak, nonspecific binding to the -GCTCG- DNA hairpin in ITC titration but strong binding affinity for the AT site sequence with 1:1 stoichiometry. ITC results for the two compounds with -AATT- (Figure 5D and 5F) indicate a strong binding primary site with a $\Delta H = -6.0$ kcal/mole (Table 1) and weaker secondary binding, as in SPR, that can not be accurately fit to obtain thermodynamic constants. The binding of DB1111 and DB1164 to -GCTCG- is essentially nonspecific, with low ΔH values. DB1164 binds to the -AATT- sequence slightly more strongly than DB1111 with a more favorable ΔH (Table 1). In summary, ITC results clearly show that DB1242 binds specifically as a 2:1 complex to -GCTCG- but shows very weak binding to AT sequences in agreement with footprinting and SPR results. DB1111 and DB1164, again as with SPR and footprinting results, bind much more weakly to -GCTCG- than to -AATT-. Clearly the central pyrimidine of DB1242 provides unusual and very favorable interaction ability for the -GCTCG-sequence.

CD Spectroscopy: Evaluation of the Binding Mode

CD spectroscopy was used to obtain information on the binding mode of the linear compounds. Positive induced signals in CD spectroscopy are generally obtained for

compounds that bind in the DNA minor groove and this pattern provides a method for evaluating solution binding modes.[43] DB1242 has an absorbance peak near 300nm that overlaps with the long wavelength region of the DNA CD spectrum. For this reason the free DNA spectrum was subtracted from that for the compound-DNA complex, but a spectrum for free DNA is included with the difference CD spectra for reference in Figure 6. As expected for a minor groove complex, there is a strong positive induced CD spectrum seen near 300 nm on binding of DB1242 to -GCTCG-. In agreement with a 2:1 binding mode, the CD signal increases up to a ratio of 2. Above a ratio of 2, the change in CD signals on addition of DB1242 decreases to zero. Because of the concentrations used in the CD experiments, essentially all of the DB1242 added at low ratios of compound to DNA is bound and cooperativity is not detected. The results with the -AATT- sequence are quite different with only small induced CD signals on addition of DB1242 to the DNA.

With DB1111 and DB1164 (Figure 1) the induced CD signals for complex formation with -GCTCG- are much smaller than observed for DB1242 (Figure 6). Interestingly, all of the linear compounds give induced CD signals with the -AATT- sequence that are smaller than for classically curved minor groove binding compounds such as DB75.[59] This has been observed previously with other linear minor groove binding agents[12, 14] and appears to be a characteristic of such compounds with AT sequences. The strong positive induced CD signals with DB1242 and -GCTCG- show that the weak CD signals are not characteristic of all sequences and are in agreement with the impressive binding of DB1242 with the -GCTCG- sequence as observed by other methods.

Molecular Docking Studies

To help understand the experimentally observed differences in interaction of the compounds of Figure 1 with the -GCTCG- sequence, a modeling investigation of potential stacked dimer structures that could interact favorably with the DNA minor groove was carried out. The molecular geometries of DB1242, DB1164 and DB1111 are significantly different. It is well known that the two phenyls of biphenyl are twisted with fslightly over 40° in the gas phase and slightly under 40° in condensed phases[22, 44] due to steric repulsion of the two pairs of ortho phenyl hydrogens at the bond connecting the two phenyls. We have shown in a previous crystal structure that a phenyl-pyridine type junction, such as that in DB1164, has a twist angle between 10°–20° in the solid state.[45] The lower torsional angle is presumably due to single proton pair repulsion as versus two similar repulsions in biphenyl type systems. Based on extension of these results, the angle for the phenyl-pyrimidine system of DB1242, with no proton pair repulsion, would be expected to be near 0° and Hartree-Fock calculations support this hypothesis. The calculations predict torsional angles of 44–45° for the biphenyls of DB1111, near 30° for the phenyl-pyridine in DB1164, and 0° for the phenyl-pyrimidine in DB1242. In all calculations the phenyl-amidine twist is near 40° (Figure 7A). Even with a reasonable error in the predicted angles, these results indicate that the phenyl-pyrimidine region of DB1242 provides a region of planar surface, with relatively little energy cost, that can slide into the minor groove with little to no steric repulsion. Such a surface is not present in the other linear compounds of Figure 1 and the energy cost to obtain a planar stacking surface would be substantially increased for the more highly twisted compounds.

The next step in the model analysis was to construct stacked dimer structures. Three qualitative criteria were used to form the stacked dimers: (i) the stacked structure should have an approximately concave type shape to match the minor groove; (ii) the amidine groups should be separated as far as possible to minimize electrostatic repulsion among the four charges and (iii) hydrogen bonds should be formed with amidines and heterocyclic nitrogens where possible. Features that can explain the unique experimental binding differences with -GCTCG- were readily incorporated into a stacked dimer model for

DB1242 (Figure 7B). The two amidines on the inner face and ends of the stacked DB1242 dimer are in an optimum position to interact with bases at the floor of the minor groove. The amidines on the outside, convex, face of the dimer are in a position and have an appropriate dihedral angle to form hydrogen bonds with the pyrimidine nitrogens on the outside face of the other stacked molecule of DB1242, to give two intermolecular amidine-pyrimidine hydrogen bonds within the dimer (Figure 7B). As can be seen from Figure 7A, the pyrimidine nitrogens have a negative electrostatic potential that interacts favorably with the positive potential of the amidines. The pyrimidine nitrogens of the dimer that are on the concave face are in position to accept hydrogen bonds from the G-NH_2 groups in the minor groove.

To determine whether the DB1242 dimer model could interact favorably with the minor groove in the -GCTCG- sequence, docking studies were carried out. The docking studies were guided by the experimental results and the goal was to determine whether reasonable DB1242 dimer-DNA structures could be constructed. The docking procedure (Sybyl-Flexidock) and starting DNA structure are described in the Methods Section. Possible binding conformations of sterically acceptable complexes for observation of intermolecular molecule to molecule and molecule to DNA interactions were obtained from the docking experiments. The docked structures converged on a set of similar low energy conformations for the DB1242-DNA complex and an example is shown in Figure 7C. As in the original proposed dimer structure (Figure 7B), two hydrogen bonds, amidine–NH to pyrimidine nitrogen, can be seen within the conformation of the DB1242 dimer. Recognition of the minor groove of the -GCTCG- sequence by the dimer is through four hydrogen bonds, two in the upper molecule (amidine-DNA) and two in the lower molecule (one amidine-DNA and a pyrimidine nitrogen-DNA, Figure 7C). The dimer appears to have some conformational freedom that gives the stacked DB1242 dimer complex stability by allowing strong hydrogen bonding to orient the docked structure. These hydrogen bonds help explain the experimental binding constant and large negative binding enthalpy for DB1242 with the -GCTCG- sequence.

To further analyze the correlation between docked structures and binding constants, DB1111 and DB1164 (Figure 1) were docked into the same DNA minor groove -GCTCG- sequence (Figure 7D). When docked, DB1111 (Figure 7D) is observed to have predominantly amidine to base edge interactions in the minor groove of DNA. The two DB1111 compounds in a stacked dimer cannot form intermolecular hydrogen bonds. The lack of molecule to molecule interaction provides an understanding of the low experimental binding constant for DB1111. DB1164 can bind in four different orientations depending on the locations of the pyrimidine nitrogen. When a pyrimidine nitrogen of DB1164 points out, it can hydrogen bond to an amidine on the other molecule of the DB1164 dimer. When the nitrogen points in, it can accept a hydrogen bond for a G-NH_2 group. In all DB1164 dimer dockings two hydrogen bonds were formed, amidine–NH to pyrimidine nitrogen. Each docked conformation appears possible, thus, this DB1164 dimer is most likely not bound into one specific orientation. There are fewer hydrogen bonds than with the DB1242 complex and they appear to be weaker interactions than those formed in the DB1242 complex.

DISCUSSION

The dicationic polyamide, netropsin, diamidines, pentamidine and berenil, were identified as DNA minor groove binding agents in some of the earliest studies of unfused aromatic cation-DNA complexes, and they have served as model compounds in many studies on nucleic acid interactions.[10, 19, 46–49] The fit of the compounds into the minor groove was visualized in molecular detail in the x-ray structure by Dickerson and coworkers of netropsin bound to the same-AATT- sequence of the self-complementary duplex, d(CGCG-AATT-

$CGCG)_2$.[49] Neidle and coworkers have provided highly informative x-ray structures for pentamidine, berenil, DB75 and a number of other diamidines bound at the -AATT- site. All of these compounds have a concave shape that allows them to match the curvature of the helical minor groove and slide deeply into the groove. Hydrogen bond donating groups on the concave face of these compounds contact hydrogen bond acceptors on the A and T base edges that are exposed at the floor of the groove in the -AATT- site.[9] The minor groove width in AT sequences can narrow to the width of the unfused aromatic groups in the compounds without a large energy penalty[41, 50] and the compounds make extensive energetically favorable contacts with the walls of the minor groove. The positive charges on the amidines or other groups provide electrostatic contributions to the complex energetics through phosphate interactions. The discovery of many similar AT specific minor groove binding agents led to a model of the key requirements for minor groove compound design: a concave shape to match the groove structure; hydrogen bond donating groups that are placed at inner-face compound positions to index appropriately with acceptors on base edges at the floor of the groove; and positive charges for solubility and DNA phosphate interactions.[51] Displacement of bound water from the minor groove in AT sequences provides a general favorable contribution to the binding entropy.[49, 52–54]

The netropsin structure led the Lown[48, 55] and Dickerson[49] groups to propose an extension of the model in which GC base pair specificity could be designed into minor groove binding polyamides by replacing some of the hydrogen bond donating groups on the concave face of the compounds with hydrogen bond acceptors. Such acceptors would be sterically able to accommodate the extra size of the G-NH_2 group, which hydrogen bonds with C in the minor groove, and could form an extra hydrogen bond with the –NH_2 group. Because DNA sequences with GC or mixed GC and AT base pairs do not adopt a narrow minor groove as easily as for pure AT sequences, the initial compounds did not have a good fit to minor groove and they had limited success. The discovery of a stacked dimer structure for a complex of distamycin in the minor groove by Pelton and Wemmer,[56, 57] however, suggested a way around this difficulty by use of dimers to fit into a widened minor groove. This lead to successful construction of polyamides, based on distamycin and netropsin, by Lown, Dervan, Lee and others,[11, 48, 55, 58] which could bind to the minor groove in GC and mixed sequences as a stacked dimer. In principle the dimer concept could be applied to other minor groove binding agents, such as pentamidine, berenil, Hoechst 33258, however, a general construction model for dimer complexes of such compounds or analogs has not been developed.

Although the dimer design concept with polyamides has worked well, the goal of designing clinically useful agents based on polyamides has not been realized. Minor groove targeting diamidines, however, have provided clinically useful agents against several microorganism-caused diseases, and an orally effective prodrug of DB75 (Figure 1) is in phase III clinical trials against trypanosome caused sleeping sickness.[7, 8] It seems certain that with their low toxicity and excellent cell uptake, these compounds could yield agents with broader therapeutic applications if methods of targeting additional DNA sequences could be developed. To help discover additional DNA molecular recognition mechanisms through both monomer and dimer complexes of diamidines, we have taken a two-fold experimental discovery approach: (i) vary the concave shape of the diamidines and (ii) include a variety of hydrogen bond acceptors and heterocyclic units in the basic compound structure. The initial evaluation of the new agents is by DNaseI footprinting, which allows combinatorial-type analysis of recognition at sequences of different lengths.[27] It provides a method to discover compounds with new and unusual binding specificity that do not obey the classical model.

The linear compounds of Figure 1 were designed to completely remove the concave shape required in the classical model for minor groove binding. DB1111 has a triphenyl aromatic

system while the other linear compounds have one (DB1164) or two nitrogen hydrogen bond acceptors in a central heterocycle. The predictions for these compounds were that they would bind significantly more weakly to AT sequences than the classical minor groove agents but they would have the capability to recognize additional sequences through the nitrogen heterocycles. DNaseI footprinting analysis confirmed the reduced binding of the linear compounds to AT base pairs but did indeed produce a very exciting result for DB1242 (Figure 2). The strong footprint at -GCTCG- for DB1242 is a unique GC-rich DNA recognition sequence with only one AT base pair, in contrast to the usual required four ATs with no GC base pair for the diamidines. The recognition is surprisingly specific for the DB1242 heterocycle and none of the other linear compounds show a detectible footprint at -GCTCG- under the experimental conditions. The differential cleavage plots in Figure 2 illustrate clearly that (i) all of the linear compounds and DB75 (Figure 1) footprint at AT sites of four or more base pairs; (ii) the footprints for DB1242 are quite weak at most AT sequences and are significantly weaker than for the footprint at -GCTCG-; (iii) all of the compounds footprint more weakly than DB75 in AT sequences as would be expected from their linear shape; (iv) on a relative scale the linear compounds footprint slightly better than expected at sites with a TA step which might indicate that they prefer a wider groove than classical minor groove binding agents. The requirement of the wider groove for these compounds agrees with their highly twisted conformation (below and Figure 7).

Several important conclusions can be derived from the quantitative analysis of DB1242 binding to the -GCTCG- sequence by biosensor-SPR methods. It is clear that two molecules bind to each-GCTCG- site and that the binding is highly cooperative with K2 over 500 times larger than K1. Binding of DB1242 to the -AATT- site is quite weak, as predicted from the footprinting results, and is in the same affinity range as nonspecific binding (Table 1). This is in strong contrast to other DNA minor groove binding diamidines. DB1111 and DB1164 were also evaluated by SPR methods and the results are reversed from DB1242. Both compounds bind at the nonspecific affinity level to the -GCTCG- sequence but bind more strongly to -AATT-. It is clear that the pyrimidine heterocycle of DB1242, when stacked in a 2:1 complex, provides a unique interaction at -GCTCG-. ITC results confirm the 2:1 binding of DB1242 to the GC sequence and show that formation of the 2:1 complex is enthalpy driven, suggesting that intermolecular interactions such as hydrogen bonding are dominant in complex formation. Binding of DB75 and related concave diamidines to AT sites in a 1:1 complex, in contrast, is entropy driven.[52] Thus, both the interaction sequence and the thermodynamic driving force are completely different for DB1242 relative to classical minor groove agents. Interestingly, the DB75 benzimidazole derivative, DB293 (Figure 1), also forms a 2:1 complex in the sequence -ATGA- that has one GC base pair.[28, 42] DB293 does bind well to AT sites, however, with close to the same binding constant as for -ATGA-, and it has the classical minor groove concave shape. Thermodynamic analysis of the DB293 interaction with the -ATGA- site indicates that formation of the 2:1 complex is enthalpy driven. In contrast, binding of DB293 to the -AATT- site is entropy driven as with other AT specific minor groove binding agents.

The strong positive induced CD signal for the DB1242 complex at -GCTCG- indicates a stacked minor groove dimer complex as expected for an unfused aromatic diamidine. These clear experimental observations lead to important questions about the recognition of the GC rich site by DB1242, in contrast to the many diamidines and other minor groove agents investigated to date. We have conducted a preliminary analysis of molecular models of DB1242 relative to the other linear compounds of Figure 1 to obtain a better idea of why the pyrimidine ring of DB1242 may provide a unique recognition module for -GCTCG-. Some very interesting molecular structural differences among the compounds and a plausible molecular explanation for all of our observations were suggested by the models.

Molecular features that can explain the unique interaction of DB1242 with -GCTCG- were readily incorporated into a stacked dimer model (Figure 7). The two amidines on the inner face and ends of the stacked dimer are in an optimum position to interact with bases at the floor of the minor groove. The amidines on the outside, convex, face of the dimer are in a position and have an appropriate twist to form hydrogen bonds with the pyrimidine nitrogens on the outside face of the other stacked molecule of DB1242, to give two intermolecular amidine-pyrimidine hydrogen bonds within the dimer. The pyrimidine nitrogens have a negative electrostatic potential that interacts favorably with the positive potential of the amidines (Figure 7B). The pyrimidine nitrogens on the concave face of the dimer are in position to accept hydrogen bonds from the G-NH_2 groups in the minor groove. In the 5-GCTCG-3′•′-CGAGC-5′ double helix there are four G-NH_2 that could potentially donate hydrogen bonds to the pyrimidine N. This interaction with the G-NH_2 groups dictated stacking orientation (i) in the dimer (Figure 7B). The opposite stacking orientation places the pyrimidine nitrogens too close together to provide interactions with the G bases of the -GCTCG- sequence. The center of the stacked dimer fits snugly against the groove floor and provide shape-specific recognition of the AT base pair in the center of the -GCTCG- sequence.

In Figure 7C the 5′-G5C6T7C8G9-3′ sequence is on the right and G5, in the upper right, is colored yellow. The upper amidine of the top molecule of the DB1242 dimer forms a hydrogen bond to the keto of the C of the G5-C base pair. The amino N of the G that is hydrogen bonded to C6 is 3.4 Å from the pyrimidine N of the upper DB1242. This is long for a hydrogen bond but the interaction certainly stabilizes the complex. There is a hydrogen bond between the top amidine of the lower molecule of the dimer and the upper pyrimidine (outer N), and the same amidine forms a hydrogen bond to the keto of C8. The G9 amino group forms a hydrogen bond to the inner pyrimidine N of the lower molecule of the dimer (G9 is colored magenta in Figure 7C). A hydrogen bond between the lower amidine of the top molecule of the dimer and the outer pyrimidine N of the other DB1242 can also be seen. The bottom amidine of the lower molecule of the dimer is away from the bases at the floor of the groove but is close to the phosphate of A10 (3.4 Å from amidine N to the phosphate O at the bottom left of Figure 7C). Thus the requirements for all of the base pairs of the -GCTCG- sequence are explained by the dimer model and the docked structure explains the specific and unusual recognition of this sequence by DB1242.

The modeling results provided a better understanding of the experimental results. As can be seen from Table 1, the first molecule on DB1242 to bind at -GCTCG- has a small Gibbs energy of binding with small favorable ΔH and ΔS values. The second molecule to bind, however, locks both molecules into the site and forms the extended contacts and numerous hydrogen bonds in an optimized structure. The final complex has a very favorable binding enthalpy but an unfavorable ΔS for binding. Although the proposed model is preliminary and hypothetical at this point, it does explain all experimental observations for binding of DB1242 to -GCTCG-. In addition, the model provides a very clear explanation for why the pyrimidine ring of DB1242 leads to a strong interaction at the GC sequence but why neither DB1111, without a ring nitrogen, or DB1164, with a single ring nitrogen, can bind as strongly to -GCTCG-. The DB1242 stacked dimer complex provides ideas for design of new compounds for specific recognition of a broad range of DNA sequences that were not previously thought possible with heterocyclic diamidines.

Acknowledgments

This work was supported by grants (to W. D. W. and D. W. B.) from the National Institutes of Health (GM61587 and AI064200) and (to M.-H. D.-C) from the Association pour la Recherche contre le Cancer (ARC-3717). We thank the Institut de Recherches sur le Cancer de Lille and the Conseil Régional Nord-Pas-de-Calais for a PhD fellowship to Paul Peixoto.

References

1. Russ AP, Lampel S. Drug Discov Today. 2005; 10:1607–10. [PubMed: 16376820]
2. Hajduk PJ, Huth JR, Tse C. Drug Discov Today. 2005; 10:1675–82. [PubMed: 16376828]
3. Lansiaux A, Tanious F, Mishal Z, Dassonneville L, Kumar A, Stephens CE, Hu Q, Wilson WD, Boykin DW, Bailly C. Cancer Res. 2002; 62:7219–29. [PubMed: 12499262]
4. Lansiaux A, Dassonneville L, Facompre M, Kumar A, Stephens CE, Bajic M, Tanious F, Wilson WD, Boykin DW, Bailly C. J Med Chem. 2002; 45:1994–2002. [PubMed: 11985467]
5. Susbielle G, Blattes R, Brevet V, Monod C, Kas E. Curr Med Chem Anticancer Agents. 2005; 5:409–20. [PubMed: 16101491]
6. Mathis AM, Bridges AS, Ismail MA, Kumar A, Francesconi I, Anbazhagan M, Hu Q, Tanious FA, Wenzler T, Saulter J, Wilson WD, Brun R, Boykin DW, Tidwell RR, Hall JE. Antimicrob Agents Chemother. 2007 May.
7. Tidwell, RR.; Boykin, DW. Dicationic DNA minor-groove binders as antimicrobial agents. In: Demeunynck, M.; Bailly, C.; Wilson, WD., editors. DNA and RNA Binders: From Small Molecules to Drugs. Vol. 2. WILEY-VCH; Weinheim: 2003. p. 414-460.
8. Wilson WD, Nguyen B, Tanious FA, Mathis A, Hall JE, Stephens CE, Boykin DW. Curr Med Chem Anti-Canc Agents. 2005; 5:389–408.
9. Neidle S. Nat Prod Rep. 2001; 18:291–309. [PubMed: 11476483]
10. Bailly C, Chaires JB. Bioconjug Chem. 1998; 9:513–538. [PubMed: 9736486]
11. Dervan PB, Edelson BS. Curr Opin Struct Biol. 2003; 13:284–99. [PubMed: 12831879]
12. Miao Y, Lee MP, Parkinson GN, Batista-Parra A, Ismail MA, Neidle S, Boykin DW, Wilson WD. Biochemistry. 2005; 44:14701–8. [PubMed: 16274217]
13. Ismail MA, Arafa RK, Brun R, Wenzler T, Miao Y, Wilson WD, Generaux C, Bridges A, Hall JE, Boykin DW. Bioorg Med Chem. 2006; 13:7434–45.
14. Nguyen B, Lee MP, Hamelberg D, Joubert A, Bailly C, Brun R, Neidle S, Wilson WD. J Am Chem Soc. 2002; 124:13680–13681. [PubMed: 12431090]
15. Cooper A, Johnson CM, Lakey JH, Nollmann M. Biophys Chem. 2001; 93:215–30. [PubMed: 11804727]
16. Cooper A. Biophys Chem. 2005; 115:89–97. [PubMed: 15752588]
17. Nguyen B, Hamelberg D, Bailly C, Colson P, Stanek J, Brun R, Neidle S, Wilson WD. Biophys J. 2004; 86:1028–1041. [PubMed: 14747338]
18. Goodsell D, Dickerson RE. J Med Chem. 1986; 29:727–733. [PubMed: 2422377]
19. Cory M, Tidwell RR, Fairley TA. J Med Chem. 1992; 35:431–438. [PubMed: 1738139]
20. Nguyen B, Tardy C, Bailly C, Colson P, Houssier C, Kumar A, Boykin DW, Wilson WD. Biopolymers. 2002; 63:281–97. [PubMed: 11877739]
21. Im H, Bernstein ER. J Chem Phys. 1988; 88:7337–47.
22. Ando S, Hironaka T, Kurosu H, Ando I. Magnetic Resonance in Chemistry. 2000; 38:241–50.
23. Turchaninov VK, Vokin AI, Korostova. Russian Chemical Bulletin. 1997; 46:1407–14.
24. Ismail MA, Arafa RK, Brun R, Wenzler T, Miao Y, Wilson WD, Generaux C, Bridges A, Hall JE, Boykin DW. Bio Med Chem. 2005; 13:6718–6726.
25. Fasman, GD. Handbook of Biochemistry and Molecular Biology, Nucleic Acids. 3. Vol. 1. CRC Press; Cleveland, OH: 1975. p. 589
26. Wang L, Bailly C, Kumar A, Ding D, Bajic M, Boykin DW, Wilson WD. Proc Natl Acad Sci U S A. 2000; 97:12–6. [PubMed: 10618362]
27. Bailly, C.; Kluza, J.; Martin, C.; Ellis, T.; Waring, MJ. DNase I Footprinting of Small Molecule Binding Sites on DNA. In: Herdewijn, P., editor. Oligonucleotide Synthesis: Methods and Applications. Vol. 288. Humana Press; Totowa, NJ: 2004. p. 319-342.
28. Mallena S, Lee MP, Bailly C, Neidle S, Kumar A, Boykin DW, Wilson WD. J Am Chem Soc. 2004; 126:13659–69. [PubMed: 15493923]
29. Nguyen B, Tanious FA, Wilson WD. Methods. 2007; 42:150–61. [PubMed: 17472897]
30. Davis TM, Wilson WD. Anal Biochem. 2000; 284:348–353. [PubMed: 10964419]

31. Spartan '04. Wavefunction Inc; Irvine, CA: 2004.
32. SYBYL Molecular Modeling Software. 7.2. Tripos Inc; St. Louis, MO: 2006.
33. Gasteiger J, Marsili M. Tetrahedron Letters. 1980; 36:3219–28.
34. Baruah H, Bierbach U. Nucleic Acids Res. 2003; 31:4138–46. [PubMed: 12853631]
35. Fengler A, Brandt W. J Mol Model. 1999; 5:177–188.
36. Lee T, Cho M, Ko SY, Youn HJ, Baek DJ, Cho WJ, Kang CY, Kim S. J Med Chem. 2007; 50:585–9. [PubMed: 17266209]
37. Chen J, Wang J, Xie X. J Chem Inf Model. 2007 ASAP Article.
38. Wilson WD, Tanious FA, Ding D, Kumar A, Boykin DW, Colson P, Houssier C, Bailly C. J Am Chem Soc. 1998; 120:10310–10321.
39. Abu-Daya A, Brown PM, Fox KR. Nucleic Acids Res. 1995; 23:3385–3392. [PubMed: 7567447]
40. Bostock-Smith CE, Harris SA, Laughton CA, Searle MA. Nucleic Acids Res. 2001; 29:693–702. [PubMed: 11160891]
41. Tanious FA, Hamelberg D, Bailly C, Czarny A, Boykin DW, Wilson WD. DNA sequence dependent monomer-dimer binding modulation of asymmetric benzimidazole derivatives. J Am Chem Soc. 2004; 126:143–53. [PubMed: 14709078]
42. Wang L, Carrasco C, Kumar A, Stephens CE, Bailly C, Boykin DW, Wilson WD. Biochemistry. 2001; 40:2511–21. [PubMed: 11327873]
43. Rodger, A.; Nordén, B. Circular Dichroism and Linear Dichroism. Oxford University Press; New York: 1997.
44. Sancho-Garcia JC, Cornil J. J Chem Phys. 2004; 121:3096–101. [PubMed: 15291619]
45. Wilson WD, Strekowski L, Tanious FA, Watson RA, Mokrosz JL, Strekowska A, Webster GD, Neidle S. J Am Chem Soc. 1988; 110:8292–8299.
46. Waring MJ, Bailly C. J Mol Recognit. 1997; 10:121–7. [PubMed: 9408827]
47. Bailly C, Perrine D, Lancelot JC, Saturnino C, Robba M, Waring MJ. Biochem J. 1997; 323:23–31. [PubMed: 9173886]
48. Lown JW, Krowicki K, Balzarini J, De Clercq E. J Med Chem. 1986; 29:1210–4. [PubMed: 3027326]
49. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE. Proc Natl Acad Sci U S A. 1985; 82:1376–1380. [PubMed: 2983343]
50. Hamelberg D, McFail-Isom L, Williams LD, Wilson WD. J Am Chem Soc. 2000; 122:10513–10520.
51. Wilson, WD. Reversible interactions of nucleic acids with small molecules. In: Blackburn, GM.; Gait, MJ., editors. Nucleic Acids in Chemistry and Biology. 2. Oxford University Press, Inc; New York: 1997. p. 329-374.
52. Mazur S, Tanious FA, Ding D, Kumar A, Boykin DW, Simpson IJ, Neidle S, Wilson WD. Journal of Molecular Biology. 2000; 300:321–337. [PubMed: 10873468]
53. Haq I, Ladbury JE, Chowdhry BZ, Jenkins TC, Chaires JB. J Mol Biol. 1997; 271:244–57. [PubMed: 9268656]
54. Breslauer KJ, Remeta DP, Chou WY, Ferrante R, Curry J, Zaunczkowski D, Snyder JG, Marky LA. Proc Natl Acad Sci U S A. 1987; 84:8922–8926. [PubMed: 2827160]
55. Kissinger K, Krowicki K, Dabrowiak JC, Lown JW. Biochemistry. 1987; 26:5590–5. [PubMed: 2823885]
56. Pelton JG, Wemmer DE. Proc Natl Acad Sci U S A. 1989; 86(15):5723–7. [PubMed: 2762292]
57. Pelton JG, Wemmer DE. Biochemistry. 1988; 27:8088–96. [PubMed: 3233197]
58. Mrksich M, Wade WS, Dwyer TJ, Geierstanger BH, Wemmer DE, Dervan PB. Proc Natl Acad Sci U S A. 1992; 89:7586–90. [PubMed: 1323845]
59. Laughton CA, Tanious F, Nunn CM, Boykin DW, Wilson WD, Neidle S. Biochemistry. 1996; 35:5655–61. [PubMed: 8639524]

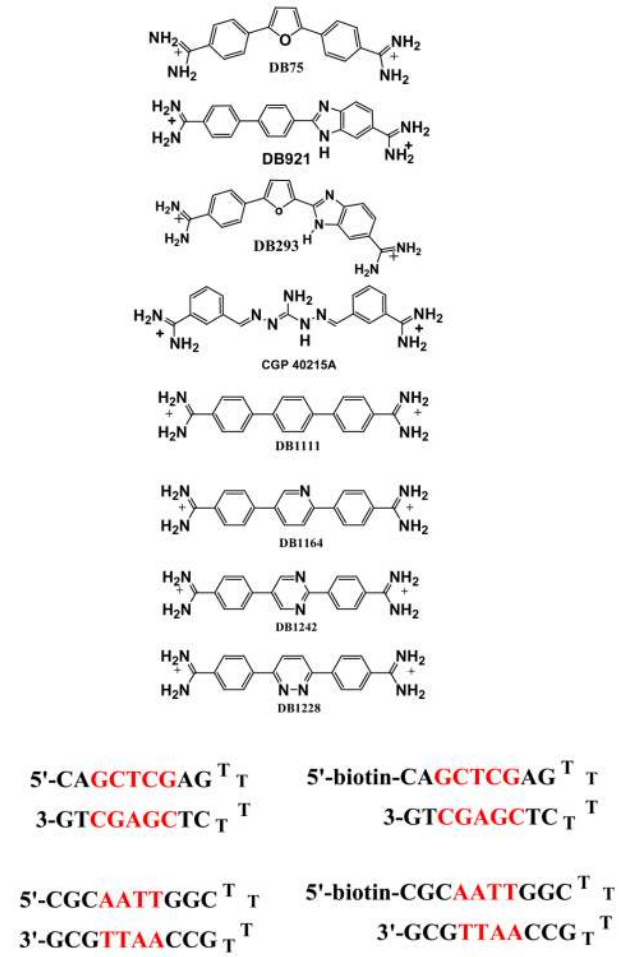

Figure 1.
Compound structures and DNA oligomer sequences used in this study.

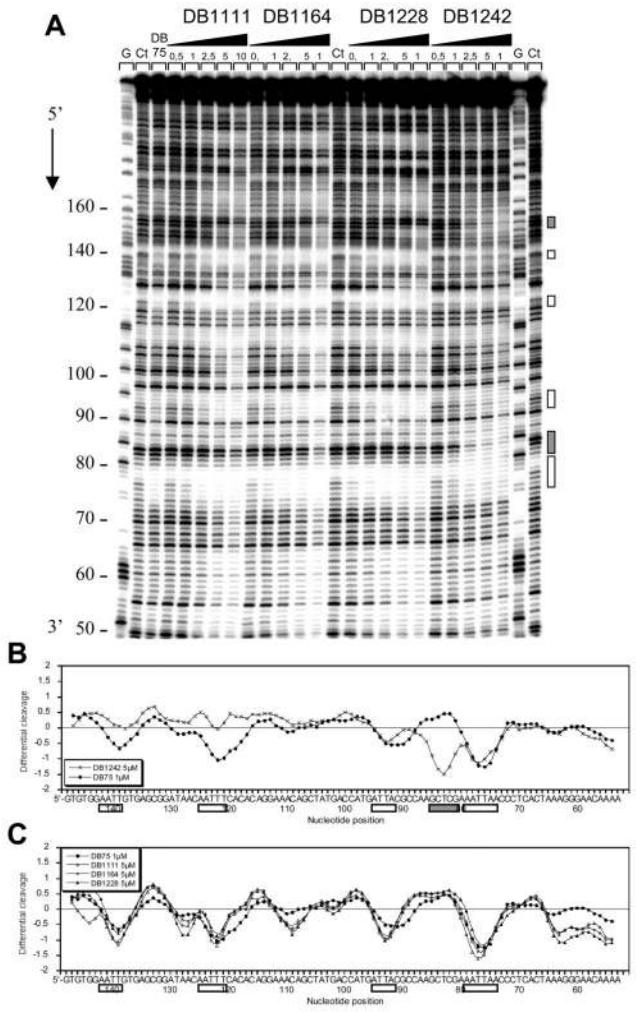

Figure 2.

DNase I footprinting titration experiments. The pBS plasmid was isolated and purified from E. coli using Qiagen columns. The 265 bp DNA fragment was prepared by 3′-[^{32}P]-end labeling of the EcoRI-PvuII double digest of the pBS plasmid (Stratagene) using α-[^{32}P]-dATP and AMV reverse transcriptase. (A) The products of the DNase I digestion were resolved on an 8% polyacrylamide gel containing 8 M urea. Drug concentrations are at the top of the lanes. Tracks labeled G represent dimethylsulfate piperidine markers specific for guanines. Differential cleavage plots compare the susceptibility of the DNA to cutting by DNase I in the presence of (B) DB75 and DB1242 (C) DB75, DB1111, DB1164 and DB1228. Deviation of points toward the lettered sequence (negative values) corresponds to a ligand-protected site and deviation away (positive values) represents enhanced cleavage. The vertical scale is in units of ln(fa) - ln(fc), where fa is the fractional cleavage at any bond in the presence of the drug and fc is the fractional cleavage of the same bond in the control.

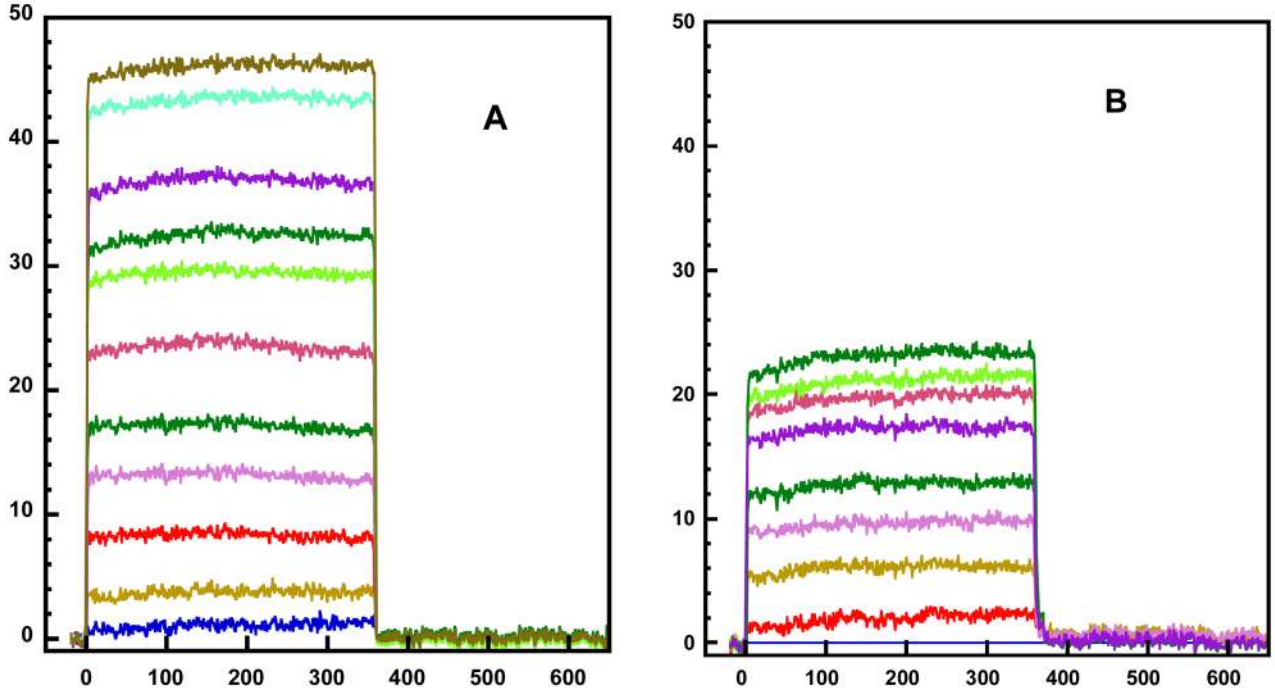

Figure 3.

SPR sensorgrams for DB1242 with (A) -GCTCG- and (B) -AATT- hairpin DNA (figure 1). The compound concentrations were 0.1, 0.3, 0.5, 1.0, 1.4, 1.8, 2.2, 2.4, 2.8, 3.5, 4.0 to 4 μM from bottom to top in (A) and (B). The experiments were carried out in cacodylic acid buffer at 25 °C.

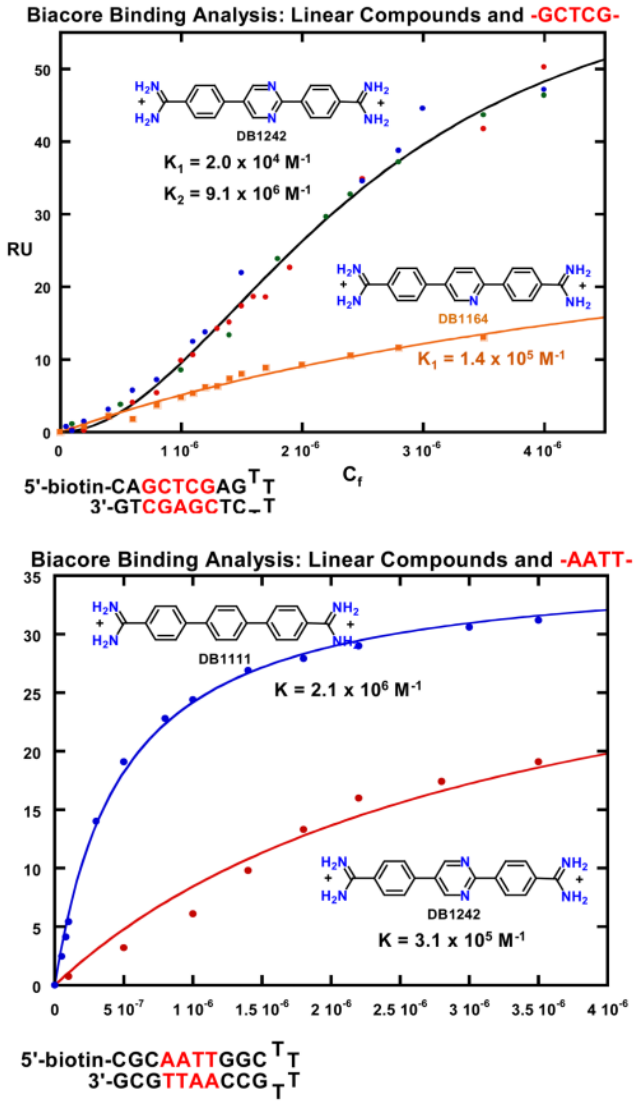

Figure 4.

RU values from the steady state region of SPR sensorgrams are plotted against the unbound compound concentration, C_f (flow solution): (A) DB1242 and DB1164 binding to -GCTCG- DNA hairpin and (B) DB1242 and DB1111 with the AATT DNA hairpin. The data in A were fitted to a two site model and that in B were fitted to one site model using equation 1.

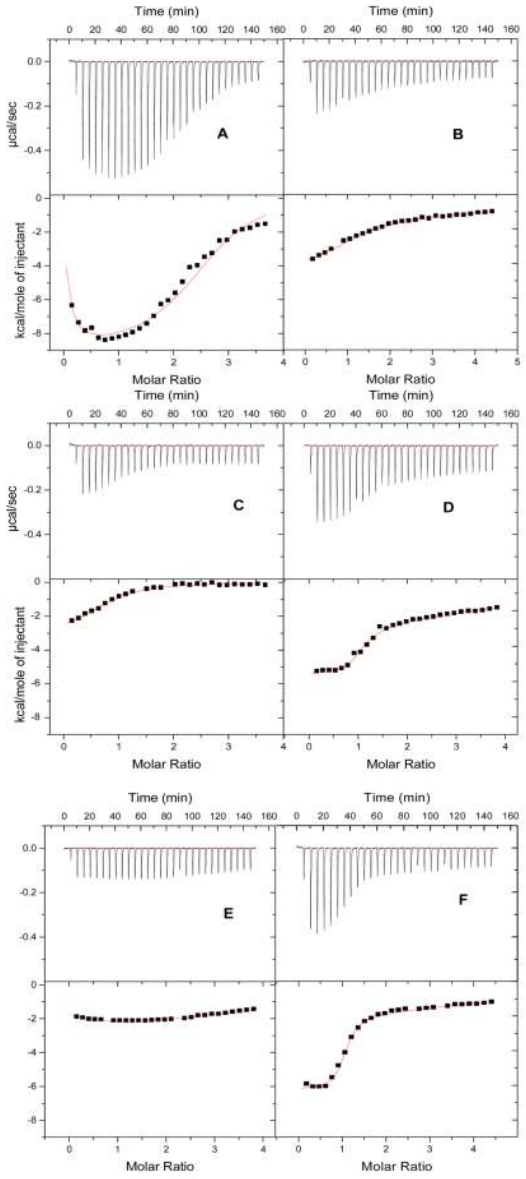

Figure 5.

ITC curves (12μM hairpin duplex) for the binding of DB1242 to the (A) –GCTCG- and (B) -AATT-hairpin; DB1111 to the (C) -GCTCG- and (D) -AATT- hairpin; DB1164 to the (E) -GCTCG- and (F) -AATT- hairpin. In each panel the top plot is the baseline corrected experimental data. For the lower plots results were converted to molar heats and plotted against the compound to DNA molar ratio. The same buffer conditions were used as in Figure 3.

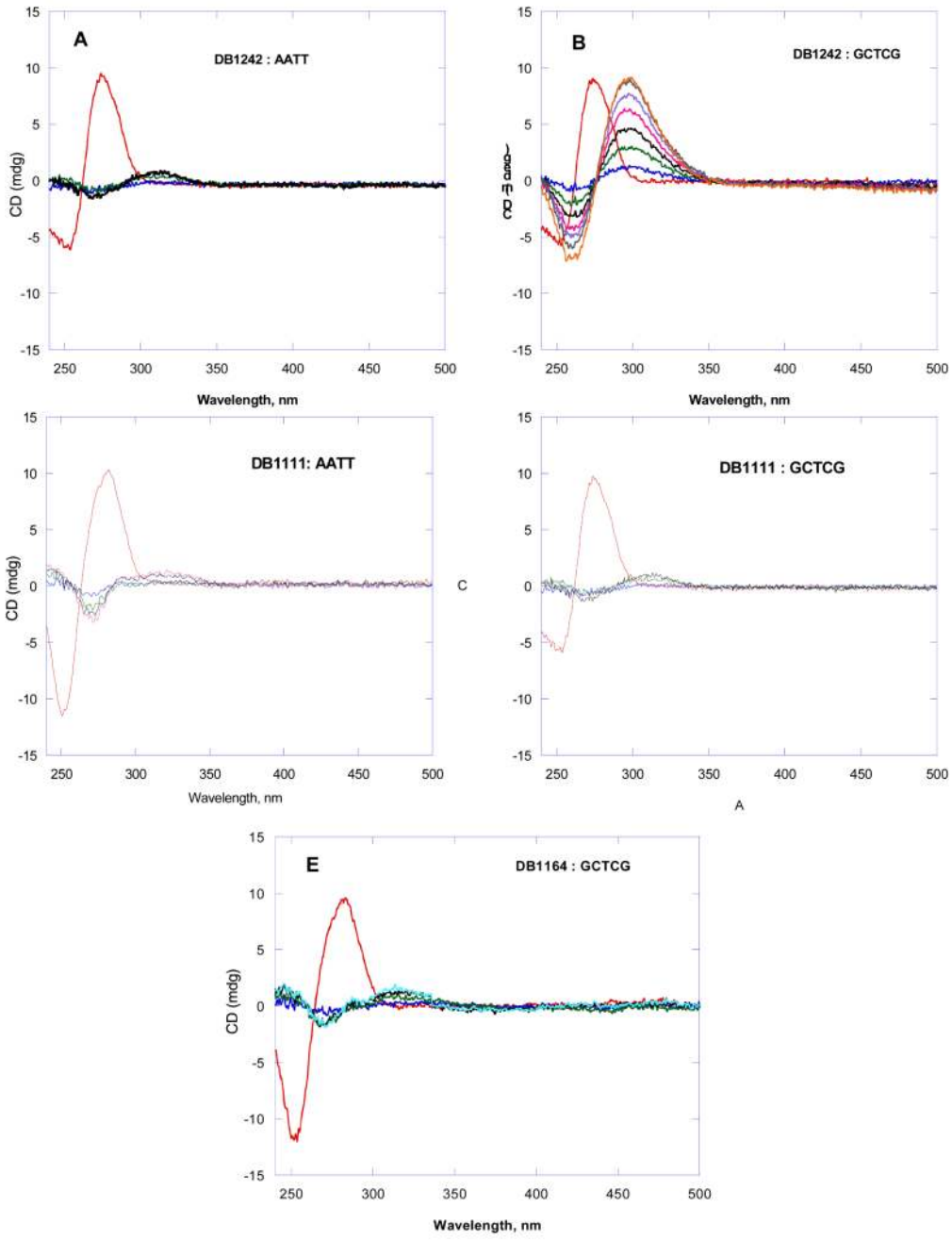

Figure 6.

Difference CD spectra for DB1242, DB1111 and DB1164. In each panel a CD spectrum for DNA is shown along difference CD spectra for the titration of compound into the DNA solution. (A) DB1242 with -AATT-: ratios of the compound to DNA hairpin from bottom to top at 300 nm are 0.5, 2.0, 2.5 and 3.0. (B) DB1242 with -GCTCG-: ratios of compound to DNA from bottom to top are 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8 (C) DB1111and -AATT-: ratios of compound to DNA from bottom to top are 0.5, 1.0, 1.5 (D) DB1111 with -GCTCG-: ratios of compound to DNA from bottom to top are 0.5, 1.0, 2.0, 3.0 (E) DB1164 with -GCTCG-: ratios of compound to DNA from bottom to top are 1.0, 1.5, 2.0, 2.5, 3.0. The same buffer conditions were used as in Figure 3. Because of the relatively

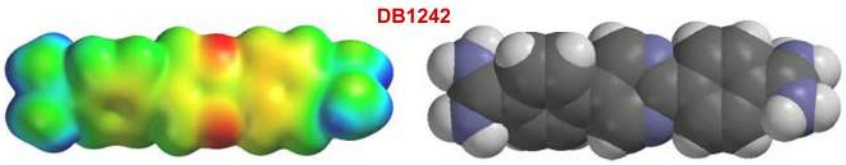

Figure7A

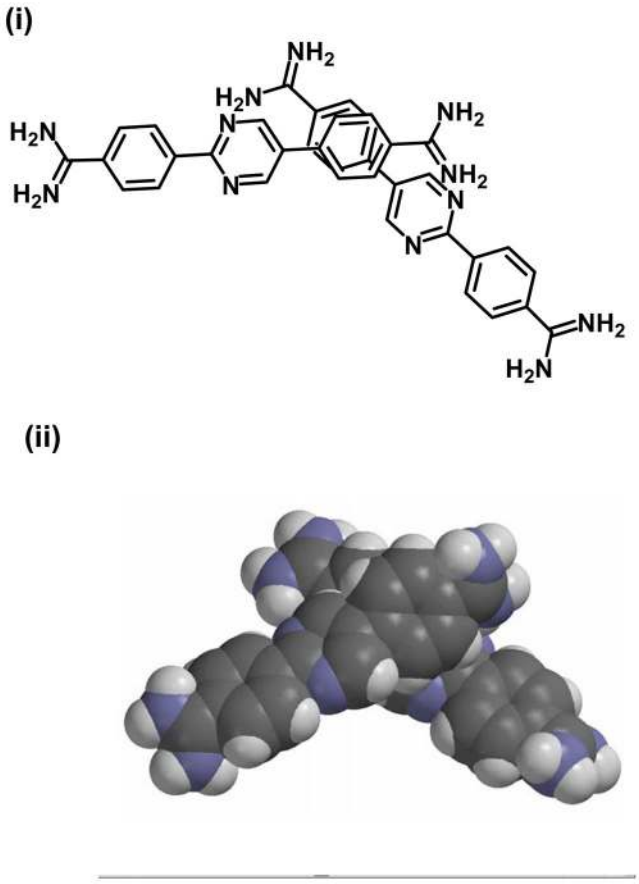

Figure7B

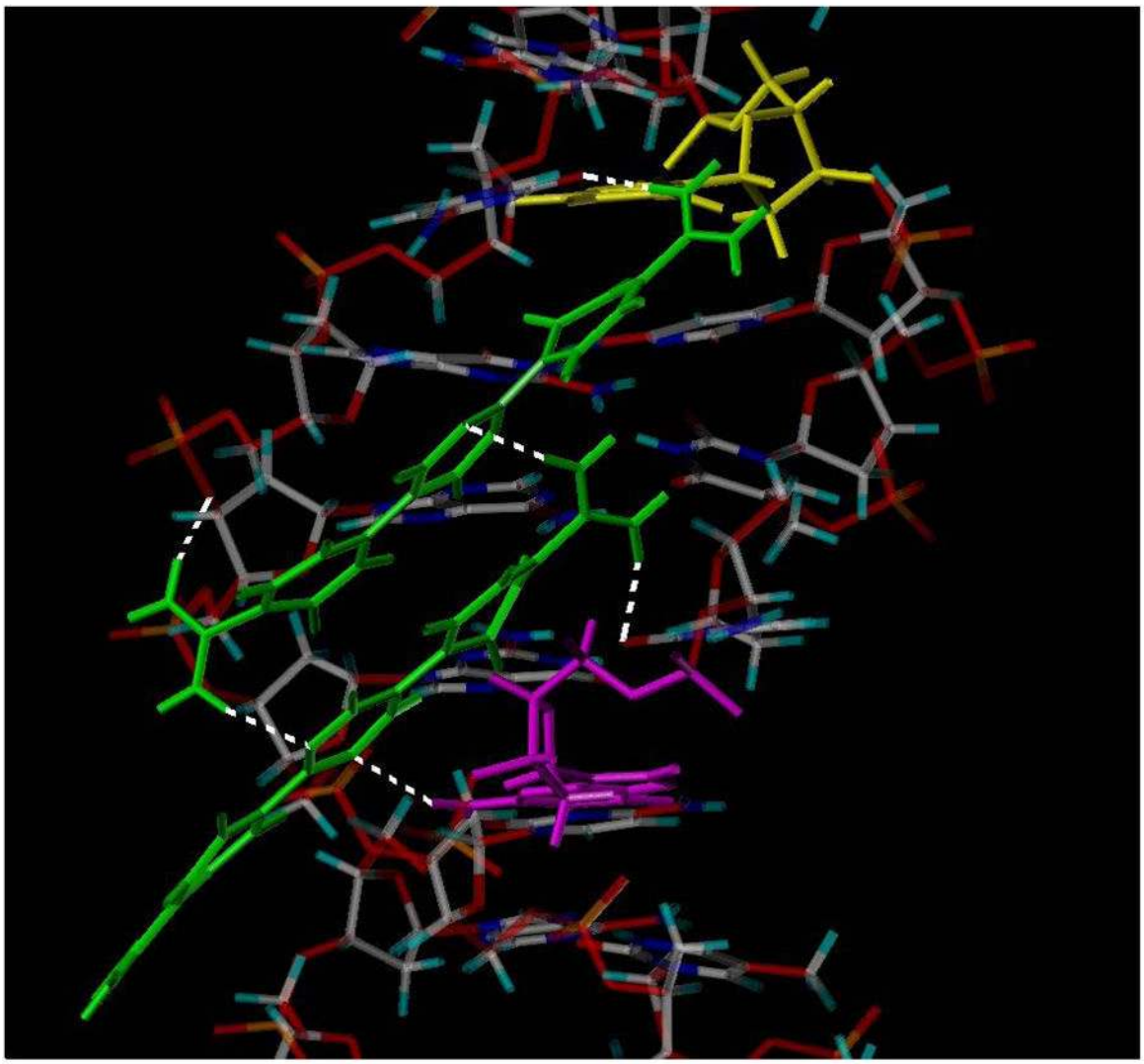

Figure 7C

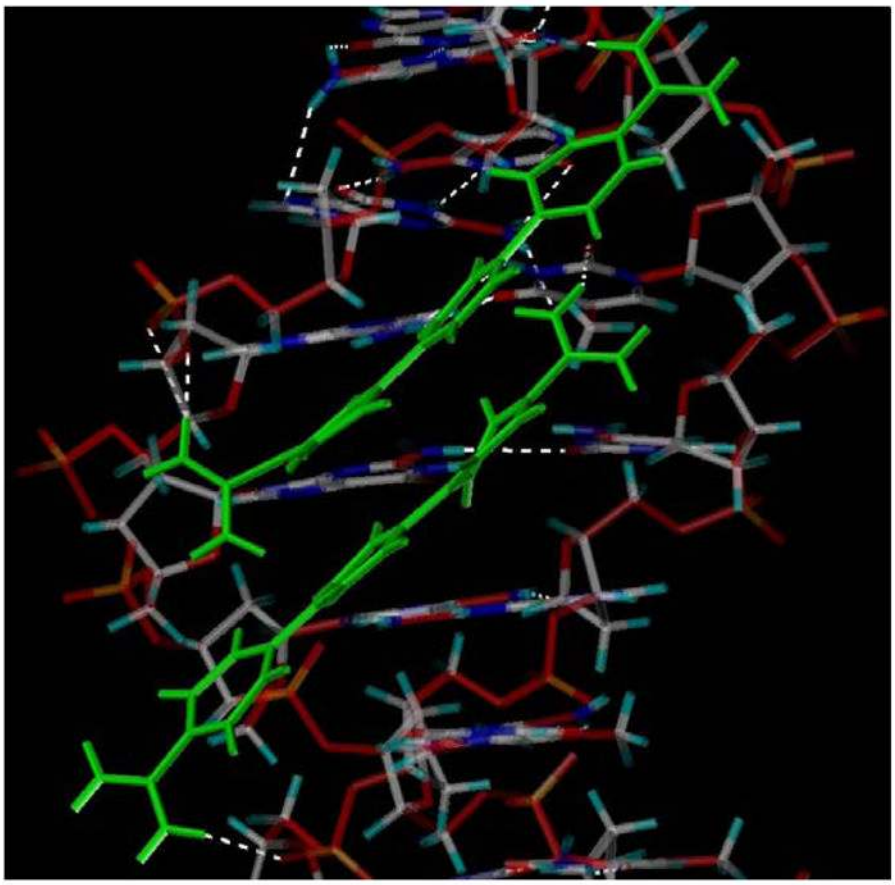

Figure 7D

Figure 7.
(A) Geometry-optimized models for DB1242.[31] A color coded electron density map is on the left and a space-filling model with atom colors is displayed on the right (carbon-black, nitrogen-blue, and hydrogen-white). (B) Stacked dimer models for DB1242: (i) stick model and (iii) space filling model. (C) Flexidock generated DB1242 dimer (green) in complex with DNA minor groove binding site-GCTCG-. A representative low energy structure is shown for this DNA dimer complex. The 3′G is presented in magenta and the 5′G is shown in yellow; key hydrogen bonds are displayed in white. Note that two hydrogen bonds are formed between the molecules of the dimer. (D) Flexidock generated DB1111 dimer (green) in complex with DNA minor groove binding site-GCTCG-. Example structures are shown for this DNA complex with hydrogen bonds in white. Note that there are no hydrogen bonds between the dimer molecules.

Table 1

Binding constants and Experimental Thermodynamic Values at 25 °C for Interaction of DB1111, DB1242 and DB1164 with GCTCG and AATT DNA Hairpin.

Compd	DNA	K_1 (M^{-1})	K_2 (M^{-1})	ΔH_1 (Kcal/mol)	ΔH_2 (Kcal/mol)	$T\Delta S_1$ (Kcal/mol)	$T\Delta S_2$ (Kcal/mol)	ΔG_1 (Kcal/mol)	ΔG_2 (Kcal/mol)
DB1111	GCTCG*	1.5×10^5	----	−3.1	---	−4.0	-----	−7.1	-----
	AATT*	2.1×10^6	1.2×10^5	−5.7	----	−3.0		−8.7	−6.9
DB1242	GCTCG**	2.0×10^4	9.1×10^6	−3.3	−15.8	−2.6	−6.3	−5.9	−9.5
	AATT*	3.1×10^5	9.0×10^4	−4.7		−2.9	-----	−7.6	−6.7
DB1164	GCTCG*	1.4×10^5	----	−3.0	----	−4.0		−7.0	-----
	AATT*	5.8×10^6	2.1×10^5	−6.2	----	−3.0	-----	−9.2	−7.2

Experiments were carried out in Cacodylic acid buffer at 25 °C. Values of ΔH are obtained from calorimetry; K are from SPR; TΔS are calculated from the relationship ΔG=ΔH − TΔS.

The error for ΔH is 2% to 4%

*Error for K,ΔG and TΔS is 15%

**Error for K,ΔG and TΔS is 30%