

Designed Compounds for Recognition of 10 Base Pairs of DNA with Two AT Binding Sites

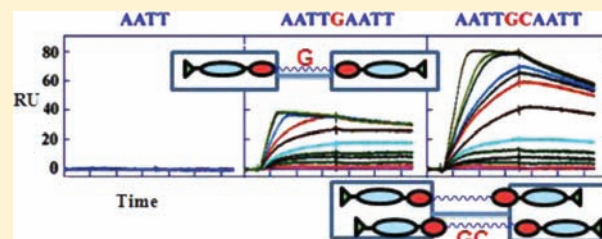
Yang Liu,[†] Yun Chai,^{†,§} Arvind Kumar,[†] Richard R. Tidwell,[‡] David W. Boykin,[†] and W. David Wilson^{*,†}

[†]Department of Chemistry, Georgia State University, Atlanta, Georgia 30302-4098, United States

[‡]Department of Pathology, University of North Carolina, Chapel Hill, North Carolina 27599, United States

S Supporting Information

ABSTRACT: Short AT base pair sequences that are separated by a small number of GCs are common in eukaryotic parasite genomes. Cell-permeable compounds that bind effectively and selectively to such sequences present an attractive therapeutic approach. Compounds with linked, one or two amidine-benzimidazole-phenyl (ABP) motifs were designed, synthesized, and evaluated for binding to adjacent AT sites by biosensor-surface plasmon resonance (SPR). A surprising feature of the linked ABP motifs is that a set of six similar compounds has three different minor groove binding modes with the target sequences. Compounds with one ABP bind independently to two separated AT sites. Unexpectedly, compounds with two ABP motifs can bind strongly either as monomers or as cooperative dimers to the full site. The results are supported by mass spectrometry and circular dichroism, and models to explain the different binding modes are presented.



INTRODUCTION

The concept of specific control of cellular gene expression by designed compounds is a key goal of chemical biology and offers many advantages in the development of new types of drugs as well as agents for biotechnology. The control could be affected, for example, by directly targeting genes with designed oligomers, such as PNA,^{1,2} or with a variety of small molecules.^{3–5} Sequences for most of the 20–25 thousand human genes are now known and all of them are potentially susceptible to external control.⁶ The sequences of the genomes of a large number of disease causing microorganisms have also been determined and offer selective strategies to control or destroy the organisms.^{7,8} Designed compounds that have biological activity against either duplex,^{4,5,9–12} triplex and triplet-repeat DNA,^{13,14} or quadruplex DNA structures^{15–19} have also been identified and are under development.

The design of optimum compounds to target DNA for a therapeutic outcome must obey several rules. The compounds obviously must be cell permeable and this limits the group types that can be combined on a core compound template. The compounds must be large enough to give sufficient affinity and selectivity to generate activity while maintaining low toxicity. The cost of synthesis can also be a factor, especially for compounds that are designed for use against diseases such as malaria and sleeping sickness that affect many of the world's most financially depressed regions.

With these constraints as limiting features, we have focused in this work on the design of relatively simple compounds that can selectively target a 10 base pair sequence in the minor groove of DNA. To uniquely bind to a site in the entire human genome, about 15–16 base pairs must be in the binding site.²⁰

If we focus on genes and their control sequences while recognizing that not all genes are expressed in a specific cell, that number is reduced. Although the design concepts are general and could be extended to target DNA in a variety of cells, we have specifically focused on the mitochondrial kinetoplast genome of trypanosomes as an optimum, initial test system.^{11,21–23} The kinetoplast consists of thousands of interlocked AT-rich DNA minicircles that code for guide RNAs that control mRNA editing.^{1,24} The mRNAs are coded by maxicircle DNAs that are also part of the kinetoplast matrix. The possibility of targeting the kinetoplast as a unique DNA target structure is recognized,^{11,21–23} but new drugs to selectively target the unusual structure have not yet reached the final approval stage.²⁵ For the small mitochondrial genomes of these parasitic microorganisms, a 10 base pair target site can give the desired selectivity.²⁰ Trypanosomes cause sleeping sickness (*Trypanosoma brucei*) in Africa²⁵ and Chagas disease (*Trypanosoma cruzi*) from the southern United States to Argentina.²⁶ The organisms infect and kill many thousands of people each year and current drugs to treat the associated diseases have serious limitations from toxicity to resistance.^{11,25,27}

Both the kinetoplast of trypanosomes and the nuclear genome of malaria are quite AT rich with significant regions of AT base pairs separated by a small number of GC base pairs.^{9,11,21,22,24} To simplify and reduce the costs of synthesis, symmetric compounds have been prepared (Figure 1). With appropriate linking groups, the symmetry of such compounds

Received: December 13, 2011

Published: February 27, 2012

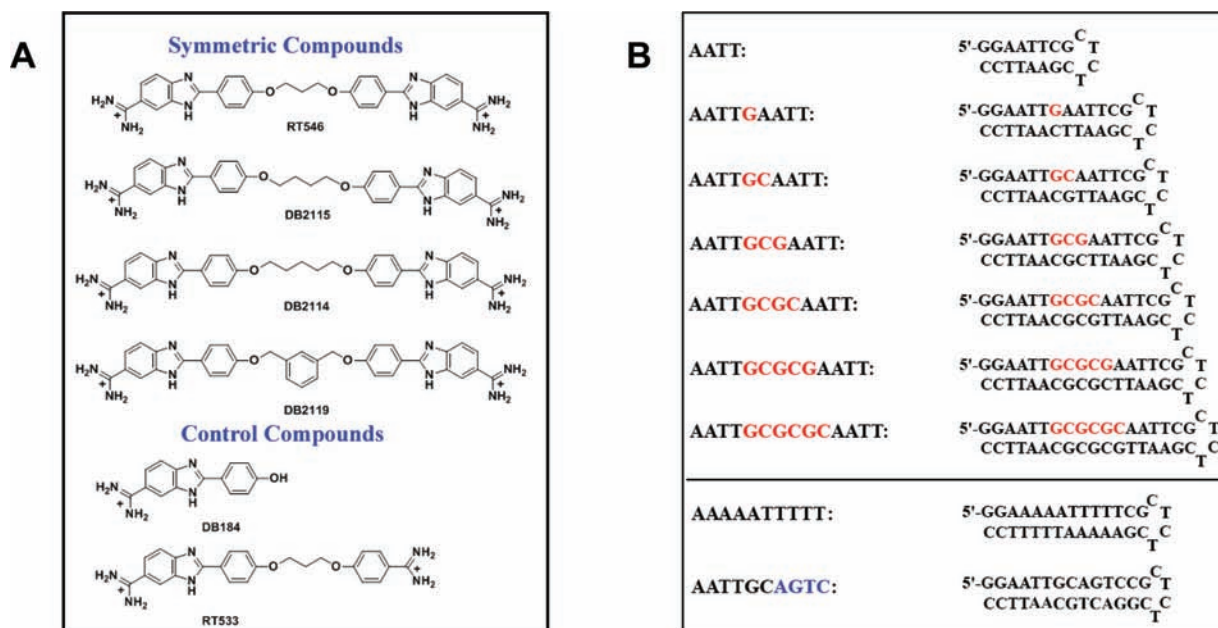


Figure 1. Structure of the compounds and the DNA sequences used in this study. For SPR experiments, 5'-biotin labeled DNA sequences are used.

makes them ideal for targeting two AT sites separated by GC base pairs. The amidine-benzimidazole-phenyl (ABP) motif is effective for targeting AT sites in DNA and we have incorporated that motif in our design strategy. In an initial set of test compounds, the symmetric compounds RT546, DB2114, DB2115, and DB2119 were designed with two ABP motifs, to specifically target two AT sequences with a flexible linker to connect them across one or more GC base pairs (Figure 1). Compounds DB184 and RT533, with single ABP motifs, are included as controls. The concept of connecting recognition units to recognize two or more sites in proteins or DNAs has been used in other systems.²⁸ In DNA, combilexins connect an intercalator and minor groove binding agent;^{29–31} bis- or polyintercalators can intercalate and bind in both the major and minor grooves;^{32–35} minor groove alkylating groups can be placed on two linker groove recognizing units to alkylate at two sites;^{12,36} and linked polyamide minor groove agents can also bind at adjacent sites.³⁷

The key questions that we wish to address with these compounds are (i) what level of binding affinity can be obtained with compounds of this type and DNA sequences containing AT base pair sites separated by one or more GC base pairs which are common in the parasitic organisms (Figure 1); (ii) do the compounds fold back through the flexible linkers to bind to single AT sites or can a single compound bind to both of the AT sites as desired; and (iii) for compounds that can bind simultaneously to both sites, what correlation is there between linker length, rigidity and the number of GC pairs between the AT sites? These questions can be answered with the designed set of compounds in Figure 1. The results presented here are quite surprising and show that even this small set of compounds has unique interactions with DNA including three completely different binding modes, even with the relatively simple set of DNA sequences in Figure 1.

MATERIALS AND METHODS

Materials. Syntheses of DB2114, DB2115, and DB2119 were done much the same as with published compounds DB184, RT546, and RT533,^{38,39} and the detailed syntheses are described in the Supporting

Information. The purity of all compounds was verified by NMR and elemental analysis.

In SPR experiments, 5'-biotin labeled hairpin DNA oligomers were used (DNA sequences shown in Figure 1B). In mass spectrometry (MS), circular dichroism (CD), and T_m experiments, the hairpin DNA oligomers used were AATT [5'-GGAATTCGCTCTCGAATTCC-3'], AATTGAATT [5'-GGAATTGAATTCGCTCTCGAATTCAATTCC-3'] and AATTGCAATT [5'-GGAATTCGCAATTCGCTCTCGAATTGCAATTCC-3'] with the hairpin loop sequences underlined. All DNA oligomers were obtained from Integrated DNA Technologies, Inc. (IDT, Coraville, IA) with reverse-phased HPLC purification and mass spectrometry characterization. The CCL10 buffers used in CD and T_m experiments contained 0.01 M cacodylic acid, 0.001 M EDTA, and 0.1 M NaCl, pH 6.25. The SPR experiments were performed in filtered, degassed CCL10 buffer with $5 \times 10^{-3}\%$ (v/v) Surfactant P20 and 0.1 M ammonium acetate, pH 7.0, was used in MS experiments.

Biosensor-Surface Plasmon Resonance (SPR) Studies. SPR measurements were performed with a four-channel Biacore T100 or T200 optical biosensor system (Biacore, GE Healthcare, Inc.). 5'-Biotin labeled DNA samples [AATT, AATTGAATT, AATTGCAATT, AATTGCGAATT, AATTGCGCAATT, AATTGCGCGAATT, AATTGCGCGCAATT hairpins] were immobilized onto streptavidin-coated sensor chips (Biacore SA) as previously described.^{40,41} Three flow cells were used to immobilize the DNA oligomer samples, while a fourth cell was left blank as a control. SPR binding analysis was performed with multiple injections of different compound concentrations over the immobilized DNA surface at a flow rate of 25 $\mu\text{L}/\text{min}$ and 25 $^\circ\text{C}$. For steady-state analysis, 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 one- ($K_2 = 0$) or two-site interaction models:

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

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.^{41,42} Kinetics fits were used when no steady state was reached as previously described.^{41,43}

Electrospray Mass Spectrometry (ESI-MS) Studies. ESI-MS experiments were performed on a Q-TOF micro (Waters Micromass, Manchester, U.K.) with its standard ESI source and general published procedures for DNA.^{44,45} The electrospray source was operated in the

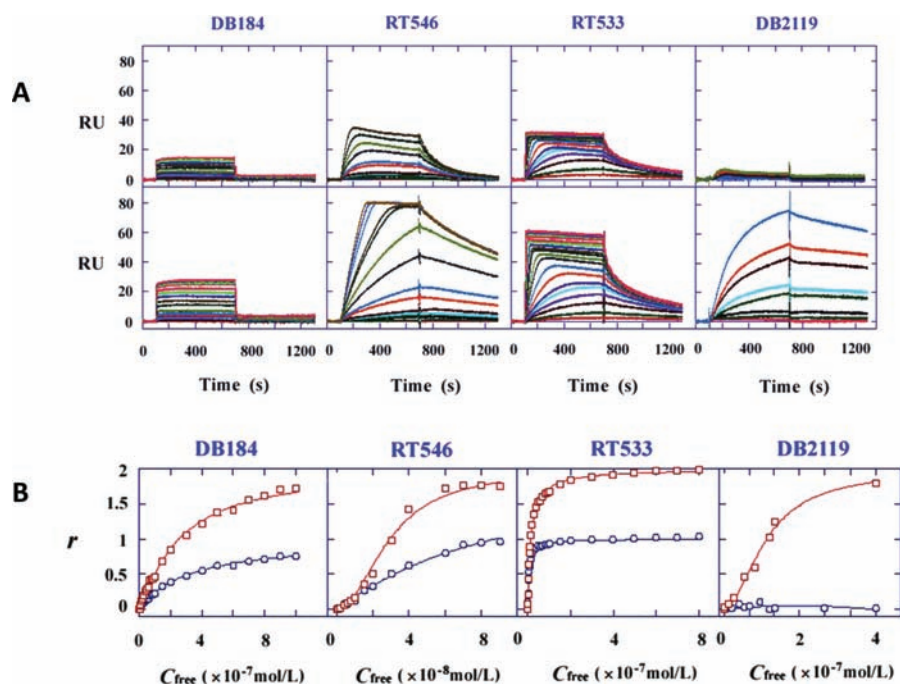


Figure 2. SPR binding affinity: (A) SPR sensorgrams for the interaction of selected compounds with AATT (upper panel) and AATTGCAATT (lower panel) DNA sequences. (B) Comparison of the SPR binding affinity for AATT (blue circles) and AATTGCAATT (red squares) DNA sequences with different compounds. RU values from the steady-state region of SPR sensorgrams are converted to r ($r = RU/RU_{max}$) and are plotted against the unbound compound concentration (flow solution) for DB184, RT546, RT533, and DB2119. The lines are the best fit values to a single site or two-site interaction models and K values are in Table 1.

negative ion mode with a needle voltage of -2.2 kV. The cone voltage was set to -30 V, and the RF lens1 voltage to -60 V. The hexapole collision voltage of 3 V and the ion energy of 1.5 V were used for full scan MS. Source block and desolvation temperatures were set to 70 and 100 °C, respectively. Spectra were acquired from 200 to 3200 m/z and only a portion of the mass range was shown for clarity. DNA oligomers were dialyzed in 0.1 M ammonium acetate, pH 7.0, using Spectra/Por 7 dialysis membranes (molecular weight cut off: 1000 Da; Spectrum Laboratories, Inc., CA) to remove the nonvolatile ions. Experiments were generally conducted at a concentration of 5×10^{-6} M for hairpin DNA and the ratios of compound to DNA (1:1, 2:1, and 3:1) were obtained by adding compound to DNA solution.

CD Titration Studies. CD spectra were recorded using a Jasco J-810 instrument with a 1 cm cell and a scan speed of 50 nm/min with a response time of 1 s. The spectra from 500 to 220 nm were averaged over five scans. A buffer baseline scan was collected in the same cuvette and subtracted from the average scan for each sample. The titration experiments were performed at 25 °C. The desired ratios of compound to DNA were obtained by adding compound to the cell containing a constant amount of DNA, and in all cases, the absorbance was kept below 1.0. Data processing and plotting were performed with Kaleidagraph software.

Molecular Docking. Molecular docking and visualization experiments were performed with the SYBYL-X1.2 software package on a Windows 8 processor Workstation.⁴⁶ The initial DNA duplexes 5'-d(GGAATTGAATTCG)-3' and 5'-d(GGAATTGCAATTCG)-3' were constructed in the Biopolymer-Build DNA Double Helix module employing regular B-DNA parameters. To accommodate two ligands binding in the minor groove of the duplex 5'-d-(GGAATTGCAATTCG)-3', the define distance-dependent constraints option was used to widen the minor groove. Two different models were constructed. The middle six base pairs were widened for the overlapped model (see Results for model descriptions), while only four base pairs were widened for the offset model. The groove was then allowed to decrease to the original width. These three DNA sequences were next energy minimized for a maximum of 100 iterations using the conjugate gradient algorithm and Tripos force

field, with a termination gradient of 0.1 kcal/(mol Å). RT546 was assigned Gasteiger-Hückel charges and minimized using the Tripos force field until a terminating conjugate gradient of 0.01 kcal/(mol Å) or the maximum 1000 iterations was reached.⁴⁷

To help position the bound compound in the AATTGCAATT site, the crystal structure, 3GJH, a 6-amidine-2-(4-amidino-phenyl)indole in complex with 5'-d(CGCGAATTCGCG)-3' from the protein data bank, was selected. The DNA duplex containing the AATTGCAATT site was aligned to the crystallized sequence such that AATT sites were overlaid and the DNA sequence from 3GJH was removed. This process was then repeated at the second AATT. A protomol was generated using a ligand-based approach with the Surflex-Dock module. During the protomol generation, the two parameters "Threshold" and "Bloat" were set to 0.5 and 2, respectively. "Threshold" determines how much the protomol extends into the minor groove, while "Bloat" impacts how far the protomol expands in three-dimensional space. All other parameters were left at the default values. After generation of the protomol, the GeomX module was used to dock RT546 (Figure 1).^{48,49}

The two AATTGCAATT duplexes were used with the RT546 dimer and the compound was manually docked to the widened minor groove. The complex was minimized with DNA fixed until a derivative of 0.01 kcal/(mol Å) was reached. Before each docking, the RT546 dimer was moved to a separate molecular area, so that the ligands can be moved independently of the DNA.⁴⁷ The genetic algorithm of the Flexdock module was employed implementing five different random numbers and 678 000 generations.^{47,50} Both the ligands and the bound DNA were permitted torsional flexibility in the docking process. Atomic charges were computed using the Kollman all-atom for DNA and Gasteiger-Hückel for the ligands. All the possible hydrogen-bond sites were selected for the DNA-compound complex. From each docking, the 20 lowest-energy structures were selected.

RESULTS

Biosensor-Surface Plasmon Resonance: Determining Binding Affinity, Stoichiometry, and Cooperativity. SPR experiments were used to quantitatively compare binding

Table 1

DNA	DB184		RT533	
	K^a (mol ⁻¹)	CF ^b	K^a (mol ⁻¹)	CF ^b
AATT	$K = 1.9 \times 10^6$	—	$K = 1.7 \times 10^8$	—
AATTGCAATT	$K_1 = 6.8 \times 10^6$ $K_2 = 2.2 \times 10^6$ $K = 3.9 \times 10^6$	1.3	$K_1 = 1.1 \times 10^8$ $K_2 = 6.4 \times 10^6$ $K = 2.6 \times 10^7$	0.2
AATTGCAATT	$K_1 = 5.4 \times 10^6$ $K_2 = 1.1 \times 10^6$ $K = 2.4 \times 10^6$	0.8	$K_1 = 1.7 \times 10^8$ $K_2 = 2.9 \times 10^7$ $K = 7.0 \times 10^7$	0.7
AATTGCCAATT	$K_1 = 7.6 \times 10^6$ $K_2 = 2.2 \times 10^6$ $K = 4.1 \times 10^6$	1.1	$K_1 = 1.1 \times 10^8$ $K_2 = 4.9 \times 10^7$ $K = 7.3 \times 10^7$	1.8
AATTGCGCAATT	$K_1 = 6.2 \times 10^6$ $K_2 = 1.4 \times 10^6$ $K = 2.9 \times 10^6$	0.9	$K_1 = 8.1 \times 10^7$ $K_2 = 6.9 \times 10^7$ $K = 7.5 \times 10^7$	3.4
AATTGCGCAATT	$K_1 = 6.5 \times 10^6$ $K_2 = 1.8 \times 10^6$ $K = 3.4 \times 10^6$	1.1	$K_1 = 6.4 \times 10^7$ $K_2 = 5.8 \times 10^7$ $K = 6.1 \times 10^7$	3.6
AATTGCGCGCAATT	$K_1 = 6.5 \times 10^6$ $K_2 = 1.4 \times 10^6$ $K = 3.0 \times 10^6$	0.9	$K_1 = 7.2 \times 10^7$ $K_2 = 4.6 \times 10^7$ $K = 5.8 \times 10^7$	2.6
DNA	RT546		DB2119	
	K^a (mol ⁻¹)	CF ^b	K^a (mol ⁻¹)	CF ^b
AATT	$K = 6.8 \times 10^6$	—	$K < 10^4$	—
AATTGCAATT	$K = 5.1 \times 10^7$	—	$K = 3.3 \times 10^6$	—
AATTGCAATT	$K_1 = 1.1 \times 10^6$ $K_2 = 9.4 \times 10^8$ $K = 3.2 \times 10^7$	3400	$K_1 = 2.4 \times 10^6$ $K_2 = 5.9 \times 10^7$ $K = 1.2 \times 10^7$	98
AATTGCCAATT	$K_1 = 3.4 \times 10^6$ $K_2 = 2.7 \times 10^8$ $K = 3.0 \times 10^7$	310	$K_1 = 8.3 \times 10^6$ $K_2 = 3.5 \times 10^7$ $K = 1.7 \times 10^7$	17
AATTGCGCAATT	$K_1 = 4.8 \times 10^6$ $K_2 = 6.2 \times 10^7$ $K = 1.7 \times 10^7$	52	$K_1 = 7.3 \times 10^5$ $K_2 = 2.5 \times 10^6$ $K = 1.3 \times 10^6$	13
AATTGCGCGAATT	$K_1 = 1.2 \times 10^7$ $K_2 = 6.0 \times 10^7$ $K = 2.6 \times 10^7$	21	$K < 10^5$	—
AATTGCGCGCAATT	$K_1 = 1.2 \times 10^7$ $K_2 = 5.1 \times 10^7$ $K = 2.4 \times 10^7$	17	$K < 10^5$	—

^a $K = (K_1 K_2)^{0.5}$. ^bCF (Cooperativity Factor) = $(K_2/K_1) \times 4$.

affinity, stoichiometry, and cooperativity of compound interactions with immobilized AATT and AATTGCAATT sequences (Figure 2).^{40,41,43} Essentially, the same moles of the DNA oligomers were immobilized on the surface of each sensor chip so that the sensorgram saturation levels can be compared directly for stoichiometry. The differences in interaction strength and binding stoichiometry for four example compounds are visualized in a plot of r (moles of compound bound/mol of hairpin DNA) versus C_{free} , the free compound concentration (Figure 2B). The plots were best fit with a one-site binding model for AATT and a two-site binding model for AATTGCAATT (K values are summarized in Table 1).

The relative values of the macroscopic equilibrium constants, K_1 and K_2 , also reflect the cooperativity of the interaction. A cooperativity factor to assess the degree of cooperativity is defined as $CF = (K_2/K_1) \times 4$. CF is equal to 1 for interactions with no cooperativity, greater than 1 for positive cooperativity, and less than 1 for negative cooperativity. As can be seen in Figure 2 and Table 1, DB184 binds AATT as a monomer and AATTGCAATT at two sites without cooperativity, as expected. In a similar manner, RT533 binds AATT very strongly as a monomer, while it binds AATTGCAATT as a 2:1 complex with no significant cooperativity. Surprisingly, the linked

compound, RT546, binds in a completely different manner, as a weak monomer to AATT, but as a strong dimer with positive cooperativity to AATTGCAATT (CF > 3000). DB2119, with a more rigid linker, does not bind to AATT while it also binds to AATTGCAATT as a dimer with positive cooperativity, as with RT546. These SPR results are supported by thermal melting studies. DB184 has a very small increase in T_m in agreement with its weak binding by SPR. Both RT546 and DB 2119 have biphasic curves in agreement with their positive binding cooperativity. RT533 has a monophasic curve in agreement with its lack of binding cooperativity (Supporting Information, Figure S1). T_m curves for all of the two-site compounds do not have an observable high temperature baseline and the T_m values cannot be determined accurately. It is clear, however, that the T_m increases scale in the same general order as binding constants.

SPR sensorgrams were also obtained for two additional DNAs; a sequence of similar length without a GC linker, A₅T₅, and a sequence with one AATT binding site mutated to AGTC, AATTGCAGTC (Figure S2, Supporting Information). RT546 and DB2119 bind strongly as monomers to A₅T₅ (Figure S2, Supporting Information). RT533 binds A₅T₅ as a 2:1 complex with negative cooperativity. With AATTGCAGTC, RT546 and

RT533 bind as monomers with binding constants very close to the values for a single site AATT (Table 1 and Figure 2). As with AATT, no binding is observed for DB2119 with AATTGCAGTC (Figure S2, Supporting Information).

In additional studies with DNA sequences that have an extended GC linker length, it was observed that the linker length has no significant effect on binding of the control compounds, DB184 and RT533 (Table 1). As the GC length increases, RT546 still forms a 2:1 complex but with decreasing cooperativity. Figure S3 (Supporting Information) shows that both the association and dissociation rates increase for RT546 binding with 0 to 3 GC. DB2119, however, loses significant binding affinity and cooperativity with increasing GC length and has no detectable binding for the 5 and 6 GC base pair spacer (Table 1).

To investigate how the linker length in RT546 affects the DNA recognition pattern, two longer compounds with one (DB2115) and two (DB2114) additional $-\text{CH}_2-$ groups were synthesized. For the linker $[-\text{O}-(\text{CH}_2)_n-\text{O}-]$, the binding affinity and cooperativity for DNA sequences can vary by a surprisingly large amount as n increases from 3 to 5 (Figure 3, Table S1,

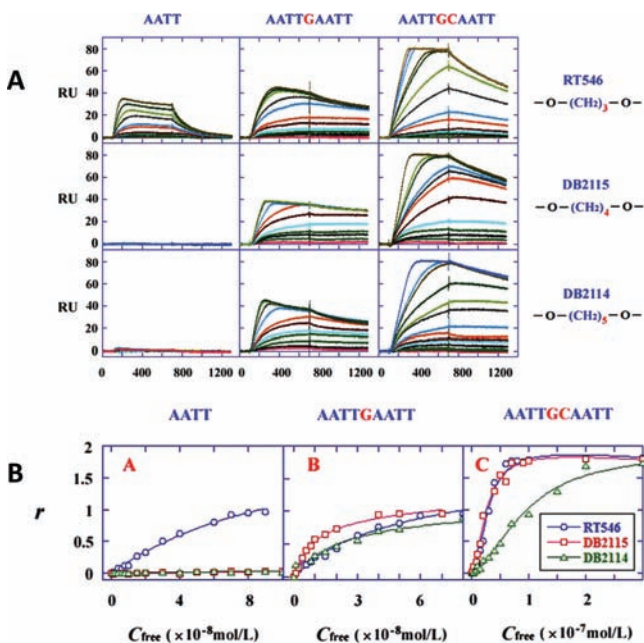


Figure 3. The effect of compounds linker length changes on DNA binding: (A) SPR sensorgrams for the interaction of DNA sequences (AATT, AATTGAATT, and AATTGCAATT) binding with RT546 $[-\text{O}-(\text{CH}_2)_3-\text{O}-]$, DB2115 $[-\text{O}-(\text{CH}_2)_4-\text{O}-]$, and DB2114 $[-\text{O}-(\text{CH}_2)_5-\text{O}-]$. (B) Comparison of the SPR binding affinity for DNA sequences binding with compounds with different linker length. The lines are the best fit values to a single site or two site interaction models.

Supporting Information). RT546, $n = 3$, binds AATT as a weak monomer, for example, while no binding is observed for the longer compounds, DB2115 and DB2114 ($n = 4$ or 5). With AATTGAATT, all three compounds bind strongly as monomers and with AATTGCAATT all three bind as cooperative 2:1 dimers. As can be seen in Figure 3B and Table S1, the three compounds slightly lose binding affinity for AATTGCAATT when the linker length is increased and the cooperativity for binding decreases significantly (about 100 times) as n increases from 3 to 5. Apparently, RT546 has an optimum linker length in this general structure and has the

highest binding affinity and cooperativity for a 2:1 complex with AATTGCAATT.

Electrospray Ionization Mass Spectrometry (ESI–MS): Characterizing the Binding Stoichiometry and Cooperativity.

ESI–MS experiments allow the resolution of complex mixtures and determination of stoichiometries and any binding cooperativity for complexes that are present simultaneously in an injected sample.^{44,45} Figure S4A (Supporting Information) shows ESI–MS spectra of AATT with the three linked compounds RT546, RT533, and DB2119 at different compound to DNA ratios. A high intensity 1:1 complex is detected for RT533, only a low intensity 1:1 complex peak is observed for RT546, and no peak is observed for DB2119. Figure S4B (Supporting Information) shows ESI–MS spectra of AATTGAATT with the same compounds. Peaks for both a 1:1 (labeled as blue) and a 2:1 complex (labeled as red) can be observed for RT533, even at low compound to DNA ratios. As the ratio is increased, the intensity of the 2:1 complex peak increases, but no higher order complexes are observed. For RT546 and DB2119, only a 1:1 complex peak is observed as the ratio increases, indicating a monomer complex. ESI–MS results for AATTGCAATT with the three compounds RT546, RT533, and DB2119 are summarized in Figure S4C (Supporting Information). For RT546, the peak for the 2:1 complex is much larger than the 1:1 complex peak indicating strong cooperative dimer formation. With RT533, the peak for the 1:1 complex is much larger than the peak for the 2:1 complex at low compound to DNA ratios. With increasing ratio, the peak for the 2:1 complex increases, and the 1:1 complex peak decreases. This clearly shows that RT533 binds to AATTGCAATT as a 2:1 complex with no cooperativity. For DB2119, only the peak of the 2:1 complex is detected as the compound to DNA ratio increases, indicating that DB2119 binds to AATTGCAATT as a dimer with very strong positive cooperativity. To visualize the different stoichiometries for AATTGCAATT and AATT binding with the three linked compounds more easily, ESI–MS spectra of three DNA sequences with RT546, RT533, and DB2119 at a 2:1 ratio (compound to DNA) are compared in Figure 4. All of these results are in excellent agreement with the findings from SPR.

Circular Dichroism (CD): Probing the Binding Mode and Binding Ratio.

CD spectra monitor the asymmetric environment of the compounds when bound to DNA and therefore can be used to obtain information on the binding mode.⁵¹ Free RT546 does not exhibit CD signals; however, upon addition of the compound to AATTGCAATT, substantial, positive induced CD signals arise in the compound absorption region between 300 and 400 nm (Figure S5A, Supporting Information). These positive induced CD signals for the complexes are characteristic patterns for minor groove binding in AT sequences.⁵¹ CD changes are significantly smaller for AATT than for AATTGCAATT. CD spectra for the complexes of RT533, RT546, and DB2119 with AATTGCAATT sequence are compared in Figure S5B (Supporting Information) and display large positive induced CD spectral changes, as expected, for minor groove complexes. For all three compounds, the saturation maximum is 2 compounds/hairpin for AATTGCAATT, and these values are consistent with the results from SPR experiments. The induced CD signals for DB2119 with a more rigid linker are larger than for the more flexible compounds, suggesting a stronger dipole interaction with the bases in this compound.⁵¹

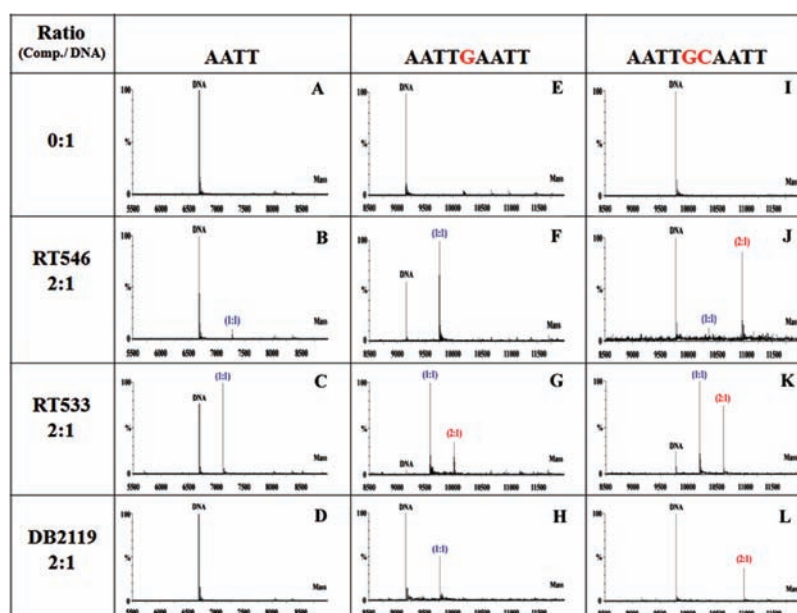


Figure 4. Comparison of ESI–MS spectra of DNA sequences with RT546, RT533, and DB2119 at a 2:1 ratio (compound to DNA): (A–D) the AATT DNA sequence; (E–H) the AATTGAATT DNA sequence; (I–L) the AATTGCAATT DNA sequence. More detailed MS spectra are in Supporting Information.

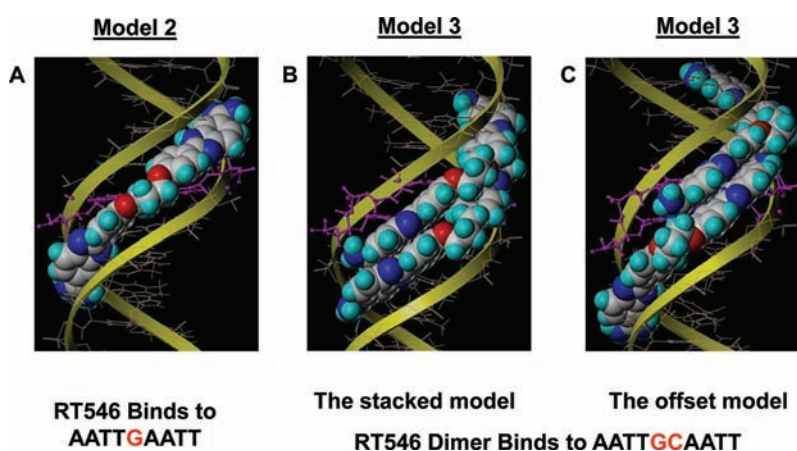


Figure 5. Models for RT546 docked into the (A) AATTGAATT site, Model 2 in Figure 6; and (B and C) the AATTGCAATT sequence (see Model 3 in Figure 6). The complex in (B) is highly overlapped while in (C) the complex is offset. For clarity, the terminal bases are not displayed in the images, and the images are of only the lowest-energy conformation. The DNA backbone is shown by yellow ribbons and the GC base pairs are displayed by ball and stick type in magenta, while the ligand is displayed as spacefill type in CPK colors. More detailed modeling Figures of the models are in Supporting Information, Figure S6.

Molecular Docking. To provide ideas for a better understanding of the two site binding modes of symmetric compounds, RT546 was docked into the two AATT sites with G and GC spacing sequences (Figure 5) as described in the Materials and Methods (additional views of the models are shown in Figure S6). The monomer docking at AATTGAATT is similar to classical minor groove binding compounds (Figure 5A). The groove is relatively narrow along the entire two site length. The ABP modules fit nicely into the AATT sites with amidine and benzimidazole H-bonds to the floor of the groove. The trimethylene linker binds into the groove at the GC base pair as expected. Although minor variations of this model are possible, and certainly occur on a dynamic basis, any major movement of RT546 in this site results in a significant loss of binding interactions and is unlikely.

For the AATTGCAATT site, two different models with favorable interactions for the RT546 2:1 complex are presented. In Figure 5B, the stacked molecules are slightly offset such that the amidine of the top molecule of the figure stacks over the benzimidazole of the lower molecule. At the other end of the dimer, the stacking is reversed. This allows amidine and benzimidazole –NHs of the lower molecule to H-bond to AT base pairs at the floor of the groove, but the imidazole –NH and phenyl of the upper molecule are pulled into the GC region of the sequence. The alkyl group of the upper molecule is at the top GC base pair in the site. For the lower molecule, the stacking is reversed. All four amidine groups in the stacked dimer are able to H-bond to AT base edges at the floor of the groove. The central six base pairs of the minor groove in the AATTGCAATT sequence must be widened to accommodate the stacked dimer. The groove then rapidly narrows to the

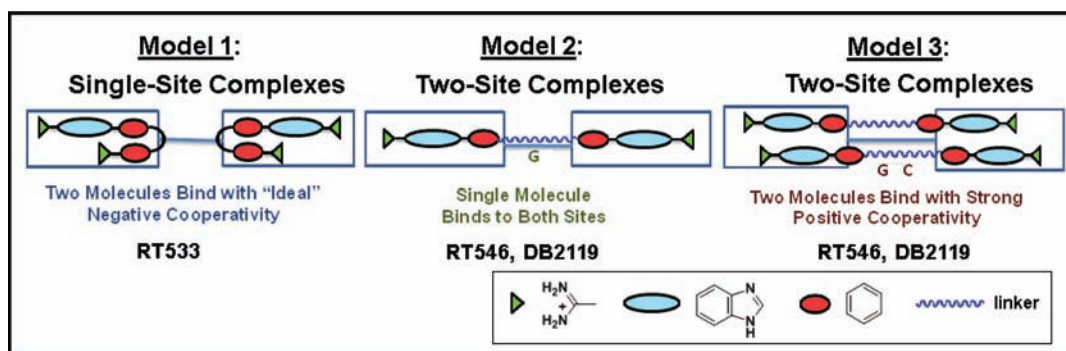


Figure 6. Schematic representation of three different two-site binding modes for the compounds of Figure 1. Model 1: single-site complexes, two molecules bind with “ideal” negative cooperativity. Model 2: two-site complexes, single molecules bind to the entire sequence. Model 3: two-site complexes, two molecules bind with strong positive cooperativity to the entire sequence. The coloring scheme is green triangles for amidine groups, light blue ovals for benzimidazole groups, red ovals for phenyl groups, and dark blue lines for the linker.

normal B-form width at the ends of the AT binding site and flanking sequences.

In Figure 5C, the AATT regions of groove are narrow, as in all known structures, while the groove widens at the GC site, also as expected. To accommodate two stacked RT546 molecules into this asymmetric groove structure, the compound stacking must be significantly offset. This moves the amidines far apart and reduces electrostatic repulsion. It allows stacking of two benzimidazole-phenyl units in the wide GC groove. Both amidine and benzimidazole $-NH$ s in the GC center point toward the GN3 (amidine) or C keto (benzimidazole NH) groups and are within H-bonding distance. The phenyl oxygens of these stacked systems are rotated such that the methylene and remaining ABP units are brought to the center of the AATT sites. The amidine and benzimidazole $-NH$ s that are not in the GC sequence H-bond to the floor of the groove at AN3 and T keto groups. The central four base pairs have a wide groove to accommodate the stacked modules and the groove then quickly narrows to contact the remaining reactions of the compounds in the AT sequences of the offset dimer. The two stacked dimer complexes at the AATTGCAATT sequence were docked to provide the best interactions with the DNA sequence. Movement of the compounds away from these positions results in a significant loss of H-bonding and is unlikely. To optimize H-bonding with base pair edges at the floor of the groove, the compounds in the dimer must be moved in specific jumps from interactions with one base pair to the next. This restriction means that there are very few optimized dimer positions in the two site sequences and the two shown have the best energetics.

DISCUSSION

The six compounds in Figure 1 were designed to probe the types of linked structures that can bind strongly to the minor groove in adjacent AT sites of the kind found in the DNA of parasitic microorganisms. The results provide important and somewhat surprising information on how compound molecular structure affects binding to complex, closely related DNA sequences. The fundamental unit in the compound design is an ABP, amidine-benzimidazole-phenyl, motif and all compounds have at least one copy of the motif (Figure 1, Tables 1 and S1). As expected from their shape, chemical groups, and charge, CD studies show that all of these compounds bind to the DNA minor groove.⁵¹ The results of biosensor-SPR and mass spectrometry experiments with oligomer DNA models for

two-site kinetoplast DNA binding sites, however, clearly reveal that the different compounds have quite different structures in complex with DNA.

The simple control compound, DB184, binds relatively weakly to all of the DNAs. It binds to one site with the AATT DNA and to two sites with no cooperativity with the two site DNAs by all methods, as expected (Figure 2). This compound, thus, has essentially “ideal” independent, multisite binding properties⁴² and is a useful control for comparison with the more complex results for the other compounds (Figure 2). With the linked compounds, a surprisingly large variation in interactions with the model DNAs of Figure 1 is obtained. The analysis with different DNAs makes it possible to dissect the binding modes of the single ABP module control compound, RT533. It binds to AATT as a monomer, about 100 times stronger ($K > 10^8 \text{ M}^{-1}$) than DB184, suggesting that the phenylamidine can fold back to stack on the benzimidazole-phenyl module with favorable interactions in a single GAATTC site. A schematic model to illustrate this binding mode is shown in Figure 6 (Model 1). We suggest that the stacked part of the complex will be closer to the wider GC minor groove and will be quite dynamic. While the compound may be able to fit the AATT site without stacking, this would require unfavorable conformational changes in the linker, and given the strong binding, such a complex seems unlikely. RT533 binds to all of the two site sequences as a strong 2:1 complex indicating that the compound can bind to individual AATT sites in the two site sequences in much the same way as with the single AATT. It has significant negative cooperativity in the interaction with the closely spaced sites in AAAAATTTTT and AATTGAATT, perhaps due to electrostatic repulsion and unfavorable steric effects. The negative cooperativity decreases with the GC spacer length increase in AATTGCAATT and no significant cooperativity is observed for the DNAs with longer GC sequences. On the whole, the compound binds similarly to DB184 except that it has larger K values.

The symmetric compounds have two ABP modules and they exhibit very different binding modes with the different DNA sequences. RT546, which has an $O-(CH_2)_3-O$ linker, is the simplest of these compounds and in contrast to RT533 binds to the single AATT DNAs in a relatively weak 1:1 complex. The restricted conformational space of this compound does not allow it, or any other symmetric ABP derivative, to fold into a favorable complex with single AATT sites. RT546, however, binds to AATTGAATT as a strong 1:1 complex, again in

contrast to RT533 which binds as a 2:1 complex. These results again indicate that RT546 is too large to fold back and bind to a single AATT site, as in Figure 6, Model 1. The results indicate, however, that it can extend its linker through the minor groove at a GC base pair to bind effectively as a monomer to both AATT sites in the AATTGAATT sequence, as in Model 2 of Figure 6. The compounds were designed to bind in this manner, ABP units at AATT and the linker at the GC base pair, and the observed complex is quite satisfying.

To provide a molecular model to help visualize the interactions in the 1:1 complex, RT546 was docked into the AATTGAATT sequence (Figure 5A). The docked model for the interaction of RT546 with the single GC sequence is similar to classical minor groove complexes at pure AT sites. The single GC between the narrow AATT units does not significantly widen the minor groove but forms a binding site for the somewhat more bulky methylene groups. The lack of widening of the minor groove by a single GC in A-tract sequences may be a common feature of DNA complexes. A similar lack of groove widening by a single GC was noted with the Fis transcription factor–DNA complex while insertion of three GC base pairs did widen the groove with substantial loss of Fis binding affinity.⁵² There are limited favorable interactions in this GC part of the complex in Figure 5A and current design efforts are focused on possible GC recognition units to improve the interaction. The ABP modules fit into the AATT sites with van der Waals, H-bonding, and water displacement contributions to binding affinity. The modules do not, however, completely fill the length of the AATT site and it is clear that the affinity for the AATT sequences in AATTGAATT can be also improved.

Analysis of the binding results with AATTGCAATT provided an unexpected surprise. The expected result was either a 1:1 complex whose strength depended on how the linking chain could fit into the groove at GC, or no complex if the linker did not fit. Instead, we observed a highly cooperative 2:1 complex. Complexes at the same ratio but with decreasing positive cooperativity were observed for the other DNAs in Table 1. A schematic model to explain these results is shown in Figure 6 (Model 3). The stacked dimer in Model 3 is an entirely new binding mode for dicationic minor groove compounds that bind in A-tract type sequences. In the molecular docking results in Figure 5, two models for the 2:1 complex of RT546 at AATTGCAATT are described. The results show that the minor groove with two GC base pairs between AATT widens such that a stacked complex of RT546 is more energetically favorable than the monomer binding with a single GC. This widening also agrees with the Fis–DNA results for more than one GC between AT sequences.⁵² The two complexes for RT546 are both able to form a significant number of favorable interactions with base edges at the floor of the minor groove as well as cation- π interactions in the stacked dimer. The offset model of Figure 5C appears more favorable than the overlapped model in Figure 5B for two primary reasons: (i) the charged amidines are close in Figure 5B and this will give significant electrostatic repulsion; and (ii) to accommodate the model in Figure 5B, the AATT minor groove must widen to bind the stacked dimer units but this sequence strongly resists dimer complexes in other systems. Wemmer and co-workers, for example, observed monomer binding in AATT and only found the distamycin dimer in longer sequences of A-tract type structures.⁵³ Monomer binding is observed in all AATT minor groove complexes in the protein

database. The model in Figure 5C retains a narrow groove in the AATT sites but has the expected wider GC groove. The models again suggest that both GC and A-site recognition can be significantly improved. The AT recognition part of the complex does not completely fill the site and expansion of the ABP unit can improve these interactions. For stacked dimer recognition of the GC spacer, we must make nonsymmetrical molecules, as with expanded RT533 analogues that have an H bond acceptor for interaction with the G-NH₂. The other molecule of the dimer should have a donor for interaction with the complementary C=O group. Initial examples of such molecules are in preparation.

Another important observation with RT546 is that as the GC linker length increases, K_1 for the cooperative complex increases while K_2 decreases. As a result, the overall binding constant ($K = (K_1 \times K_2)^{1/2}$) does not significantly change but the cooperativity factor ($CF = K_2/K_1 \times 4$) decreases. This is probably caused by decreased stacking of the ABP modules with decreased van der Waals interactions (K_2 decrease), but favorable, decreased amidine electrostatic repulsion (K_1 increase). Adjustments to optimize binding energetics are also possible in the alkyl linking chain.

It is somewhat surprising that in the symmetric compounds, changing the length of the linking chains generally causes a fairly small difference in DNA affinity (Tables 1 and S1). With the AATTGAATT sequence and the alkyl linked compounds, for example, the five methylene compound, DB2114, has slightly weaker binding than the four and three methylene derivatives (Table S1). In contrast, the more rigid five carbon linker in DB2119, with a central phenyl, causes a 10-fold decrease in binding affinity. This finding suggests that the length and/or flexibility of the phenyl linker are not optimum for binding with a single GC base pair spacer. With the AATTGCAATT sequence, DB2119 binds much more strongly and clearly has a better length for this sequence. It binds similarly to the three GC spacer DNA but the binding constant decreases markedly for the longer DNAs in contrast to the flexible linker compounds. The phenyl linker has much larger restrictions on its two site interactions, but with appropriate optimization, this could provide a powerful ability to target specific DNA sites, especially since this compound does not significantly bind to single AATT sites. The flexible compound RT546 has much more similar affinities for all of the DNA GC spacer lengths and can thus adapt its cooperative dimer formation to a range of sequences. Such compounds might be best for targeting kinetoplast DNA where binding to multiple two-site sequences is an advantage.

There are two key points to consider in evaluation of the interaction of these first-generation compounds with two closely spaced AT sites in terms of designing the second generation of linked modules for binding to two closely spaced AT sites. First, longer flexible linkers (as in DB2114) have weaker binding probably due to loss of conformations and entropy on forming a complex. Such linkers should be avoided when possible. Second, the current linkers in the ABP compounds do not have favorable GC interactions and serve primarily to let the two ABP modules fit into the AT sites. Optimization of compounds with appropriate linker lengths that have specific GC recognition motifs is, thus, a requirement for second generation compounds. Synthesis of second generation compounds, based on these design principles, is in progress and should provide new agents with both increased affinity and specificity.

In summary, linked two-site binding compounds such as those in Figure 1 can have three major, different types of complexes with A-tract sequences in DNA. There are two particularly unexpected features of the results. First is the switch from monomer binding by the compounds with two ABP motifs, when there is a single GC between AATT target sites, to dimer binding with a two GC spacer in the DNA sequence. Clearly, the widening of the groove with the two GCs allows the compound to improve the total free energy of binding by inserting two stacked molecules into the binding site. Second, the AATT site has typically maintained a narrow, A-tract type minor groove, and a cooperative minor groove dimer has not previously been observed for this sequence. This suggests that the wider central GC sequence may allow compound module stacking while maintaining a relatively narrow minor groove with only single compound modules inserted, as in Model 3 in Figure 5C. The compounds in this research have provided exciting new information on recognition of adjacent AT binding sites as well as numerous suggestions for design of second generation agents that have improved affinity and selectivity.

■ ASSOCIATED CONTENT

● Supporting Information

Synthesis and characterization of symmetric compounds; figures for comparison of thermal melting; SPR experiments for A₅T₅ and AATTGCAGTC binding; the effect of GC length change on RT546 binding; ESI-MS spectra; comparison of CD spectra; effect of compound linker length on the DNA recognition pattern. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

wdw@gsu.edu

Notes

The authors declare no competing financial interest.

[§]On leave from Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

■ ACKNOWLEDGMENTS

This research was supported by National Institutes of Health grant AI064200. Biacore instruments were purchased with partial support from the Georgia Research Alliance.

■ REFERENCES

- (1) Mohammed, H. S.; Delos Santos, J. O.; Armitage, B. A. *Artif. DNA PNA XNA* **2011**, *2*, 43.
- (2) Hu, J.; Corey, D. R. *Biochemistry* **2007**, *46*, 7581.
- (3) Højfeldt, J. W.; Van Dyke, A. R.; Mapp, A. K. *Chem. Soc. Rev.* **2011**, *40*, 4286.
- (4) Tietjen, J. R.; Donato, L. J.; Bhimisaria, D.; Ansari, A. Z. *Methods Enzymol.* **2011**, *497*, 3.
- (5) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521.
- (6) He, Y.; Hoskins, J. M.; McLeod, H. L. *Trends Mol. Med.* **2011**, *17*, 244.
- (7) Kissinger, J. C. *Trends Parasitol.* **2006**, *22*, 240.
- (8) Mu, J.; Seydel, K. B.; Bates, A.; Su, X. Z. *Curr. Genomics* **2010**, *11*, 279.
- (9) Purfield, A. E.; Tidwell, R. R.; Meshnick, S. R. *Malar. J.* **2009**, *8*, 104.

- (10) Franks, A.; Tronrud, C.; Kiakos, K.; Kluza, J.; Munde, M.; Brown, T.; Mackay, H.; Wilson, W. D.; Hochhauser, D.; Hartley, J. A.; Lee, M. *Bioorg. Med. Chem.* **2010**, *18*, 5553.
- (11) Wilson, W. D.; Tanious, F. A.; Mathis, A.; Tevis, D.; Hall, J. E.; Boykin, D. W. *Biochimie* **2008**, *90*, 999.
- (12) Brucoli, F.; Hawkins, R. M.; James, C. H.; Wells, G.; Jenkins, T. C.; Ellis, T.; Hartley, J. A.; Howard, P. W.; Thurston, D. E. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3780.
- (13) Ryan, B.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J.M.. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11497.
- (14) Hebert, M. D. *Biochimie* **2008**, *90*, 1131.
- (15) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261.
- (16) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593.
- (17) Collie, G. W.; Sparapani, S.; Parkinson, G. N.; Neidle, S. *J. Am. Chem. Soc.* **2011**, *133*, 2721.
- (18) Koirala, D.; Dhakal, Soma; Ashbridae, B.; Sannohe, Yuta; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S. *Nat. Chem.* **2011**, *3*, 782.
- (19) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L. *Biochimie* **2008**, *90*, 131.
- (20) Dervan, P. B. *Science* **1986**, *232*, 464.
- (21) Roy Chowdhury, A.; Bakshi, R.; Wang, J.; Yildirim, G.; Liu, B. Y.; Pappas-Brown, V.; Tolun, G.; Griffith, J. D.; Shapiro, T. A.; Jensen, R. E.; Englund, P. T. *PLoS Pathog.* **2010**, *6*, e1001226.
- (22) Motta, M. C. *Curr. Pharm. Des.* **2008**, *14*, 847.
- (23) González, M.; Cerecetto, H. *Expert Opin. Ther. Pat.* **2011**, *21*, 699.
- (24) Lukeš, J.; Guilbride, D. L.; Votýpka, J.; Zíková, A.; Benne, R.; Englund, R. T. *Eukaryotic Cell* **2002**, *1*, 495–502.
- (25) Paine, M. F.; Wang, M. Z.; Generaux, C. N.; Boykin, D. W.; Wilson, W. D.; De Koning, H. P.; Olson, C. A.; Pohlig, G.; Burri, C.; Brun, R.; Murilla, G. A.; Thuita, J. K.; Barrett, M. P.; Tidwell, R. R. *Curr. Opin. Invest. Drugs* **2010**, *11*, 876.
- (26) Le Loup, G.; Pialoux, G.; Lescure, F. X. *Curr. Opin. Infect. Dis.* **2011**, *24*, 428.
- (27) Brun, R.; Don, R.; Jacobs, R. T.; Wang, M. Z.; Barrett, M. P. *Future Microbiol.* **2011**, *6*, 677.
- (28) Corson, T. W.; Aberle, N.; Crews, C. M. *ACS Chem. Biol.* **2008**, *3*, 677.
- (29) Bourdouxhe-Housiaux, C.; Colson, P.; Houssier, C.; Waring, M. J.; Bailly, C. *Biochemistry* **1996**, *35*, 4251.
- (30) David-Cordonnier, M. H.; Hildebrand, M. P.; Baldeyrou, B.; Lansiaux, A.; Keuser, C.; Benzschawel, K.; Lemster, T.; Pindur, U. *Eur. J. Med. Chem.* **2007**, *42*, 752.
- (31) Zhang, R.; Wu, X.; Guziec, L. J.; Guziec, F. S.; Chee, G. L.; Yalowich, J. C.; Hasinoff, B. B. *Bioorg. Med. Chem.* **2010**, *18*, 3974.
- (32) Wakelin, L. P. G. *Med. Res. Rev.* **1986**, *6*, 275.
- (33) Holman, G. G.; Zewail-Foote, M.; Smith, A. R.; Johnson, K. A.; Iverson, B. L. *Nat. Chem.* **2011**, *3*, 875.
- (34) Zolova, O. E.; Mady, A. S. A.; Garneau-Tsodikova, S. *Biopolymers* **2010**, *93*, 777.
- (35) Fox, K. R.; Gauvreau, D.; Goodwin, D. C.; Waring, M. J. *Biochem. J.* **1980**, *191*, 729.
- (36) Rahman, K. M.; James, C. H.; Thurston, D. E. *Nucleic Acids Res.* **2011**, *39*, 5800.
- (37) Kers, I.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3339.
- (38) Czarny, A.; Wilson, W. D.; Boykin, D. W. *J. Heterocycl. Chem.* **1996**, *33*, 1393.
- (39) Tidwell, R. R.; Geratz, J. D.; Dann, O.; Volz, G.; Zeh, D.; Loewe, H. *J. Med. Chem.* **1978**, *21*, 613–623.
- (40) Liu, Y.; Wilson, W. D. *Methods Mol. Biol.* **2010**, *613*, 1.
- (41) Nguyen, B.; Tanious, F. A.; Wilson, W. D. *Methods* **2007**, *42*, 150.
- (42) Van Holde, K. E.; Johnson, W. C.; Ho, P. S. In *Principles of Physical Biochemistry*; Pearson Prentice Hall: Upper Saddle River, NJ, 2006.

- (43) Nanjunda, R.; Munde, M.; Liu, Y.; Wilson, W. D. In *Methods For Studying DNA/Drug Interactions*; CRC Press-Taylor & Francis Group: Boca Raton, FL, 2011.
- (44) Rosu, F.; De Pauw, E.; Gabelica, V. *Biochimie* **2008**, *90*, 1074.
- (45) Lombardo, C. M.; Martínez, I. S.; Haider, S.; Gabelica, S. V.; Pauw, E. D.; Moses, J. E.; Neidle, S. *Chem. Commun.* **2010**, *46*, 9116.
- (46) SYBYL *Molecular Modeling Software*, Version XI.2; Tripos, Inc.: St. Louis, MO, 2010.
- (47) Collar, C. J.; Lee, L.; Wilson, W. D. *J. Chem. Inf. Model.* **2010**, *50*, 1611.
- (48) Holt, P. A.; Chaires, J. B.; Trent, J. O. *J. Chem. Inf. Model.* **2008**, *48*, 1602.
- (49) Bates, P. J.; Laber, D. A.; Miller, D. M.; Thomas, S. D.; Trent, J. O. *Exp. Mol. Pathol.* **2009**, *86*, 151.
- (50) Liu, Y.; Collar, C. J.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D. *J. Phys. Chem. B.* **2008**, *112*, 11809.
- (51) Ardhammar, M.; Norden, B.; Kurucsev, T. In *DNA-Drug interactions*; Wiley: New York, 2000.
- (52) Stella, S.; Cascio, D.; Johnson, R. C. *Genes Dev.* **2010**, *24*, 814.
- (53) Wemmer, D. E.; Geierstanger, B. H.; Fagan, P. A.; Dwyer, T. J.; Jacobsen, J. P.; Pelton, J. G.; Ball, G. E.; Leheny, A. R.; Chang, W. H.; Bathini, Y.; Lown, J. W.; Rentzeperis, D.; Marky, L. A.; Singh, S.; Kollman, P. In *Structural Biology: The State of the Art*; Adenine Press: Albany, NY, 1993.