Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c6ob02390h

Systematic synthetic and biophysical development of mixed sequence DNA binding agents†

Ananya Paul,[a] Arvind Kumar,[a] Rupesh Nanjunda,[a,b] Abdelbasset A. Farahat,[a,c] David W. Boykin[a] and W. David Wilson*[a]

Received 2nd November 2016,
Accepted 13th December 2016

DOI: 10.1039/c6ob02390h

www.rsc.org/obc

It is now well established that, although only about 5% of the human genome codes for protein, most of the DNA has some function, such as synthesis of specific, functional RNAs and/or control of gene expression. These functional sequences open immense possibilities in both biotechnology and therapeutics for the use of cell-permeable, small molecules that can bind mixed-base pair sequences of DNA for regulation of genomic functions. Unfortunately very few types of modules have been designed to recognize mixed DNA sequences and for progress in targeting specific genes, it is essential to have additional classes of compounds. Compounds that can be rationally designed from established modules and which can bind strongly to mixed base pair DNA sequences are especially attractive. Based on extensive experience in design of minor-groove agents for AT recognition, a small library of compounds with two AT specific binding modules, connected through linkers which can recognize the G·C base pairs, were prepared. The compound-DNA interactions were evaluated with a powerful array of biophysical methods and the results show that some pyridyl-linked compounds bind with the target sequence with sub-nanomolar K_D, with very slow dissociation kinetics and 200 times selectivity over the related sequence without a G·C base pair. Interestingly, a set of compounds with AT module connected by different linkers shows cooperative dimer recognition of related sequences. This type of design approach can be expanded to additional modules for recognition of a wide variety of sequences.

Introduction

As part of a long-term project to prepare an entirely new class of DNA minor groove binding agents that can recognize mixed AT and G·C base pair (bp) sequences of DNA, a variety of modular units that can be combined in different ways to recognize specific DNA sequences are being developed. A major goal of this project is to prepare compounds that can inhibit specific transcription factors and control cellular gene expression by binding to the DNA minor groove.[1–5] In *in vitro* applications such specific recognition compounds will allow new diagnostic applications while *in vivo* they could provide new therapeutics and gene-specific probes.[6–12] The goal to modulate transcription factor activity has been of significant interest, but has been quite difficult to reach and transcription factors have been described as "undruggable".[13–15] Only polyamides have had significant success in systematic, mixed-sequence DNA recognition and transcription factor inhibition.[16–20] This is in spite of significant efforts in chemical biology for over a decade to find methods to systematically modulate cellular transcription factor activity. Complementing the successful polyamides with compounds that have different recognition modules with enhanced cellular uptake and improved solution properties would be an important major advance in the effort to target specific protein binding to nuclear DNA.[1,7,21–24]

Because of their synthetic accessibility,[25,26] favorable cell uptake,[7,9] demonstrated clinical activity,[27–29] and excellent solution properties,[30] our initial design efforts for new types of mixed-sequence DNA binding compounds have focused on heterocyclic cations.[26,31,32] One design concept includes three parts, two AT sequence recognition modules that flank a central G·C bp recognition unit. This is a common sequence in the mitochondrial kinetoplast DNA of a number of parasitic organisms as well as in human gene promoter sequences.[33,34] We recently reported that DB2120 with a central pyridine linker unit flanked by phenyl-benzimidazole-based modules for AT recognition (Fig. 1) could strongly and selectively recog-

[a]*Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30303-3083, USA. E-mail: wdw@gsu.edu; Fax: +1 404-413-5505; Tel: +1 404-413-5503*
[b]*Janssen Research and Development, 1400 McKean Rd, Spring House, PA 19477, USA*
[c]*Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt*

†Electronic supplementary information (ESI) available: Compound synthesis, and characterization data used in this article, supporting figures (Fig. S1–S3 and Tables S1, S2). See DOI: 10.1039/c6ob02390h

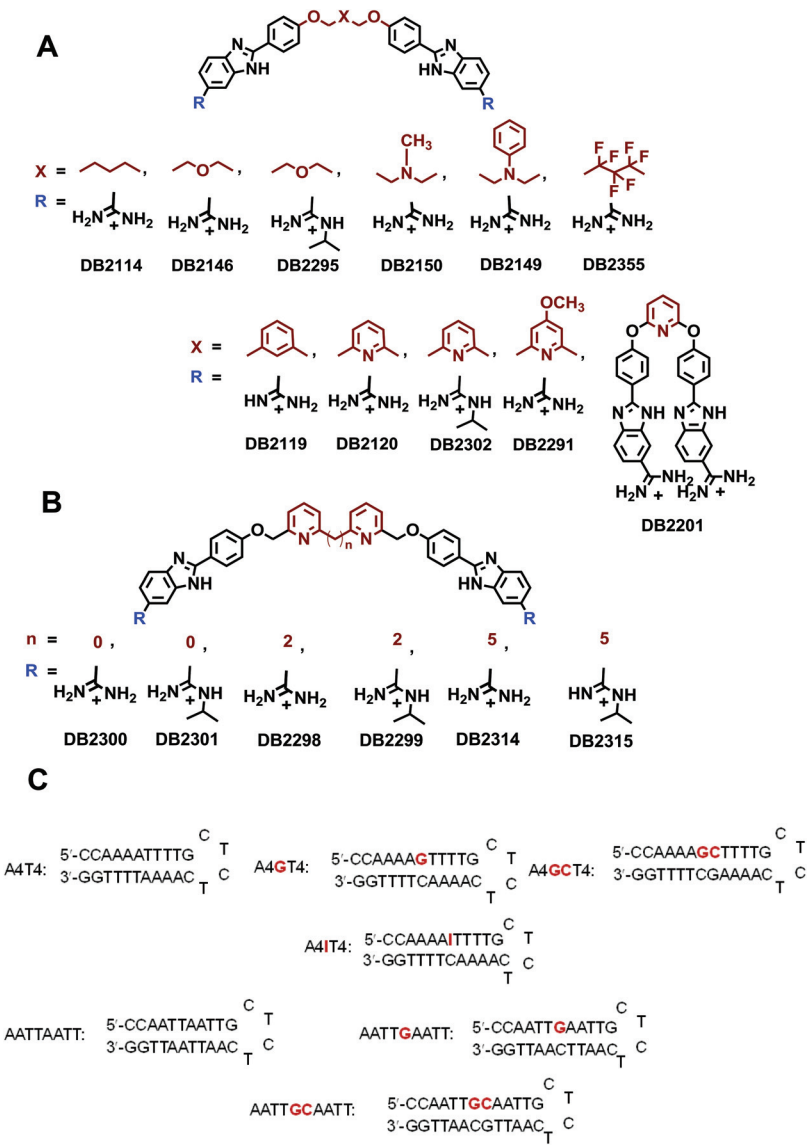

Fig. 1 The structure of compounds and DNA sequences used in this study. For SPR experiments, 5′-biotin labeled DNA sequences are used for SPR studies.

nize a single G base in a G·C bp flanked by AT sequences with an nM affinity.[31] It is an important new design finding and complements other new minor groove binders that can also target mixed sequences of DNA.[35,36] These compounds provide strong evidence that new heterocyclic derivatives, with appropriate positioning of H-bond acceptors to interact with the G amino –NH which projects into the minor groove, can be successfully designed. Although our understanding of AT sequence binding by minor groove agents is extensive, the understanding of mixed base pair sequence recognition by minor groove binders is much more limited. This lack of appropriate design models limits the use of the DNA minor groove as a receptor in drug design and development efforts, and it is essential to expand our knowledge of mixed sequence recognition by synthetic heterocycles.

To help overcome this limitation and develop a better understanding of DNA mixed-sequence molecular recognition a small library of new modular compounds, generally focused around DB2120, were prepared (Fig. 1A and B) and tested with a panel of mixed sequence DNAs (Fig. 1C). The primary goal of this work is to develop a detailed understanding of both the DNA sequence and compound properties that lead to strong DNA binding and G·C bp recognition. The new compounds were designed and synthesized based on the three-component, linked-module concept and have H-bond acceptors for interaction with the G-NH in the minor groove and specific G·C recognition (Fig. 1A and B). For specific AT sequence recognition, the compounds have terminal amidine-benzimidazole-phenyl (ABP) motifs that function well in strong binding to pure AT sequences.[26] The ABP modules are linked with alkyl

or alkyl aromatic modules with H-bond acceptors for potential recognition of one or more G·C bps (Fig. 1). Specific goals of these studies were to determine (1) how linker length and rigidity affect the ability of ABP modules to recognize two AT sites separated by G·C bps; (2) how different H-bond acceptors, which can appropriately function in the minor groove, provide specific binding for G·C bps; and (3) how compounds that recognize more than a single G·C bp can be designed. Detailed biophysical studies of the interactions of all compounds with a variety of DNA sequences have been conducted. While many similarities were discovered between the results for recognition of G·C bps in different sequences, there were some surprising differences that we have been able to explain from the sequence-dependent properties of the DNA minor groove. We have been able to rationally design and prepare new compounds that can selectively recognize long, mixed bps DNA sequences. While there remains much to learn, the new results offer clear directions and leads for the development of a novel DNA minor-groove-binding language.

Results and discussion

Compound design and preparation

Previously, symmetric ABP motifs linked through flexible alkyl linkers (DB2114, DB2115, Fig. 1) of variable length have been used with a set of DNA sequences to evaluate length and flexibility effects in recognition of two AT sequences separated by one or more G·C bps.[26] To explore these variations in greater detail the linker variations with the ABP motifs have been expanded. By introducing a pyridyl group connected with and without $-CH_2O-$ moieties as a linker, DB2370, DB2120 and DB2201 were prepared (Fig. 1 and S1†). Because DB2120 is excellent for binding at AT sites separated by a single G·C bp, we synthesized two additional derivatives, one with an isopropylamidino, DB2302 and another with methoxy ($-OCH_3$) substituted pyridyl, DB2291. Detailed binding studies, with different biophysical techniques, were conducted for these compounds and a set of mixed DNA sequences. General syntheses of new compounds (DB2146, DB2295, DB2150, DB2149, DB2355, DB2201, DB2302, DB2291, DB2300, DB2301, DB2298, DB2299, DB2314 and DB2315) are described in the Experimental section and ESI.† The syntheses of DB2114, DB2119 and DB2120 have been previously reported.[26,31]

Thermal melting (T_m): ranking the relative stabilities

To examine the consequences of changing the compound linkers for relative binding to DNA, all newly synthesized compounds were screened qualitatively by thermal melting experiments. The ΔT_m values for DB2114 and DB2119 with AATTGAATT and AATTGCAATT (Table 1) (Fig. S2†) are in good agreement with previously reported values.[26] With the A4T4 and A4GT4 sequences, both compounds showed thermal stabilities that are similar to the AATT sequences (ESI,† Table 1). No significant binding-induced increase in the thermal stability of the compounds with pure A-tract sequences containing two G·C bp was observed. The presence of the additional G·C bp within the A-tracts acts as a strong negative regulator for binding of the compounds in this set. Higher T_m values with A4GCT4 and AATTGCAATT (Table 1), were observed for the *N*-methyl functionalized analog, DB2150, perhaps due to the additional charge on the sp^3 *N*. Replacement of *N*-methyl with an *N*-phenyl in DB2149 caused a large decrease in overall binding affinities for all six DNA sequences. This effect arises from steric hindrance with the minor groove by the phenyl which is rotated out of the plane of the ABP module.

To investigate the behavior of a more hydrophilic linker, the central $-CH_2-$ in DB2114 was replaced with oxygen in DB2146 and in DB2295. Due to the structural similarities (Fig. 1A) with the parent molecule, DB2114, DB2146 and DB2295 both showed similar T_m results with pure or mixed narrow A-tract sequences (Table 1). However, for both ligands, a very interesting increase in thermal stability has been observed for the AATTGCAATT sequence (ΔT_m 12 °C and 8 °C for DB2146 and DB2295 respectively) that is discussed in more detail below.

Other analogs linking two highly AT specific ABP groups by modified aliphatic linkers, DB2355 and DB2150, failed to demonstrate any preference for single G·C bp containing binding sites (Table 1). The lack of sequence specificity with G·C bp containing sequences could be due to the favorable structural orientation of these flexibly linked molecules at the narrow minor groove of A-tract sequences. The decrease of binding affinity by incorporation of a bulky *N*-Ph group in

Table 1 Thermal melting studies (ΔT_m; °C)[a] of aliphatic linker linked bis-benzimidazole compounds with mixed DNA sequences

	DB2114	DB2146	DB2295	DB2150	DB2149	DB2355
A4T4	13	19	17	14	9	19
A4GT4	11	11	16	13	7	16
A4GCT4	1	8	5	4	2	4
AATTAATT	16	18	17	18	8	22
AATTGAATT	10	10	11	9	3	13
AATTGCAATT	5	12	8	10	1	13

[a] $\Delta T_m = T_m$ (the complex) − T_m (the free DNA). The listed values are for 2 : 1 [ligand]/[DNA] ratio and an average of two independent experiments with a reproducibility of ±0.5 °C.

DB2149 indicates not only flexibility but steric crowding by the linker plays very critical roles in DNA minor groove recognition.

The essential breakthrough in developing strong and selective single G·C bp recognition in this series was achieved with the pyridyl derivative of DB2119, DB2120.[31] In our initial report, it was observed that a phenyl (DB2119) to pyridyl (DB2120) change resulted in a striking increase in the thermal stability of the single G·C bp containing A4GT4 motif. On the other hand, the significantly lower thermal stability for the A4T4–DB2120 complex clearly indicates a high sequence-selective binding for DB2120 (Fig. 1A and Table 2).

A major new set of derivatives with two pyridyl groups linked with an alkyl chain (Fig. 1B and S1†), has been developed. These compounds were synthesized with the goal of selective recognition of two G·C bps with flanking AT bps. The flexible linker should allow one pyridyl to H-bond with G-NH_2 on the top strand, while rotation of the alkyl linker should allow the second pyridyl group to H-bond with a G-NH_2 on the bottom strand in a monomer minor groove complex. Perhaps not surprisingly, the compounds without a flexible linker (DB2300 and DB2301) showed poor binding to all sequences (Table S1†). The compounds with n = 2 (DB2298, DB2299) (Fig. 1B) showed moderate binding but, in a disappointing result, no selectivity for the G·C bp DNA sequences and poor selectivity in general. The n = 5 (DB2314 and DB2315) compounds also showed poor binding, probably due to too much flexibility between the pyridine H-bond acceptor groups (Table S1 and Fig. S1†).

In summary, with the compounds in Fig. 1, the highly flexible aliphatic linked ABP molecules DB2114, DB2146, and analogs, show strong DNA minor groove binding with less sequence specificity. The thermal melting results of oxyethylene linked DB2146 and DB2295 with AATTGCAATT indicates that these compounds may form 2 : 1 complex with the wider minor groove of this sequence (ESI,† Table 1). The high T_m results of DB2120 and DB2302 with single G·C bp containing indicate that we have optimized the flexibility and hydrogen bond accepting ability of the pyridyl linker for selective recognition of single G·C bp containing mixed DNA sequences.

Table 2 Thermal melting studies (ΔT_m; °C)[a] of aromatic linker linked bis-benzimidazole compounds with mixed DNA sequences

	DB2201	DB2119	DB2120	DB2302	DB2291
A4T4	5	8	13	15	17
A4GT4	1	11	20	21	18
A4GCT4	1	3	8	11	8
AATTAATT	1	12	14	19	16
AATTGAATT	2	8	15	14	13
AATTGCAATT	3	2	3	3	4

[a] $\Delta T_m = T_m$ (the complex) − T_m (the free DNA). The listed values are for 2 : 1 [ligand]/[DNA] ratio and an average of two independent experiments with a reproducibility of ±0.5 °C. The green box indicates very strong interaction of listed compounds with the corresponding DNA sequences.

Biosensor-surface plasmon resonance (SPR): high resolution evaluation of binding affinity, kinetics, stoichiometry and cooperativity

In order to quantitatively examine the binding specificity, stoichiometry, and kinetics of the compounds of most interest (based on T_m screening results) were evaluated with a spectrum of mixed DNA sequences by biosensor-SPR methods.[37] DB2120 has unusually strong and selective binding for the single G·C bp containing A4GT4 sequence with much lower binding affinity for A4T4 and A4IT4 (Table 3). The high selectivity for the G over the I sequence is very encouraging for our design procedure.

The binding affinity and sequence selectivity for the hydrophilic oxyethylene linked ABP molecules which have the potential to H-bond with a G-NH_2, have been evaluated by SPR. DB2146 [$K_A = 1.2 \times 10^8$ M^{-1}] and DB2295, [$K_A = 1.7 \times 10^9$ M^{-1}] bind with high affinity in 1 : 1 complexes with the A4GT4 sequence (Table 3), a favorable improvement compared with the parent, DB2114. As with DB2114, however, these two compounds also showed low sequence specificity (Table 3). The interaction of DB2146 and DB2295 with AATT sequences containing one and two G·C bp, showed very interesting results as in T_m experiments. With AATTGAATT, DB2146 and DB2295

Table 3 Biosensor-SPR equilibrium binding affinities (K_A M^{-1})[a] of two-site targeting compounds with mixed DNA sequences

	Compound				
DNA	DB2120	DB2302	DB2291	DB2146	DB2295
A4T4	3.2×10^8	2.0×10^9	8.3×10^8	7.4×10^7	3.1×10^8
A4IT4	8.3×10^8	4.5×10^9	1.3×10^9	1.7×10^8	9.0×10^8
A4GT4	6.6×10^{10}	4.2×10^{11}	3.4×10^9	1.2×10^8	1.7×10^9
AATTGAATT	9.0×10^9	2.4×10^{10}	2.3×10^9	2.7×10^8	9.0×10^8
AATTGCAATT	$K_{A1} = 8.3 \times 10^6$ $K_{A2} = 5.0 \times 10^7$ $K = 2.0 \times 10^7$	$K_{A1} = 1.2 \times 10^7$ $K_{A2} = 7.1 \times 10^7$ $K = 2.9 \times 10^7$	$K_{A1} = 3.3 \times 10^6$ $K_{A2} = 2.4 \times 10^8$ $K = 2.8 \times 10^7$	$K_{A1} = 3.4 \times 10^7$ $K_{A2} = 8.3 \times 10^7$ $K = 5.3 \times 10^7$	$K_{A1} = 6.2 \times 10^6$ $K_{A2} = 2.4 \times 10^7$ $K = 1.2 \times 10^7$

[a] The listed binding affinities are an average of two independent experiments carried out with two different sensor chips and the values are reproducible within 10% in different experiments. The experiments were conducted in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) at 25 °C. $K = (K_{A1}K_{A2})^{0.5}$ for the AATTGCAATT sequence. The green box indicates very strong interactions of listed compounds with the corresponding DNA sequence.

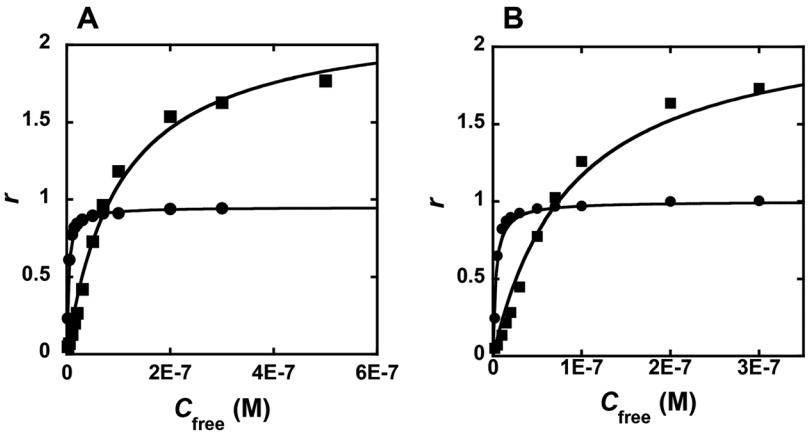

Fig. 2 Comparison of the SPR binding affinity for AATTGAATT (●) and AATTGCAATT (■) DNA sequences with (A) DB2146 and (B) DB2295. RU values from the steady-state region of SPR sensorgrams are converted to r (r = RU/RU_{max}) and are plotted against the unbound compound concentration (flow solution). The lines are the best fit values of a single site or two-site interaction models and K values are in Table 3.

formed strong monomer complexes (Fig. 2 and Table 3). Surprisingly, however, SPR sensorgrams of most of the diamidines evaluated with AATTGCAATT exhibited cooperative 2 : 1 complex formation (Fig. 2 and Table 3). It is clear from the SPR results that the AATTGCAATT promote dimer formation with GC sequences that can have a relatively wide minor groove.[38,39] This is an important result for our design studies to move to new compounds that can selectively recognize two adjacent G·C bps.

The pyridyl compounds DB2120 and DB2302 have sub-nM monomer binding with the sequence, A4GT4, with a very fast on and very slow off rates that result in large binding affinities [K_A = 6.6 × 10^{10} M^{-1}, DB2120 and K_A = 4.2 × 10^{11} M^{-1}, DB2302]. Even after a very long dissociation time for [~1 h] both DB2120 and DB2302 showed only small amounts of dissociation from DNA complexes (Fig. 3). These two compounds also showed an extraordinary 200 times binding selectivity with A4GT4 over the A4T4 sequence (Table 3 and Fig. 3). It is apparent from the sensorgrams that the binding patterns of both compounds with A4GT4 and A4T4 sequences are quite different. Even though both DB2120 and DB2302 show similar binding on-rates with the A4GT4 and A4T4 sequences, the ~300-fold slower off-rates from A4GT4 compared to the A4T4 sequence for both DB2120 and DB2302 (DB2302, k_d = 8 ± 1.8 × 10^{-6} s^{-1} for A4GT4 and k_d = 2.4 ± 1.3 × 10^{-3} s^{-1} for A4T4) (Fig. 2) result in a high sequence specificity for these two pyridyl molecules. Changing "G" (A4GT4) to "I" (Inosine, A4IT4) at the minor groove binding site decreases the binding affinity by 100 fold for both of the molecules as would be expected for the interruption of hydrogen bonding between the central pyridyl group and G-NH_2 in A4IT4 (Table 3 and Fig. 3B). The disruption of hydrogen bonding in A4IT4 increases the complex dissociation 70-fold (DB2302, k_d = 5.6 ± 1.2 × 10^{-4} s^{-1} for A4IT4) in comparison to A4GT4. The change in association constant is much less and most of the affect of the G to I change is due to the increase in k_d. The significantly lower binding affinities for A4T4 and A4IT4 sequences for both pyridyl linked molecules reveal that the pyridyl-N and G-NH_2

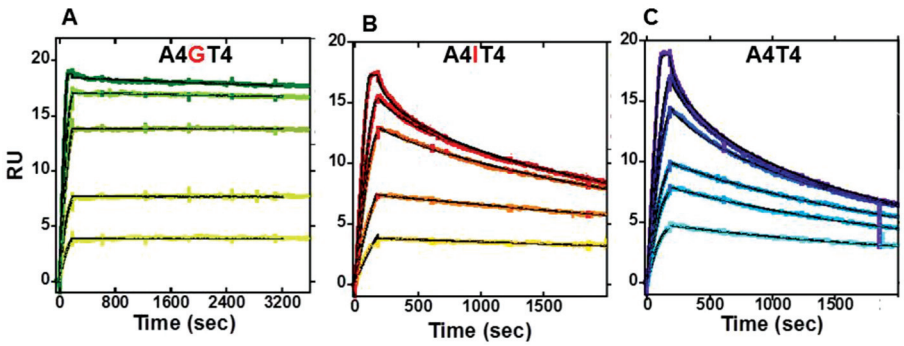

Fig. 3 Representative SPR sensorgrams for DB2302 in the presence of (A) A4GT4, (B) A4IT4 and (C) A4T4 hairpin DNAs. The data are fitted to a 1 : 1 binding function with mass transfer correction to determine equilibrium binding constants. In (A) (B) and (C) the solid black lines are best fit values for the global kinetic fitting of the results with a single site function. The injected concentrations of DB2302 with A4GT4 and A4IT4 are 3, 5, 7, 9 and 12 nM whereas in the A4T4 sequences the injected concentrations are higher; 5, 7, 9, 12, 15, and 20 respectively in 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA buffer, pH 7.4.

H-bond formation is an essential component for strong binding affinity and sequence selectivity for the A4GT4 sequence. The increment in overall binding affinity (Table 3) for isopropylamidino substituted DB2302 may be due to van der Waals interactions between the terminal isopropylamidino moiety and flanking AT bases. DB2302 has also been able to perform as another quite sequence specific ligand with the A4GT4 sequence.

The $-OCH_3$ substituted pyridyl analog; DB2291 binds almost as strongly as DB2120 to A4T4, but surprisingly showed very little sequence specificity because of a significant decrease in binding affinity for the A4GT4 sequence (Table 3). The results suggest that at pH 7.4 (experimental conditions) the pyridyl-*N* which is responsible for G-NH hydrogen bonding becomes partially protonated due to the lowering of the pyridyl pK_a by the positive inductive effect of the oxymethylene group. This observation supports the unprotonated pyridyl-*N* as a key factor for sequence selective G·C bp recognition. With the A-tract A4GCT4 sequence, no significant binding was observed for DB2120 and its analogs in agreement with T_m results.

DB2120 and DB2302 have been investigated with two additional mixed DNA sequences AATTGAATT and AATTGCAATT. The sensorgrams with AATTGAATT (Fig. 4) showed that these two molecules have very strong binding affinity towards the AATTGAATT sequence, and global kinetics fitting yielded a 1 : 1 binding mode with a $K_A = 9.0 \times 10^9\ M^{-1}$ for DB2120 and $K_A = 2.4 \times 10^{10}\ M^{-1}$ for DB2302. This is a very significant 3000 fold increase in binding affinity compared to the corresponding phenyl compound, DB2119.[25] Although the methoxy DB2291 showed less sequence specificity, it also binds very strongly with AATTGAATT with $K_A = 2.3 \times 10^9\ M^{-1}$. Surprisingly, the pyridyl linked ABP molecules bind as 2 : 1 cooperative dimers with the AATTGCAATT sequence (Table 3 and ESI,† Fig. 5). The SPR results are given in Table 3, Fig. 3 and 5 along with the previously published DB2120 results for comparison.[31]

Electrospray ionization mass spectrometry (ESI-MS): stoichiometry, cooperativity, and relative affinity

ESI-MS allows determination of relative affinities for all observed species, stoichiometry (directly from mass), and

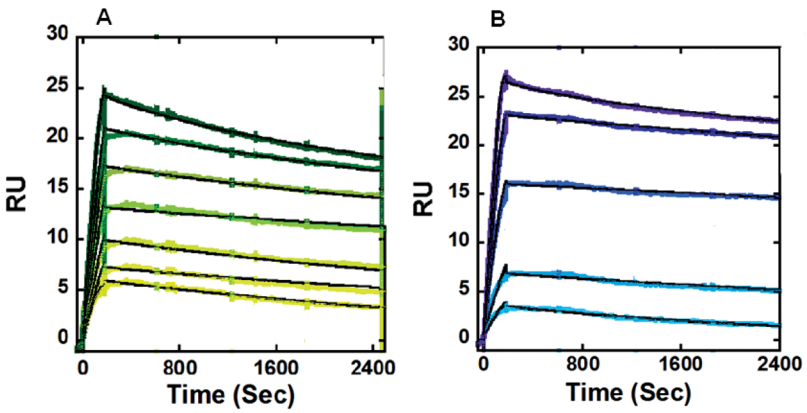

Fig. 4 Representative SPR sensorgrams for (A) DB2120, and (B) DB2302 in presence of AATTGAATT hairpin DNA. The data are fitted to a 1 : 1 binding function with mass transfer correction to determine equilibrium binding constants. In (A) and (B) the solid black lines are best fit values for the global kinetic fitting of the results with a single site function. The injected concentrations of DB2120 with AATTGAATT are 3, 5, 7, 9, 12, 15, 20 nM whereas with DB2302 with AATTGAATT they are 3, 5, 9, 12 and 15 nM respectively in 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA buffer, pH 7.4.

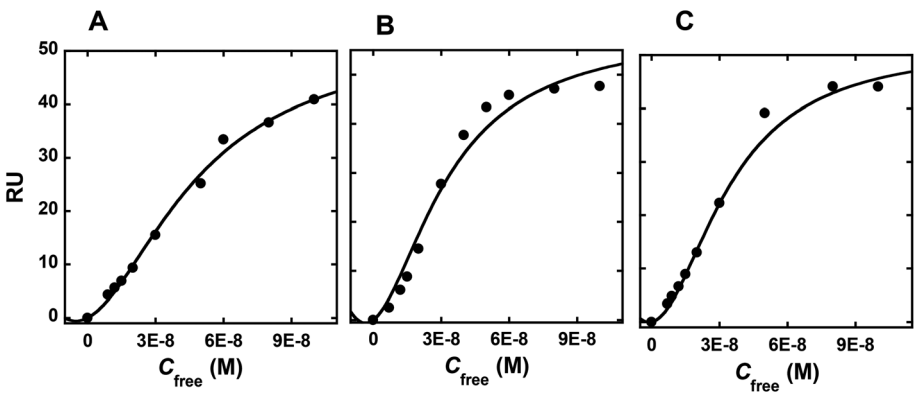

Fig. 5 SPR steady-state binding curves for AATTGCAATT DNA sequences with (A) DB2120, (B) DB2302 and (C) DB2291. The solid lines are best fits with a 2 : 1 binding model.

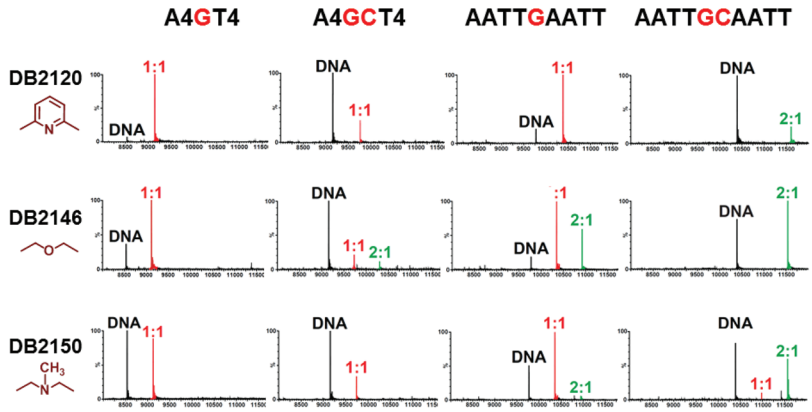

Fig. 6 Comparison of deconvoluted ESI-mass spectra of selected compounds with various DNA sequences at a 2 : 1 drug : DNA mole ratio in 150 mM NH_4OAC buffer. The peaks for the 1 : 1 and 2 : 1 complex are indicated in red and green respectively.

cooperativity when both 1 : 1 and 2 : 1 species are observed (Fig. 6). Mass spectra of DB2120 complexes show very strong monomer binding with A4GT4 and AATTGAATT sequences with high intensity 1 : 1 peaks (Fig. 6) in agreement with very strong 1 : 1 binding observed in SPR. Weak 2 : 1 complexes with DB2120 are observed for AATTGCAATT and weak 1 : 1 peaks for A4GCT4 also agree with other biophysical methods. With the AATTGAATT sequence, DB2146 forms relatively high intensity 1 : 1 and weak 2 : 1 complexes, suggesting non-cooperative dimer formation. However, with AATTGCAATT DB2146 forms a strong 2 : 1 complex (Fig. 6), in agreement with the cooperative dimer formation in SPR. The results with the AATTGCAATT sequence highlight important molecular recognition features that could be incorporated into future compound design schemes for sequences with two G·C bps. These results suggest that the binding aptitude of different linker based symmetrical ABP compounds depends on the compound structure, linker flexibility, and lone pair electron availability from the heteroatom of the linker and the width, structure and chemistry of the minor groove of duplex DNA. In summary, the mass spectrometry results support the SPR and T_m results.

Circular dichroism: probing conformational changes and binding mode

Circular dichroism (CD) is a powerful, non-invasive method to understand conformational changes of biomolecules on complex formation. CD also can provide small molecule binding modes with DNA using pattern recognition.[40,41] Fig. 7 shows the CD spectra of DB2302 with A4GT4 (Fig. 7A) and DB2146 with AATTGCAATT (Fig. 7B). The addition of DB2302 gives significantly induced CD signals (ICD) between 300 to 400 nM, where the diamidines absorb and DNA signals do not interfere, with saturation near a 1 : 1 compound to DNA ratio. This result suggests that DB2302 binds as a monomer in the minor groove of the A4GT4 sequence as expected from the structure (ESI, Fig. S1†). Moreover, small and consistent changes in the CD spectral region of DNA (230 to 290 nm) are observed with incremental titration of DB2302, suggesting

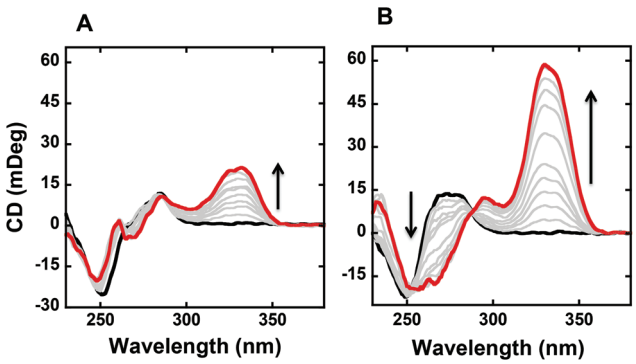

Fig. 7 Circular dichroism spectra of (A) DB2302 with A4GT4 (5 μM) and (B) DB2146 with AATTGCAATT (5 μM) sequence. (A) The ratios of DB2302 to A4GT4 from the bottom to the top are 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4. (B) The ratio of DB2146 to AATTGCAATT are 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0. Ligands were titrated until no further increase in the induced CD signals is observed. The strongly induced CD signal indicates a very strong complex formation in the DNA minor groove.

only minor conformational changes in DNA upon complex formation. On the other hand, titration of DB2146 with AATTGCAATT shows very strong positive ICD changes in the compound absorption region above 300 nm and reaches saturation near a 2 : 1 compound to DNA ratio in agreement with SPR and ESI-MS results. The presence of the isodichroic point at 288 nm wavelength indicates a two state (DNA-bound and unbound) population. There are also large ICD changes below 300 nm, however, these could be due to a DNA configuration change or ICD for the compound on binding. Some combination of these two is also possible but from the CD alone no definite conclusion can be made.

Conclusions

The ability to design compounds that target dsDNA in any sequence context using synthetic organic molecules remains

a challenging problem for the field. In addition to sequence specificity, a goal of this work is to prepare compounds that can enter cells and show target specific interactions with nuclear DNA. For this goal, we have used a heterocyclic platform that has shown good biological activity from cell studies through human clinical trials. In this report, a focused library of linked, symmetrical bis-benzimidazole-amidine molecules was designed and synthesized with different modifications in their linker. The linkers varied by substitutions that can not only alter the inherent properties of the ligand itself, such as hydrophobicity, molecular degrees of freedom and solubility but also play crucial roles in sequence-specific DNA recognition by modifications of hydrogen bonding, electrostatic and van der Waals interactions.

The biophysical results in this research revealed four key points for the rational design of organic small molecules for mixed DNA sequence recognition. Compounds that show strong binding in the minor groove provide our current best design platform. The compound sequence specificity and affinities vary depending on both the compound geometry/shape and chemical properties. Second, some pyridine compounds have an outstanding ability to recognize a G·C bp in an AT sequence context. This is a significant breakthrough in our mixed sequence DNA recognition effort. Third, alkyl-ether-linked compounds, such as DB2146, place an H-bond acceptor in the same position as the pyridyl *N* but they have much weaker binding to the single GC sequences and they are not as sequence-specific as pyridyl compounds such as DB2120. Fourth, some compounds, in particular, the alkyl-oxy linked DB2146, have strong cooperative 2 : 1 complex formation with the AATTGCAATT but not the A4GCT4 sequence. All of the compounds show favorable 2 : 1 interactions with AATTGCAATT. We believe the favorable 2 : 1 binding to this DNA sequence is due to microstructural variations in the DNA minor groove at AATTGCAATT but definitive analysis of this complex will need higher resolution structural studies. These results show that new design motifs for mixed sequence recognition of DNA are possible with the heterocyclic-amidine-platform.

Experimental section

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

Acknowledgements

This work was supported by the US NIH [Grant No. GM111749 to W. D. W. and D. W. B.]. The authors thank Carol Wilson for manuscript assistance and Dr Siming Wang for ESI-MS assistance.

Notes and references

1 M. Munde, A. Kumar, P. Peixoto, S. Depauw, M. A. Ismail, A. A. Farahat, A. Paul, M. V. Say, M.-H. D. Cordonnier, D. W. Boykin and W. D. Wilson, *Biochemistry*, 2014, **53**, 1218–1227.
2 J. Rodrıguez, J. Mosquera, J. R. Couceiro, M. E. Vázquez and J. L. Mascareñas, *Chem. Sci.*, 2015, **6**, 4767–4771.
3 R. Moretti and A. Z. Ansari, *Biochimie*, 2008, **90**, 1015–1025.
4 J. A. Raskatova, J. L. Meier, J. W. Puckett, F. Yang, P. Ramakrishnan and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 1023–1028.
5 Y. Zhang, G. Sicot, X. Cui, M. Vogel, C. A. Wuertzer, K. L. Geyda, J. Wheeler, D. A. Harki, K. A. Muzikar, D. A. Stolper, P. B. Dervan and A. S. Perkins, *Biochemistry*, 2011, **50**, 10431–10441.
6 G. N. Pandian, K. I. Shinohara, A. Ohtsuki, Y. Nakano, M. Masafumi, T. Bando, H. Nagase, Y. Yamada, A. Watanabe, N. Terada, S. Sato, H. Morinaga and H. Sugiyama, *ChemBioChem*, 2011, **12**, 2822–2828.
7 M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson and G. M. K. Poon, *Nucleic Acids Res.*, 2014, **42**, 1379–1390.
8 G. He and J. K. Bashkin, *Future Med. Chem.*, 2015, **7**, 1953–1955.
9 W. D. Wilson, F. A. Tanious, A. Mathis, D. Tevis, J. E. Hall and D. W. Boykin, *Biochimie*, 2008, **90**, 999–1014.
10 T. Synold, B. Xi, J. Wu, Y. Yen, B. Li, F. Yang, J. Phillips, N. Nickols and P. Dervan, *Cancer Chemother. Pharmacol.*, 2012, **70**, 617–625.
11 S. Kumar, N. Ranjan, P. Kellish, C. Gong, D. Watkins and D. P. Arya, *Org. Biomol. Chem.*, 2016, **14**, 2052–2056.
12 J. Bordello, M. I. Sánchez, M. E. Vázquez, J. L. MascareÇas, W. A. Soufi and M. Novo, *Chem. – Eur. J.*, 2015, **21**, 1609–1619.
13 A. N. Koehler, *Curr. Opin. Chem. Biol.*, 2010, **14**, 331–340.
14 A. Chen and A. N. Koehler, *Science*, 2015, **347**, 713–714.
15 K. R. Gore, G. N. Nawale, S. Harikrishna, V. G. Chittoor, S. K. Pandey, C. Höbartner, S. Patankar and P. I. Pradeepkumar, *J. Org. Chem.*, 2012, **77**, 3233–3245.
16 K. J. Koeller, G. D. Harris, K. Aston, G. He, C. H. Castaneda, M. A. Thornton, T. G. Edwards, S. Wang, R. Nanjunda, W. D. Wilson, C. Fisher and J. K. Bashkin, *Med. Chem.*, 2014, **20**, 338–344.
17 T. G. Edwards, K. J. Koeller, U. Slomczynska, K. Fok, M. Helmus, J. K. Bashkin and C. Fisher, *Antiviral Res.*, 2011, **91**, 177–186.
18 J. A. Raskatov, N. G. Nickols, A. E. Hargrove, G. K. Marinov, B. Wold and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 16041–16045.
19 R. D. Taylor, A. Chandran, G. Kashiwazaki, K. Hashiya, T. Bando, H. Nagase and H. Sugiyama, *Chemistry*, 2015, **21**, 14996–15003.
20 L. Han, G. N. Pandian, A. Chandran, S. Sato, J. Taniguchi, G. Kashiwazaki, Y. Sawatani, K. Hashiya, T. Bando, Y. Xu, X. Qian and H. Sugiyama, *Angew. Chem., Int. Ed.*, 2015, **30**, 8700–8703.

21 A. I. Khalaf, C. Bourdin, D. Breen, G. Donoghue, F. Scott, J. C. J. Suckling, D. MacMillan, C. Clements, K. Fox and D. A. T. Sekibo, *Eur. J. Med. Chem.*, 2012, **56**, 39–47.
22 C. Suckling, *Future Med. Chem.*, 2012, **4**, 971–989.
23 B. S. P. Reddy, S. M. Sondhi and J. W. Lown, *Pharmacol. Ther.*, 1999, **84**, 1–111.
24 M. Munde, G. M. K. Poon and W. D. Wilson, *J. Mol. Biol.*, 2013, **425**, 1655–1669.
25 M. Rahimian, A. Kumar, M. Say, S. A. Bakunov, D. W. Boykin, R. R. Tidwell and W. D. Wilson, *Biochemistry*, 2009, **48**, 1573–1583.
26 Y. Liu, Y. Chai, A. Kumar, R. R. Tidwell, D. W. Boykin and W. D. Wilson, *J. Am. Chem. Soc.*, 2012, **134**, 5290–5299.
27 J. K. Thuita, M. Z. Wang, J. M. Kagira, C. L. Denton, M. F. Paine, R. E. Mdachi, G. A. Murilla, S. Ching, D. W. Boykin, R. R. Tidwell, J. E. Hall and R. Brun, *PLoS Neglected Trop. Dis.*, 2012, **6**, e1734.
28 P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani and R. Romagnoli, *Med. Res. Rev.*, 2004, **24**, 475–528.
29 T. Wenzler, D. W. Boykin, M. A. Ismail, J. E. Hall, R. R. Tidwell and R. Brun, *Antimicrob. Agents Chemother.*, 2009, **53**, 4185–4192.
30 B. Nguyen, S. Neidle and W. D. Wilson, *Acc. Chem. Res.*, 2009, **42**, 11–21.
31 A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M.-H. D. Cordonnier, D. W. Boykin and W. D. Wilson, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 4927–4932.
32 Y. Chai, A. Paul, M. Rettig, W. D. Wilson and D. W. Boykin, *J. Org. Chem.*, 2014, **79**, 852–866.
33 J. C. Marini, S. D. Levene, D. M. Crothers and P. T. Englund, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 7664–7668.
34 T. A. Shapiro, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7809–7813.
35 A. Paul, Y. Chai, D. W. Boykin and W. D. Wilson, *Biochemistry*, 2015, **54**, 577–587.
36 N. K. Harika, A. Paul, E. Stroeva, Y. Chai, D. W. Boykin, M. W. Germann and W. D. Wilson, *Nucleic Acids Res.*, 2016, **44**, 4519–4527.
37 R. Nanjunda, M. Munde, Y. Liu and W. D. Wilson, Real-Time Monitoring of Nucleic Acid Interactions with Biosensor-Surface Plasmon Resonance, in *Methods for Studying Nucleic Acid/Drug Interactions*, ed. Y. Tor and M. Wanunu, CRC Press, Boca Raton, 2011, pp. 91–134.
38 S. P. Hancock, T. Ghane, D. Cascio, R. Rohs, R. D. Felice and R. C. Johnson, *Nucleic Acids Res.*, 2013, **41**, 16750–16760.
39 R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig, *Nature*, 2009, **461**, 1248–1253.
40 M. Eriksson and B. Nordén, *Drug-Nucleic Acid Interactions*, Elsevier, 2001, vol. 340, pp. 68–98.
41 L. H. Fornander, L. Wu, M. Billeter, P. Lincoln and B. Nordén, *J. Phys. Chem. B*, 2013, **117**, 5820–5830.