Showcasing research from Professor W. David Wilson's laboratory, Department of Chemistry, Georgia State University, Atlanta, 30303 GA USA.

Engineered modular heterocyclic-diamidines for sequence-specific recognition of mixed AT/GC base pairs at the DNA minor groove

The guiding goal of this research is to synthesize novel compounds that selectively bind to DNA sequences and modulate the activity of transcription factors. Targeting specific DNA sequences is of importance because transcription factors have proven difficult systems for the design of selective binding agents and are frequently classified as "undruggable". After trying several unsuccessful compound designs, an N-alkyl-benzodiimidazole structure was selected to successfully provide two H-bond acceptors for adjacent GG-NH groups that are difficult to target. This breakthrough now allows us to selectively target a transcription factor involved in leukemia.

Artwork by Dr. Ananya Paul.

As featured in:

See W. David Wilson *et al.*, *Chem. Sci.*, 2021, **12**, 15849.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

Check for updates

All publication charges for this article have been paid for by the Royal Society of Chemistry

Engineered modular heterocyclic-diamidines for sequence-specific recognition of mixed AT/GC base pairs at the DNA minor groove†

Pu Guo, [a] Abdelbasset A. Farahat, [ab] Ananya Paul, [a] David W. Boykin [a] and W. David Wilson *[a]

This report describes a breakthrough in a project to design minor groove binders to recognize any sequence of DNA. A key goal is to invent synthetic chemistry for compound preparation to recognize an adjacent GG sequence that has been difficult to target. After trying several unsuccessful compound designs, an *N*-alkyl-benzodiimidazole structure was selected to provide two H-bond acceptors for the adjacent GG-NH groups. Flanking thiophenes provide a preorganized structure with strong affinity, DB2831, and the structure is terminated by phenyl-amidines. The binding experimental results for DB2831 with a target AAAGGTTT sequence were successful and include a high ΔT_m, biosensor SPR with a K_D of 4 nM, a similar K_D from fluorescence titrations and supporting competition mass spectrometry. MD analysis of DB2831 bound to an AAAGGTTT site reveals that the two unprotonated *N* of the benzodiimidazole group form strong H-bonds (based on distance) with the two central G-NH while the central −CH of the benzodiimidazole is close to the −C=O of a C base. These three interactions account for the strong preference of DB2831 for a -GG- sequence. Surprisingly, a complex with one dynamic, interfacial water is favored with 75% occupancy.

Introduction

Engineering of organic compounds that can recognize mixed base pair (bp) sequences of DNA containing combinations of A·T and G·C bps has been a long-term goal of nucleic acid molecular recognition. Starting with the discovery of netropsin (Nt.) and distamycin (Dst.), various types of minor groove binding AT specific heterocyclic cations,[1] have had major use as biological stains, biotechnology reagents, and therapeutic agents. The AT-specific compounds have included synthetic agents such as Hoechst 33258, DAPI, furamidine, and recently the first metallo-minor groove binders, Ru(II)-bipyridyl complexe.[1] To broaden the uses of minor groove binders in all areas, an expanded library of sequence-specific compounds is needed. The broader specific recognition of G·C bps by organic molecules is a necessary next step in novel compound library development for minor groove compounds.

As part of a major effort to re-design AT-specific heterocyclic cations for recognition of G·C bps as well as A·T bps we now have modules that strongly and specifically bind single G·C bps in an AT sequence context.[2] Such new compounds have broad potential applications, for example, in targeting transcription factors (TFs) to modulate gene expression.[3] Mutations, aberrant regulation, or other disorders that modify the activity of TFs lead to a number of different kinds of diseases.[4] Given the variety of roles and diseases controlled by TFs, a strong interest in targeting them to modulate their activity with small molecules has developed.[4,5] A problem with this approach is that TFs have evolved to bind high molecular-weight nucleic acids but not, typically, small molecules.[5] Binding sites on TFs that interact strongly and specifically with small molecules are difficult to find and TFs are often defined as "undruggable".[5] We are pursuing an entirely different approach to target TF–DNA complexes by designed agents to bind to specific DNA sequences and to subsequently inhibit the promoter sequence interactions and functions of specific TFs.

As an example, expression of the TF PU.1 is frequently impaired in patients with acute myeloid leukemia (AML).[3,6] We developed a series of PU.1 inhibitors based on the AT-rich 5′-flanking sequence of many conserved PU.1 promoter sequences, which have a central -GGAA-.[3,7] As ligands and PU.1 do not directly compete for the same DNA binding site, inhibition of major groove binding TFs by minor groove binding small molecules is a multipart mission.[3,7] To improve our PU.1 targeting ability, additional GC-specific compounds that can recognize a broad selection of DNA sequences throughout the

aDepartment of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, 50 Decatur St SE, Atlanta, GA 30303, USA. E-mail: wdw@gsu.edu; Tel: +1 404-413-5503

bDepartment of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc04720e

PU.1 promoter as well as promoter sequences for other TFs (Fig. 1A and B) are needed. Binding specifically to such sequences will be very useful in biotechnology and in the design of new therapeutic agents.

Currently, for example, we lack specific binders which can target the conserved -GGAA- promoter binding site of PU.1. To bind strongly and specifically to different, multiple GC-containing sites, new types of cell-permeable DNA binding
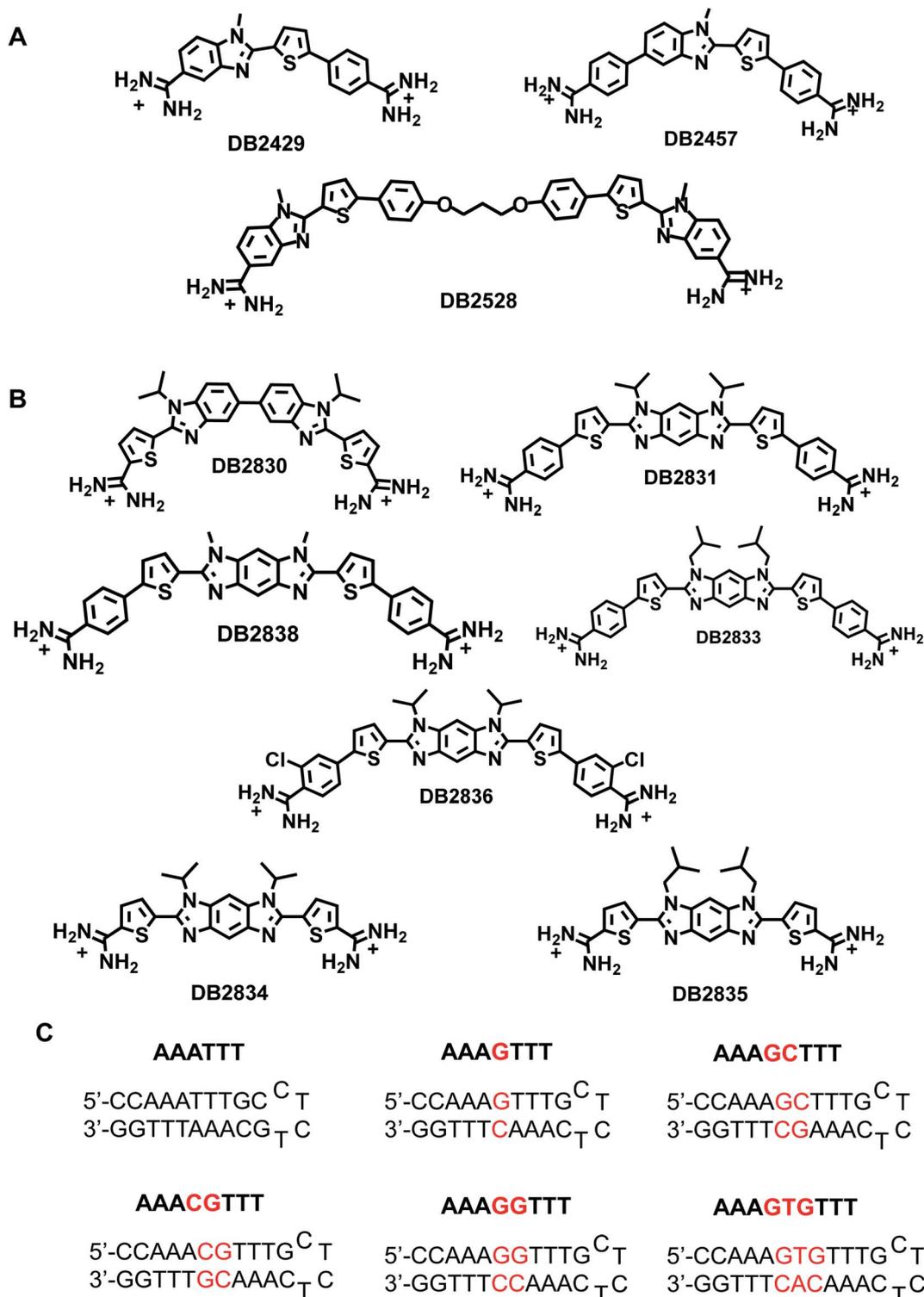
Fig. 1   (A) Chemical structures of single G·C base–pair binders; (B) chemical structure of novel two G·C base pairs binders prepared for this study; (C) the DNA sequences used in this study; DNA sequences used for SPR studies were labeled with 5′-biotin.

agents must be engineered. The conserved GGAA site is a logical next step in ETS protein targeting. The success of single G·C bp binders gives us a powerful combination method to design new types of compound structures to match the requirements of multiple G·C bps recognition.

The PU.1 recognition sequence has a number of sites with different GC-containing sequences. The central λB PU.1 promoter sequence is: $5' - \text{ATAAAA}\underline{\text{GGAA}}\text{ GTGAAAC} - 3'$, a typical 5'AT-rich PU.1 promoter sequence. Heterocyclic organic cations to target the GG sequence have not previously been possible to prepare. Targeting the central -GGAA- component of the PU.1 promoter gives us the key critical site for inhibition of ETS TF binding and is essential for complete PU.1 promoter recognition. One way to target the -GGAA- binding site is to design or combine two single G·C bp recognition units in close proximity to bind the two adjacent GC bps with flanking AT sequences.

A new design concept was used to engineer DB2830 (Fig. 1B) with two directly linked thiophene-N-i-Pr-BI modules. The compound, however, has relatively weak binding with the test GG sequence, AAAGGTTT (Fig. 1B). A curvature evaluation mechanism (described below) suggested that the DB2830 structure was too curved to bind optimally in the DNA minor groove. Considering the curvature issue of the molecules and the short distance between two G·C bps, we introduced the benzodiimidazole structure in DB2831 to provide two H-bond acceptors for G-NH$_2$ with the appropriate curvature for minor groove recognition. The compound was successful with strong and specific binding to the GG sequence. The effects of N-substituents, amidines on different aromatic groups, and chloro-substituents were also investigated by modifying the DB2831 chemical structure and are also reported here. We have used the hydrophobic N-substituents to potentially enhance the cell uptake as these compounds are targeted to inhibit the PU.1 transcription factor. The compounds are already relatively polar with two charges and have low K_D values.

The new benzodiimidazole DNA recognition structure is based on the single GC binding modules in DB2429 and DB2457 (Fig. 1A), which have a thiophene-N-alkyl-BI interaction that provides a recognition unit best characterized as a σ-hole, a successful tool for designing G-NH$_2$ minor groove binding modules.[2] Compounds that incorporate the σ-hole motif (thiophene N-RBI) are a significant step forward in our molecular design and synthesis project for the recognition of mixed bp DNA sequences.[2] The σ-hole interaction preorganizes the thiophene-N-R-BI unit for GC interaction in the minor groove. The σ-hole module is regarded as an essential component of the thiophene N-R-BI type G·C bp binders and is part of the benzodiimidazole module.

Results

Chemistry

Scheme 1 outlines the synthesis of the diamidines compounds **5a–f**. Nucleophilic displacement of the starting difluoro compound **1** with different amines in ethanol afforded the amino-substituted dinitro intermediate **2a–c**.[8] We tried the reduction of the dinitro compound using both catalytic hydrogenation and tin chloride, but the reaction was never complete and due to the insolubility of the starting dinitro compound, the impure product cannot be chromatographed. Finally, we used sodium borohydride and Pd(C) in methanol/dichloromethane as a solvent to get the intermediate tetraamine **3a–c** in good yield. This amine was subjected to an oxidative cyclization reaction using different aldehydes[9] and sodium metabisulfite in DMSO to furnish the intermediate imidazobenzimidazole **4a–f**.[10] The bis-nitriles **4a–f** were allowed to react with lithium bis(trimethylsilyl)amide in THF, followed by subsequent deprotection of the silylated amidines with ethanolic HCl (gas) to furnish the hydrochloride salts of the diamidines **5a–f**.[4,11] The synthesis of the diamidine **11** is described in Scheme 2 (ESI†). The boronic ester intermediate **7** was prepared by reaction of the starting bromo compound **6** with bis(pinacolato) diboron using bis(triphenylphosphine) palladium dichloride and potassium acetate in dioxane. The former boronic ester underwent Suzuki coupling with the bromo compound **6** to produce the bis nitro intermediate **8**. This nitro compound was reduced using sodium borohydride and Pd(C) in methanol/dichloromethane as a solvent to get the intermediate tetraamine **9** in good yield. The bisnitrile **10** was prepared by coupling of the intermediate tetramine and 5-formylthiophene-2-carbonitrile using sodium metabisulfite in DMSO. The final diamidine **11** was prepared by silylation of the bisnitrile **10** using lithium bis(trimethylsilyl)amide in THF, followed by subsequent deprotection of the silylated amidines with ethanolic HCl (gas). The purity of all final compounds had been verified by ^{1}HNMR, ESI-HRMS and elemental analysis (ESI†).

DNA thermal melting: screening for relative binding affinity

Changes in thermal melting temperature (T_m) of DNA provide an initial screening of ligands for binding affinity with different DNA sequences. Six related DNA sequences either with a pure AT (AAATTT) binding site, a single G·C bp binding site (AAAGTTT) or with different numbers of AT bps between two adjacent G·C bps (AAAGGTTT) were selected for testing (Fig. 1C). These sequences provide a systematic set to determine the relative binding selectivity of the compounds at different DNA binding sites with different DNA minor groove microstructures.

DB2830, a linked-benzimidazole–thiophene, was designed to target the two adjacent G·C bp sequence by two adjacent single G·C binding modules. This planar structure is a new development in recognition of the DNA minor groove (Fig. 1B). The compound, however, shows only a moderate binding affinity with our desired DNA binding site, AAAGGTTT ($\Delta T_m = 6$), due to an excessive curvature for the minor groove. To reduce the curvature of the compound for a better match to the minor groove surface, an entirely new structure with a more planar core, benzodiimidazole-bisthiophene was designed and synthesized, DB2831 (Fig. 1B). The benzodiimidazole-bisthiophene, with a central fused ring, is a new idea for DNA recognition, especially minor groove binding. DB2831, is found
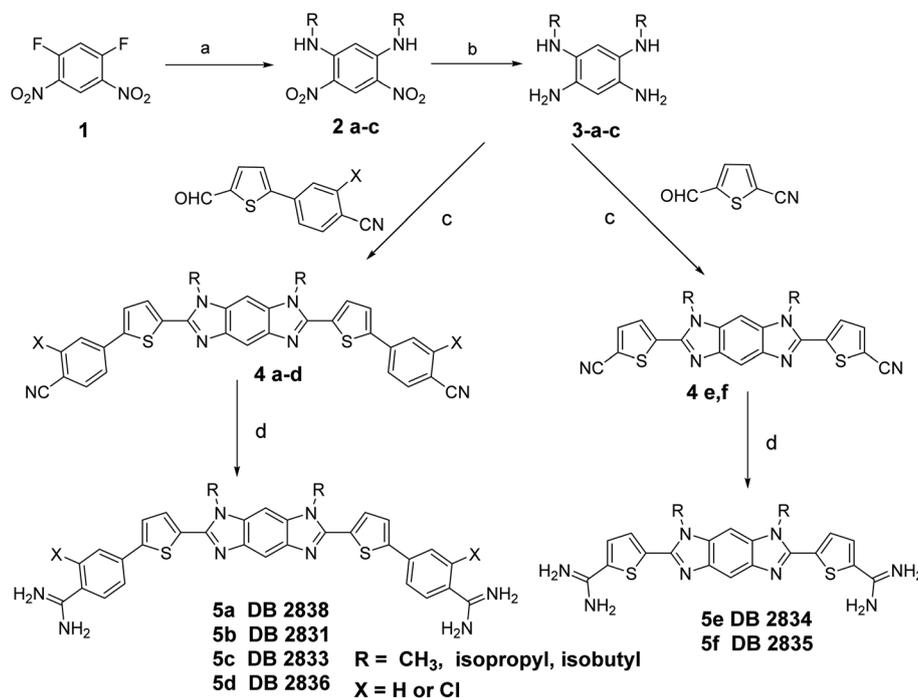
Scheme 1 Reagents and conditions: (a) amines/EtOH, rt; (b) NaBH$_4$, Pd(C), CH$_2$Cl$_2$, MeOH; (c) Na$_2$S$_2$O$_3$/DMSO, 130 °C; (d) (i) LiN(TMS)$_2$/THF, (ii) HCl/E.

to exhibit high stabilization towards AAAGGTTT ($\Delta T_m = 10$), and AAACGTTT ($\Delta T_m = 10$). In addition, DB2831 shows a weak stabilization potential for both pure AT and single G·C bp containing sequences (Table 1), illustrating the excellent sequence selectivity of the compound.

We have previously observed that, in single G·C bp binding sequences, bulky N-alkyl substituents always facilitate the sequence selectivity for comparatively wider minor groove sequences.[2] With this idea, an isobutyl derivative of DB2831, DB2833, was prepared to determine the N-alkyl substitution

effect on binding. Surprisingly, this bulky substitution caused a marked decrease in binding affinity for DB2833 (Table 1 and Fig. 1B) and this substitution was discontinued.

Truncated compounds with terminal thiophene amidines (DB2834 and DB2835) were synthesized and tested to see how the molecular size of the compound affects sequence binding affinity and selectivity. The low ΔT_m values indicate that terminal phenyl groups play significant roles for DB2831 affinity. A 2-Cl phenyl amidine proved to be an effective modification to increase the binding specificity of single G·C binders.[2] For this reason, DB2836 was designed and synthesized and showed strong binding to AAAGGTTT ($\Delta T_m = 8$) with excellent selectivity. DB2838 with N–Me substituents was also prepared and tested, but that compound provided the surprising result that no binding was detected with the selected sequences. Analysis with organic solvents suggested extensive aggregation of this compound under the experimental conditions accounting for the lack of binding.

Biosensor-SPR: ligand–DNA binding affinity and specificity studies

SPR methods have been used for quantitative determination of binding affinity and selectivity of the newly synthesized benzodiimidazole-bisthiophene (DB2831) and its derivatives with an array of different DNA sequences. A Biacore dextran-coated sensorchip (CM5) was functionalized with streptavidin and used to immobilize three different 5′biotin-labeled DNA sequences (Fig. 1C) in flow cells 2–4 while flow cell 1 was left as a blank for background subtraction. With varying concentrations of a particular compound in the flow solution, we were

Table 1 Thermal melting studies (ΔT_m, °C) of all test compounds with mixed DNA sequences[a]

	AAA	AAA	AAA	AAA	AAA	AAA
	G	GC	CG	GG	GTG
	TTT	TTT	TTT	TTT	TTT	TTT
DB2830	<1	5	2	4	6	3
DB2831	1	3	4	10	10	2
DB2833	1	<1	1	4	4	1
DB2834	<1	2	1	5	5	<1
DB2835	<1	<1	<1	4	2	1
DB2836	1	2	2	8	8	1
DB2838*	<1	1	<1	1	<1	1

[a] $\Delta T_m = T_m$ (the complex) $- T_m$ (the free DNA). 3 μM DNA sequences were studied in Tris–HCl buffer (50 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) with the ratio of 2 : 1 [ligand]/[DNA]. An average of two independent experiments with a reproducibility of 0.5 °C. Full DNA sequences as described in Fig. 1C.

able to determine comparative binding constants for all of the non-aggregated derivatives (Table 2). Sensorgrams were obtained for the compounds and are shown for DB2831 with three different DNAs in Fig. 2. DB2831 with an *N-i*-Pr substituent has exceptionally strong binding ($K_D = (2 \pm 2)$ nM) with AAAGGTTT. DB2831 and all other tested compounds show no detectible or weak binding to AAAGTTT or AAAGTGTTT in our experimental conditions. Interestingly, these sequences have only a one base pair difference from the target sequence, AAAGGTTT, with a very minimal change in minor groove microstructures. Kinetics analysis of ligand–DNA complexes was performed by global fitting with a 1 : 1 binding model for DB2831.

The SPR binding results revealed that DB2831 has an optimized size and curvature for selective recognition and strong affinity for two adjacent two G·C bps in an A-tract sequence.

Biosensor-SPR experiments are well-suited for the kinetic and thermodynamic analysis of many types of interactions. The main limitation of this method is mass transfer for tightly bound ligand–DNA interactions as observed for the DB2831-AAAGGTT complex at 100 and 200 mM NaCl concentrations (Fig. 2). Difficulties with DB2831–DNA complexes include (i) mass transfer limits on kinetics, where the rates of transfer of components from the injected solution to the immobilized component is slower than the association reaction,[12] (ii) very slow dissociation rates due to rebinding during the dissociation phase, and (iii) limited time for the association reaction due to volume limitations in the injection syringe, have been observed for our compound of interest, DB2831. To overcome the mass transfer problem for DB2831 with AAAGGTTT complex, we have conducted SPR experiments at different salt concentrations (from 100 to 400 mM NaCl concentrations) at 25 °C (Fig. 2A, B and S1†). The equilibrium association constants (K_A) obtained by 1 : 1 kinetic fit also point to the relatively rapid dissociation rates of this complex at higher salt concentrations. According to the counterion condensation theory,[13] the logarithm of the

equilibrium binding constants K_A ($=1/K_D$) (from kinetic fits) should be a linear function of the logarithm of salt is reasonable for a dication on DNA complex formation.[13] The K_A values decrease significantly as the salt concentration increases as is typical for DNA–cation complexes.[13] The slopes of the linear fits are ~1.8 which reasonable for a dication on DNA complex formation. The number of phosphate contacts (Z) between DB2831 and hairpin duplex DNA can be obtained by slope/Ψ (Ψ = fraction of phosphate shielded by condensed counterions and is 0.88 for double-stranded B-DNA),[14] and this gives a Z of 2 ± 0.2. Thus, there are two phosphate contacts between DB2831 and DNA which is a very realistic value for this dicationic molecule. The two thiophenes cause a second problem with DB2831 because of their binding to the sensorchip surface. Due to this problem, we could only work with these compounds at low concentrations with biosensor chips (Fig. 2). The sensorgrams become increasingly distorted as the concentration is increased above 30 nM for DB2831.

Removal of terminal phenyl rings of DB2831 (DB2834 and DB2835) causes a large decrease in binding ($K_D = 286$ nM). The initial compound, DB2830, also binds weakly to AAAGGTTT ($K_D = 553$ nM), which strongly supports the thermal melting results. The replacement of *i*-Pr with *i*-Bu substituents (DB2833) causes a considerable reduction in binding ability in agreement with T_m results. DB2836 with –Cl modification at the *ortho* position of the amidines keeps the excellent binding specificity, though the binding affinity with AAAGGTTT, $K_D = 62$ nM, is reduced.

Fluorescence titrations: quantification of ligand binding affinity

The binding selectivity and affinity of DB2831 were further validated by fluorescence spectroscopic titrations. Fluorescence spectroscopy is a rapid and sensitive method for quantitatively measuring the interactions between small molecules and biomolecules, such as DNA. If a change in the fluorescence intensity accompanies the binding of two species, this can be used to monitor the binding interaction and determine the stoichiometry of binding using equilibrium titration methods.[2] To evaluate the binding affinity and specificity of DB2831 with AAAGGTTT, fluorescence titrations were conducted with DB2831 and DNA sequences AAAGGTTT and AAAGTTT. As shown in Fig. 3, DB2831 shows a binding preference for AAAGGTTT with a large increment of fluorescence intensities. The fluorescence titration spectra and binding affinity fitting plots are also shown in Fig. 3. Fluorescence intensities increase when the test sequences are titrated into DB2831 solutions. With the AAAGGTTT sequence, a significant increase in the fluorescence intensity is obtained, while the AAAGTTT titration intensity increases only slightly. The binding affinity ($K_D = 5.6$ nM) between DB2831 and AAAGGTTT was calculated by fitting the fluorescence intensity *versus* DNA concentrations and validated the SPR results.

Circular dichroism (CD): determination of DNA binding mode

Circular dichroism spectroscopy is a very powerful method to evaluate the structure of optically active materials such as

Table 2 Summary of binding affinity (K_D, nM) for the interaction of all test compounds with biotin-labeled DNA sequences using biosensor-SPR method[a]

| | AAA | AAA | AAA |
| | G | GG | GTG |
	TTT	TTT	TTT
DB2830	571	553	NB
DB2831	NB	2	NB
DB2833	NB	NB	NB
DB2834	NB	286	NB
DB2835	NB	NB	NB
DB2836	NB	62	NB
DB2838	NB	NB	NB

[a] All the results in this table were obtained in Tris–HCl buffer (50 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA, 0.05% P20, pH 7.4) at a 100 μL min⁻¹ flow rate. NB means no measurable K_D under our experimental conditions, see Fig. 2D and E for examples. The listed binding affinities are an average of two independent experiments carried out with two different sensor chips and the values are reproducible within 10% experimental errors. Full DNA sequences as described in Fig. 1C.
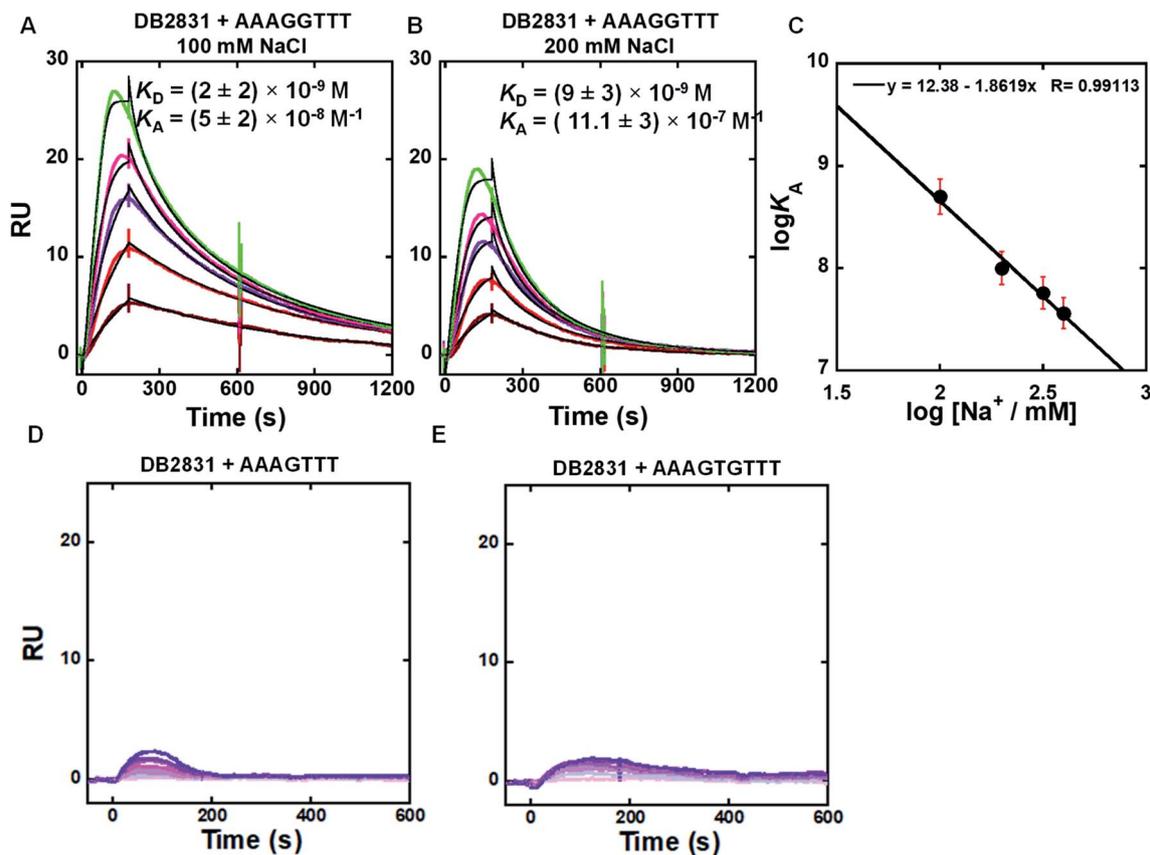
Fig. 2 (A and B) SPR sensorgrams (color) and global kinetic fits (black overlays) for DB2831 with the AAAGGTTT DNA hairpin sequence at 100 mM and 200 mM NaCl; the concentrations of DB2831 in these SPR experiments are 5–30 nM from bottom to top. (C) Salt dependence of K_A for DB2831 binding as determined by SPR. The K_A values were obtained by 1 : 1 kinetic fitting; (D and E) representative SPR sensorgrams for DB2831 in the presence of AAAGTTT, and AAAGTGTTT hairpin DNAs. The concentrations of DB2831 in these SPR experiments are 5–500 nM from bottom to top. Full DNA sequences as described in Fig. 1C.

proteins and DNA. CD spectra monitor the asymmetric environment of the binding of the ligand to DNA and, therefore, can be used to obtain information on the binding mode.[15] This method is also a very convenient and effective method of determining the saturation limit for compounds binding with

DNA sequences. It should be noted that the free DB2831 and DB2834 compounds are optically inactive and hence do not exhibit any CD spectra. However, on addition of these compounds into DNA, substantial positive induced CD signals (ICD) arose in the compound absorption region between 360
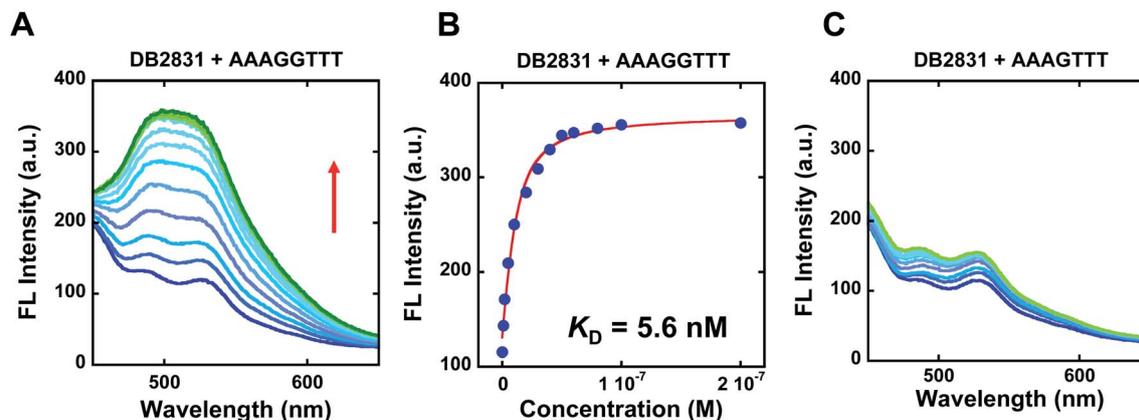

Fig. 3 Fluorescence emission spectra for 10 nM DB2831 titrated with sequence AAAGGTTT (A), or AAAGTTT (C) in TNE100 buffer at 25 °C. The excitation wavelength is 390 nm, respectively. The slit widths are [20, 20 nm]; fluorescence binding curve between 10 nM DB2831 and tested sequences in TNE100 buffer to determine equilibrium constant. Full DNA sequences as described in Fig. 1C.
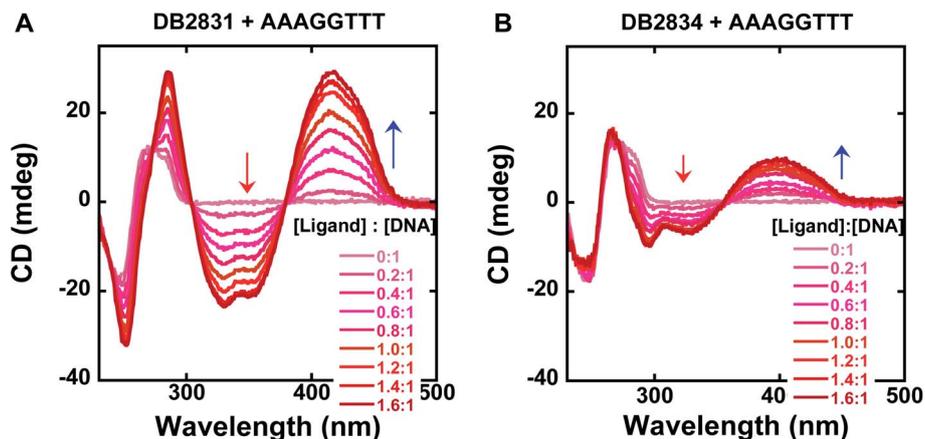
Fig. 4 Circular dichroism spectra for the titration of representative compounds, DB2831, DB2834 with a 5 μM AAAGGTTT sequence in TNE100 buffer at 25 °C. Arrows indicate the changes. Full DNA sequences as described in Fig. 1C.

and 460 nm. These positive ICD signals indicate a minor groove binding mode of the compounds. As can be seen from Fig. 4, DB2831 and DB2834 form complexes in the minor groove of the AAAGGTTT sequences with a 1 : 1 stoichiometry, as expected from the compound's structures and modeling experiments (below), in agreement with SPR results. The lower ICD of DB2834 than DB2831 also agrees that a weaker binding to AAAGGTTT. In summary, the CD titration results confirm a minor groove binding mode for the compounds in Fig. 1 with 1 : 1 binding stoichiometry.

Competition electrospray ionization mass spectrometry (ESI-MS): a direct determination of binding stoichiometry and binding specificity with relative binding affinity

Competition electrospray ionization mass spectrometry with DNA complexes[16] allows high-throughput screening of the interactions of multiple DNA sequences with libraries of minor groove binding molecules. One of the strengths of this method is the direct application in binding specificity determinations. DB2831 was tested in competition mass spectrometry with five DNA sequences: AAATTT, AAAGTTT, AAAGGTTT, AAAGCTTT,
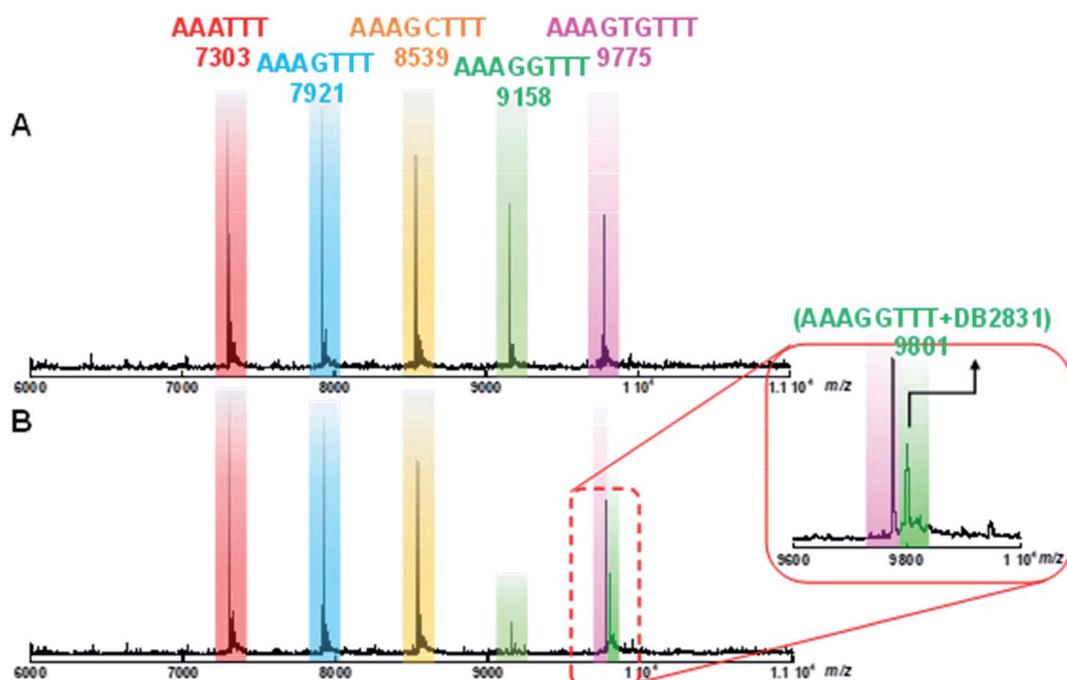

Fig. 5 ESI-MS negative mode spectra of the competition binding of sequences AAATTT; AAAGTTT; AAAGCTTT; AAAGGTTT; AAAGTGTTT: (10 μM each); with 40 μM DB2831 in buffer (50 mM ammonium acetate with 10% methanol (v/v), pH 6.8). (A) The ESI-MS spectra of free DNA mixture. (B) The ESI-MS spectra of DNA mixture with compounds. The ESI-MS results shown here are deconvoluted spectra and molecular weights are shown with each peak. The inset in red box is the expansion of (B) between 9600 and 10 000 m/z.

and AAAGTGTTT. As shown in Fig. 5A, five free DNA peaks are shown for AAATTT (m/z 7303), AAAGTTT (m/z 8539), AAAGCTTT (m/z 7921), AAAGGTTT (m/z 9158), and AAAGTGTTT (m/z 9775). Upon addition of DB2831, the intensity of the peak for AAAGGTTT (m/z 9158) decreases with the simultaneous appearance of a new peak at m/z 9801 that is characteristic of a 1 : 1 AAAGGTTT–DB2831 complex (Fig. 5B and inset). There is no appearance of other DNA and ligand complex peaks. The observed spectra clearly indicate the high specificity and affinity of DB2831 for the GG sequence and validated the results described above.

Molecular curvature evaluation

Along with stacking surface, molecular curvature is a key feature for minor groove recognition. Appropriate curvature is essential for strong H-bonding and charge interactions in the minor groove. We have established a graphical approach for the determination of comparative molecular curvature values for minor groove binding compounds. In this method, compounds are first energy minimized in the SPARTAN software package. The compounds are then compared in a graphics package such as PowerPoint. A reference circle is defined that passes through both amidine carbons (Fig. 6). The reference circle that has a radius that allows it to pass as closely as possible through the center of each molecular unit of the entire molecule and the two amidine carbons is selected as illustrated with DB2830 and DB2831 (Fig. 6). Two straight lines are then drawn from the amidine carbons to the circle point at the center of the

molecule. The angle between these two lines then defines a relative curvature value for each molecule. The values are 110° for DB2830 and 136° for DB2831.

Analysis of a range of strong binding minor groove binding compounds by this method provides a calibration value of around 140° ±5° for compounds to bind strongly in the DNA minor groove. As can be seen, DB2831 falls in this range while DB2830 is too curved to make optimum contacts with the groove. While this is a relative comparison, it is a useful number to determine before synthesis of a new compound. It is also helpful in relative comparison of compound-minor groove binding constants. It should be noted that the optimum value for curvature is a range since both DNA and the compound can make structural changes to optimize the binding energetics on complex formation.

MD simulations of DB2831–DNA complexes

To help better understand the structural basis of molecular recognition of DNA sequences with two adjacent GC bps in an AT context, molecular dynamics (MD) simulations for a complex of the novel benzodiimidazole compound, DB2831, and a B-form ds[(5′-CCAAAGGTTTCC-3′) (5′-GGAAACCTTTGG-3′)] DNA with the target GG sequence were conducted (Fig. 7). Force constants for DB2831 were determined as described previously[17] and in the Methods section (all methods are in the ESI†) and added to the force field for the simulation. The 600 ns MD simulation was performed by using the Amber 16 software package in the presence of 0.15 M Na⁺ (total 37 Na⁺ have been
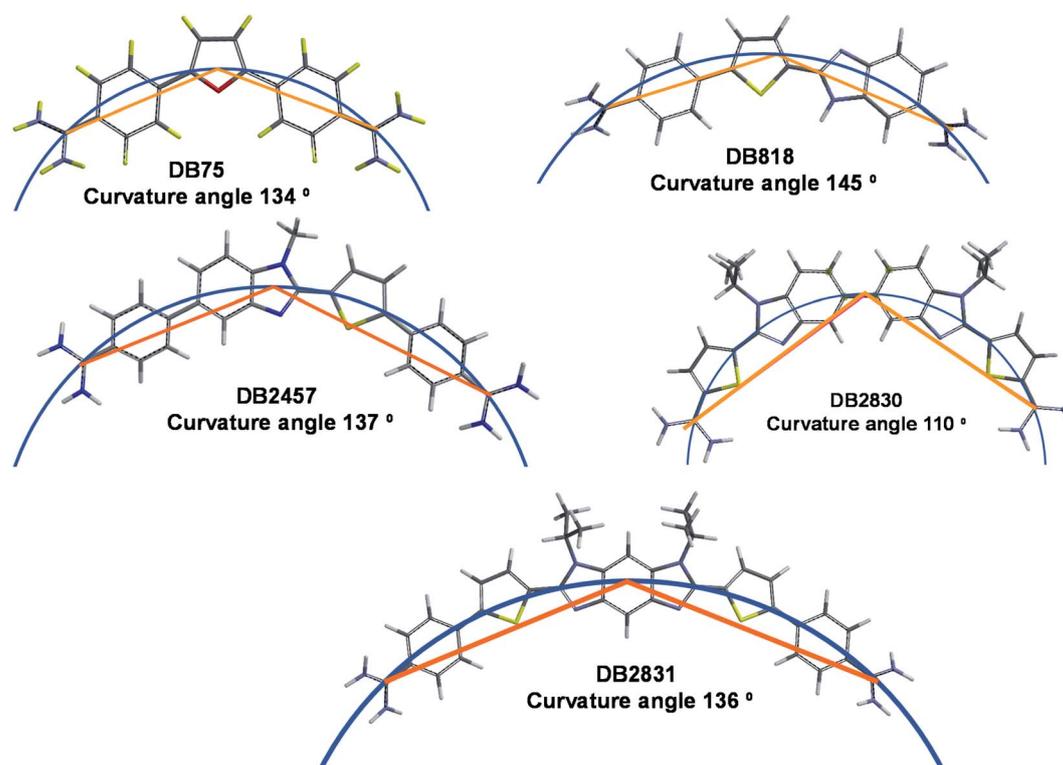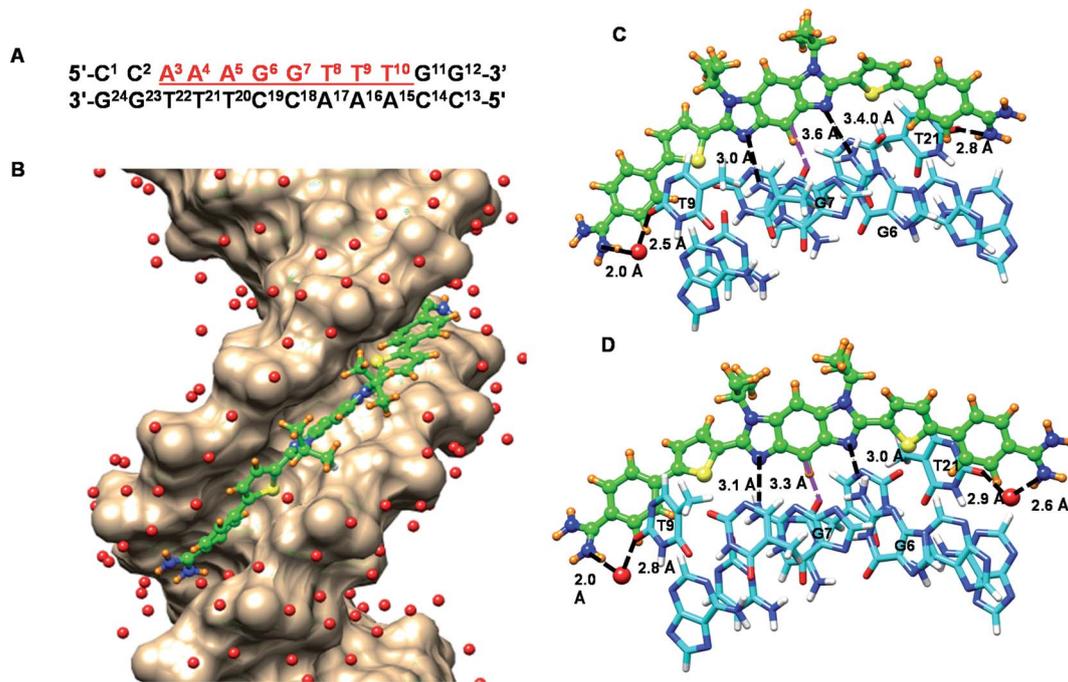

Fig. 6 Molecular curvature for selected minor groove binding compounds.

Fig. 7 (A) The DNA sequence used for MD analysis; red underlined bases indicate the binding site of the compounds; Molecular Dynamics (MD) model of DB2831 bound to an AAAGGTTT site; (B) a surface model viewed into the minor groove of the AAAGGTTT binding site with bound DB2831. The DNA bases are represented in tan- color scheme and DB2831 is light green-orange-blue-yellow (C–H–N–S) color scheme with the *N*-isopropyl groups facing out of the minor groove. The important interactions between different sections of the DB2831–DNA complex are illustrated in (C) and (D); (C) DB2831 forms three direct hydrogen bonds (black dashed lines) with DNA bases and one C–H interaction (purple dashed lines) with a DNA base. The direct interactions are (i) one of the terminal amidines with O2 of T21, (ii) and (iii) two central imidazole-Ns with two exocyclic H–N of G6 and G7. The central phenyl–C–H forms a C–H interaction with O2 of C19 (purple dashed lines). The other terminal amidine group forms an interfacial water mediated (red, ball and stick) H-bond with O2 of T9 (amidine–N–H–O–H_2–O–T9) to stabilize the compound in the minor groove. (D) DB2831 forms two direct hydrogen bonds (black dashed lines) with DNA bases and one C–H interaction (purple dashed lines) with a DNA base. The direct interactions are, two central imidazole-Ns with two exocyclic H–N of G6 and G7. The central phenyl–C–H forms a C–H interaction with O2 of C19 (purple dashed lines). The two terminal amidine groups form interfacial water mediated (red, ball and stick) H-bonds with O2 of T9 and T21 (amidine-N-H–O–H_2–O–T9 and amidine–N–H–O–H_2–O–T21) to stabilize the compound in the minor groove.

added) and to balance the excess Na$^+$, a total of 17 Cl$^-$ have been added (see Methods for details about the simulation procedure). Surprisingly the MD analysis revealed that the terminal amidine groups of DB2831 are very dynamic and three types of H-bond interactions can be formed with DNA. For about 10% of the simulation time, the two amidines form direct H-bonds to the O2 atoms of T9 and T21 with an average of 2.7–3.0 Å H-bond length. In most of the trajectory files (75%), however, one interfacial water molecule is observed. In this type of complex, one terminal amidine forms an interfacial water-mediated H-bond, amidine–water–DNA (O2 atoms of dT), and the other terminal amidine forms a direct H-bond with an O2 atom of dT at the opposite end of the complex (Fig. 7C). The amidines that form direct or water-mediated H-bonds switch in a dynamic process through the simulation. The remainder of the trajectories (15%) show two interfacial water molecules, where both terminal amidines make a connection with DNA through water-mediated H-bonding, amidine–water–DNA (Fig. 7D). The water molecules in the DB2831 binding site can adequately orient to provide favorable curvature to the DNA complex and interactions between the compound and DNA. The H-bonding ability,

flexibility, and dynamics of the bound waters help provide the high binding affinity of DB2831 to the -AAAGGTTT- binding site. The interfacial waters act as H-bond donors and acceptors to connect DB2831 and DNA bases noncovalently. Two other strong, with bonding distance ~3.0 Å, long lifetime H-bonds are formed by the two imidazole unprotonated *N* in the benzodiimidazole group of DB2831 with the central two adjacent dG-NHs that project into the minor groove (Fig. 7C and D). The strong G-NH–imidazole-N H-bonds provide high binding selectivity of DB2831 towards the AAAGGTTT sequence, which supports the ESI-MS results. Additional selectivity in binding is provided by the –CH (C2) group of the central core six-member ring of benzodiimidazole that points into the minor groove (Fig. 7C and D). This –CH forms a dynamic close interaction with the –dC=O of the dC19 base of the central G·C bps.

Phenyl –CH interactions with dT and dC –C=O and dA-N3 groups provide some additional stability for the ligand–DNA complex. No 180° rotational motions are observed for the phenyl groups of DB2831 throughout the 600 ns MD simulation, as expected for the optimum indexing of the compound. DB2831 tracks optimally along the minor groove with an

appropriate twist to match the minor groove curvature. The strong binding results from the H-bond network between the compound and DNA as well as electrostatic and van der Waals interactions. Extensive interactions are formed by the conjugated aromatic system of DB2831 with the sugar–phosphate walls of the minor groove. There is also an extensive terminal amidine–water network linking the compound to the floor of the minor groove.

Discussion

Because TFs have been difficult to target directly with small molecules, the idea of targeting TF promoters to inhibit protein–DNA complex formation with selectively engineered organic modules is an innovative alternative.[3] We have focused on the important ETS family of TFs as a test system for use of a heterocyclic diamidine molecular platform for the design and synthesis of TF inhibitors. We have previously shown that AT selective minor groove binders can inhibit HMGA minor groove DNA-binding proteins as well as the HOXA9 TF which play important roles in tumorigenesis, adipogenesis and acute myeloid leukemia.[18] The inhibition of protein–DNA binding in these systems was detected with a biosensor-surface plasmon resonance assay.[19]

The ETS family is attractive for inhibitor development because many members of the family are well-characterized and have important functions in cell biology and human diseases. Several key promoters of the PU.1 ETS TF have AT sequences on the 5′ side of the -GGAA- central, conserved recognition site.[7] The AT sequence is targeted by many known minor groove binders from netropsin to synthetic heterocyclic cations such as Hoechst dyes and heterocyclic diamidines. In an exciting development with new synthetic diamidines that have extended AT recognition sequences, we have found PU.1 inhibitors active at the cellular level, including against AML cells, as well as against an animal model of AML.[3,7] With this approach to disease treatment, the application potential of the heterocyclic diamidine platform is greatly extended. To reach the full potential of diamidines, however, it is essential to expand their sequence recognition capability past pure AT sequences. To do this an entirely new class of minor groove binders that can recognize mixed base-pair sequences, including the -ETS GGAA promoter sequence[2] is required. To accomplish this goal, we have initiated a project to add new mixed bps DNA binding motifs with variations in solubility, chemical properties, and cell uptake properties to provide the best chance of successful cellular TF inhibition.

Our design motif is built around heterocyclic diamidines that have good solubility, cell uptake, and reasonable synthesis. DB2429, DB2457 (Fig. 1A), and analogs were successful thiophene compounds in recognition of AT sequences with a single G·C bp.[2] In those compounds, the thiophene-*N*-alkyl-BI motif formed a preorganized sigma-hole stabilized conformation for minor groove-specific binding and that concept has formed the basis of the new compounds described here.[2]

With the successful preparation of new synthetic agents that recognize the AT sequence of the PU.1 promoter,[7] the next most important sequence for the design of new compounds was the central, conserved 5′-GGAA-3′ and closely flanking regions. The key to strong selective binding to this sequence for minor groove agents is the GG unit that has proven very difficult to target. For this recognition, two GC binding modules must be linked very close together. After evaluation of all of the successful, single G·C bp binding modules previously prepared, the seven compounds of Fig. 8 (from DB2818 to DB2824) linked through the benzimidazole nitrogen that faces away from the amidine were prepared. The goal with these compounds was to maintain the alkyl-benzimidazole-thiophene G·C bp recognition module in an arrangement that could stack together in the minor groove to bind to the two adjacent G·C bps unit with flanking A·T bps recognized by the terminal, substituted phenyl groups. Surprisingly, all of these compounds had very poor solution properties with extensive aggregation in aqueous solutions that prevented their successful use. Extensive compound and solution variations were not successful in providing useful non-aggregated compounds.

With the failure of this design concept, new ideas were evaluated and DB2830 with two alkyl-BI thiophene amidines in close proximity was prepared (Fig. 1B) for adjacent G·C bp binding. Although the compound was the first of our diamidines to successfully recognize adjacent GC bps, the binding was weaker than desired for cellular use. As noted in the results section, the compound is too curved for best fit to the minor groove shape and new compound structures were designed. DB2831 is a more extended and less curved structure than DB2830. There are no known benzodiimidazole minor groove binders, but modeling studies suggested that group could be a key element in minor groove binding, particularly at GG sites. A successful synthetic strategy was designed for DB2831, and analogs (Fig. 1B) and the compound displayed excellent affinity and specificity for the target 5′-AGGAA sequence of the PU.1 promoter. As described in the Results section, DB2831 had close to the ideal curvature for minor groove binding and represents a breakthrough in our design efforts.

A summary of the primary experimental results for DB2831 includes a high ΔT_m with the test -AAGGTT- sequence. Biosensor SPR studies support the strong binding of DB2831 to the test sequence with a K_D of (2 ± 2) nM. This value is confirmed by a similar K_D from fluorescence titration experiments. Competition mass spectrometry results also support strong binding and very clearly demonstrate the excellent binding selectivity of DB2831 to -AAAGGTTT-. A minor groove binding mode is indicated by CD and modeling studies. The minor groove binding is also expected from the compound structure and DNA sequences to which it binds.

MD analysis of DB2831 bound to an AAAGGTTT site reveals some very interesting features of the complex that are difficult to obtain from experimental analysis. The compound fits well between the walls of the minor groove and is able to twist to match the groove curvature (Fig. 7). The fit to the floor of the minor groove is more complex. The two unprotonated *N* of the *N*-isopropyl-benzodiimidazole group form strong H-bonds (based on distance) with the two central dG-NH (dG6 and dG7) that project into the minor groove. In addition, the central
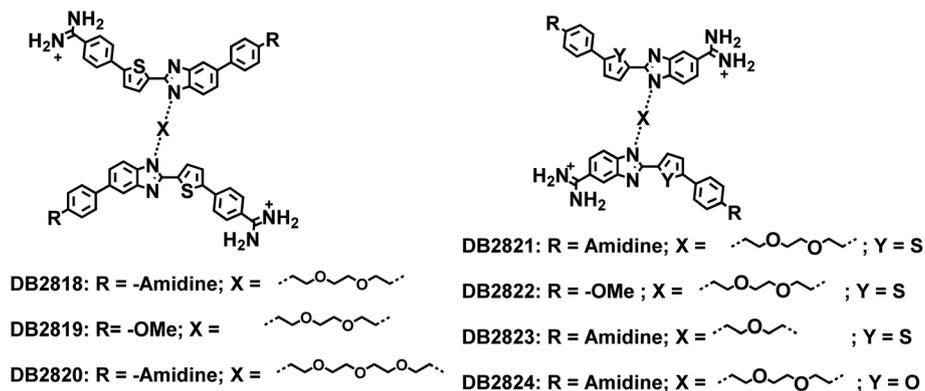
Fig. 8 Two benzimidazoles linked compounds with the goal to form a stacked complex to bind two adjacent GC bps.

–CH of the benzodiimidazole that faces into the groove is close to the –C=O of dC19 which H-bonds to dG6 and forms a stabilizing interaction. These three interactions between the benzodiimidazole and two G·C bps account for the strong preference of DB2831 with the minor groove of a -GG- sequence. Through the MD simulation, the benzodiimidazole and two thiophene groups remain in close proximity to the floor of the minor groove. The two sulfur atoms of the thiophene are an average of 3.3 ± 0.2 A from the floor of the groove and the two AT base pairs that are adjacent to the -GG- sequence. As shown in the models in Fig. 7, the terminal phenyl amidines of DB2831 are much more dynamic than the thiophene–benzodiimidazole–thiophene center of the bound molecule. A complex without any interfacial water involvement is formed with two inner facing amidine –NH groups forming H-bonds with the dT=O of dT9 and dT21. While this complex would seem to be the optimum, surprisingly it is found in only approximately 10% of the simulation. It clearly is not the minimum Gibbs energy. A complex with one interfacial water with a dynamic amidine (–NH)–water–dT=O interaction is the most favored with a 75% occupancy. The final complex has interfacial water molecules at both amidine groups to link the amidines to –dT=O groups. This complex has approximately 15% occupancy and is closer in Gibbs energy to the complex with no interfacial water. Surprisingly, the central thiophene-benzodiimidazole-thiophene center of the complex remains in a very stable position throughout the simulation. The flexibility to allow 0, 1, or 2 interfacial waters of interaction comes about due to single bond rotations of the bonds linking the terminal phenyl groups to the thiophene and amidine groups. It seems clear that dynamic, terminal interfacial water molecules can cost a minimum amount of entropy of complex formation while allowing stronger compound–DNA interactions than in their absence. A challenge for drug design will be to determine how to incorporate this type of dynamic water interaction into design efforts.

Previously, two of the N–MeBI-thiophene-phenyl units were linked with alkyl chains to create DB2528 (Fig. 1A) and analogs for the desired recognition of two GC base pairs separated by AT base pairs.[2] DB2528 strongly and specifically binds to the target sequence, AGAAACA, in agreement with the length of the three-methylene linker. Using the curvature procedure described in the Results section, we obtain a value of 124° for DB2528. DB2528, however, binds to the -AAGAAACTT binding site very strongly with a K_D of 5 nm. The shape and curvature of DB2528 clearly match the DNA minor groove to allow the N–MeBI-thiophene-phenyl units to bind to both G·C bps. This is unique to compounds with a structure like DB2528: (i) all four H-bonding groups (BI acceptor for G-NH2 and amidine donor for AT H-bonds) are at the periphery of the molecular structure and interact strongly with the floor of the minor groove; (ii) the central section of the molecule, thiophene-phenyl-linker essentially stacks with the walls of the minor groove, and this is not as distance-dependent as H-bond formation. In addition, the flexibility of the central linker allows the compound to twist to match the minor groove shape and can also alter the molecular curvature calculated by our procedure based on a rigid molecular structure. Although DB2528 has a relatively large size, its excellent solubility and fluorescence properties make it an attractive compound, along with DB2831 for recognition of two G·C bps in an AT context.

Conclusions

In summary, the binding of DB2528 and the favorable water-mediated binding of DB2831 show very clearly that curvature alone does not predict the strength of a complex that is optimum for minor groove binding. Other factors, which are now being defined, are more important for the final complex. DB2528 for 3′-site recognitions of the PU.1 promoter along with AT specific minor groove binders for 5′-site recognition now combine with DB2831 for complete recognition of the promoter sequence for the PU.1 TF.

Data availability

The MD trajectory results described in the paper are available from the authors.

Author contributions

WDW, PG, and AP designed the study and experiments. AAF, WDW and DWB designed the novel compound syntheses. WDW, PG, AP and AAF wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 (a) S. Neidle, *Nat. Prod. Rep.*, 2001, **18**, 291; (b) A. Rahman, P. O'Sullivan and I. Rozas, *Medchemcomm*, 2018, **10**, 26; (c) Z. Yu, G. N. Pandian, T. Hidaka and H. Sugiyama, *Adv. Drug Delivery Rev.*, 2019, **147**, 66; (d) M. I. Sánchez, O. Vázquez, J. Martínez-Costas, M. E. Vázquez and J. L. Mascareñas, *Chem. Sci.*, 2012, **3**, 2383; (e) M. P. Barrett, C. G. Gemmell and C. J. Suckling, *Pharmacol. Ther.*, 2013, **139**, 12; (f) C. Bailly and M. J. Waring, *Nucleic Acids Res.*, 1998, **26**, 4309; (g) A. C. Finlay, F. A. Hochstein, B. A. Sobin and F. X. Murphy, *J. Am. Chem. Soc.*, 1951, **73**, 341; (h) C. Zimmer and U. Wähnert, *Prog. Biophys. Mol. Biol.*, 1986, **47**, 31; (i) M. F. Paine, M. Z. Wang, C. N. Generaux, D. W. Boykin, W. D. Wilson, H. P. De Koning, C. A. Olson, G. Pohlig, C. Burri, R. Brun, G. A. Murilla, J. K. Thuita, M. P. Barrett and R. R. Tidwell, *Curr. Opin. Invest. Drugs*, 2010, **11**, 876; (j) B. G. Kim, H. M. Evans, D. N. Dubins and T. V Chalikian, *Biochemistry*, 2015, **54**, 3420; (k) I. Dilek, M. Madrid, R. Singh, C. P. Urrea and B. A. Armitage, *J. Am. Chem. Soc.*, 2005, **127**, 3339; (l) J. Mosquera, M. I. Sánchez, M. E. Vázquez and J. L. Mascareñas, *Chem. Commun.*, 2014, **50**, 10975.

2 (a) Y. Chai, A. Paul, M. Rettig, W. D. Wilson and D. W. Boykin, *J. Org. Chem.*, 2014, **79**, 852; (b) A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M. H. David-Cordonnier, D. W. Boykin and W. D. Wilson, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 4927; (c) P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin and W. D. Wilson, *Chem.–Eur. J.*, 2016, **22**, 15404; (d) P. Guo, A. Paul, A. Kumar, N. K. Harika, S. Wang, A. A. Farahat, D. W. Boykin and W. D. Wilson, *Chem. Commun.*, 2017, **53**, 10406; (e) P. Guo, A. A. Farahat, A. Paul, N. K. Harika, D. W. Boykin and W. D. Wilson, *J. Am. Chem. Soc.*, 2018, **140**, 14761; (f) A. A. Farahat, P. Guo, H. Shoeib, A. Paul, D. W. Boykin and W. D. Wilson, *Chem.–Eur. J.*, 2020, **26**, 4539; (g) P. Guo, A. A. Farahat, A. Paul, A. Kumar, D. W. Boykin and W. D. Wilson, *Biochemistry*, 2020, **59**, 1756.

3 I. Antony-Debré, A. Paul, J. Leite, K. Mitchell, H. M. Kim, L. A. Carvajal, T. I. Todorova, K. Huang, A. Kumar, A. A. Farahat, B. Bartholdy, S. R. Narayanagari, J. Chen, A. Ambesi-Impiombato, A. A. Ferrando, I. Mantzaris, E. Gavathiotis, A. Verma, B. Will, D. W. Boykin, W. D. Wilson, G. M. K. Poon and U. Steidl, *J. Clin. Invest.*, 2017, **127**, 4297.

4 (a) X. Jiang and Z. Yang, *OncoTargets Ther.*, 2018, **11**, 3533; (b) M. Lambert, S. Jambon, S. Depauw and M. H. David-Cordonnier, *Molecules*, 2018, **23**, 1479; (c) S. Depauw, M. Lambert, S. Jambon, A. Paul, P. Peixoto, R. Nhili, L. Marongiu, M. Figeac, C. Dassi, C. Paul-Constant, B. Billoré, A. Kumar, A. A. Farahat, M. A. Ismail, E. Mineva, D. P. Sweat, C. E. Stephens, D. W. Boykin, W. D. Wilson and M. H. David-Cordonnier, *J. Med. Chem.*, 2019, **62**, 1306.

5 (a) J. E. Darnell, *Nat. Rev. Cancer*, 2002, **2**, 740; (b) A. N. Koehler, *Curr. Opin. Chem. Biol.*, 2010, **14**, 331.

6 (a) B. Will, T. O. Vogler, S. Narayanagari, B. Bartholdy, T. I. Todorova, M. F. da Silva, J. Chen, Y. Yu, J. Mayer, L. Barreyro, L. Carvajal, D. B. Neriah, M. Roth, J. van Oers, S. Schaetzlein, C. McMahon, W. Edelmann, A. Verma and U. Steidl, *Nat. Med.*, 2015, **21**, 1172; (b) B. U. Mueller, T. Pabst, J. Fos, V. Petkovic, M. F. Fey, N. Asou, U. Buergi and D. G. Tenen, *Blood*, 2006, **107**, 3330; (c) N. Bonadies, T. Pabst and B. U. Mueller, *Blood*, 2010, **115**, 331.

7 (a) 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; (b) T. H. Pham, J. Minderjahn, C. Schmidl, H. Hoffmeister, S. Schmidhofer, W. Chen, G. Längst, C. Benner and M. Rehl, *Nucleic Acids Res.*, 2013, **41**, 6391.

8 A. A. Farahat, C. Bennett-Vaughn, E. M. Mineva, A. Kumar, T. Wenzler, R. Brun, Y. Liu, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 5907.

9 (a) A. A. Farahat, A. Kumar, M. Say, A. E.-D. M. Barghash, F. E. Goda, H. M. Eisa, T. Wenzler, R. Brun, Y. Liu, L. Mickelson, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem.*, 2010, **18**, 557; (b) D. Branowska, A. A. Farahat, T. Wenzler, R. Brun, Y. Liu, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem.*, 2010, **18**, 3551; (c) C. S. Reid, A. A. Farahat, X. Zhu, T. Pandharkar, D. W. Boykin and K. A. Werbovetz, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 6806; (d) A. A. Farahat and D. W. Boykin, *Heterocycles*, 2012, **85**, 2437; (e) A. A. Farahat and D. W. Boykin, *J. Heterocycl. Chem.*, 2013, **50**, 585.

10 (a) A. Paul, A. Kumar, R. Nanjunda, A. A. Farahat, D. W. Boykin and W. D. Wilson, *Org. Biomol. Chem.*, 2016, **15**, 827; (b) A. A. Farahat, M. A. Ismail, A. Kumar, T. Wenzler, R. Brun, A. Paul, W. D. Wilson and D. W. Boykin, *Eur. J. Med. Chem.*, 2018, **143**, 1540; (c) A. A. Farahat, S. Iwamoto, M. Roche and D. W. Boykin, *J. Heterocycl. Chem.*, 2021, 1–7.

11 A. A. Farahat, E. Paliakov, A. Kumar, A.-E. M. Barghash, F. E. Goda, H. M. Eisa, T. Wenzler, R. Brun, Y. Liu, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem.*, 2011, **19**, 2156.

12 B. Nguyen, F. A. Tanious and W. D. Wilson, *Methods*, 2007, **42**, 150.

13 (a) P. L. DeHaseth, T. M. Lohman and M. T. Record Jr, *Biochemistry*, 1977, **16**, 4783; (b) W. D. Wilson, C. R. Krishnamoorthy, Y. H. Wang and J. C. Smith, *Biopolymers*, 1985, **24**, 1941; (c) T. M. Lohman, P. L. deHaseth and M. T. Record Jr, *Biophys. Chem.*, 1978, **8**, 281; (d) S. Wang, A. Kumar, K. Aston, B. Nguyen,

J. K. Bashkin, D. W. Boykin and W. D. Wilson, *Chem. Commun.*, 2013, **49**, 8543.

14 (*a*) N. Salim, R. Lamichhane, R. Zhao, T. Banerjee, J. Philip, D. Rueda and A. L. Feig, *Biophys. J.*, 2012, **102**, 1097; (*b*) P. L. Privalov, A. I. Dragan and C. Crane-Robinson, *Nucleic Acids Res.*, 2011, **39**, 2483.

15 L. H. Fornander, L. Wu, M. Billeter, P. Lincoln and B. Norden, *J. Phys. Chem. B*, 2013, **117**, 5820.

16 S. Laughlin and W. D. Wilson, *Int. J. Mol. Sci.*, 2015, **524**, 506.

17 (*a*) J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247; (*b*) N. K. Harika, M. W. Germann and W. D. Wilson, *Chem.–Eur. J.*, 2017, **23**, 17612; (*c*) P. Athri and W. D. Wilson, *J. Am. Chem. Soc.*, 2009, **131**, 7618; (*d*) N. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza and J. Šponer, *J. Am. Chem. Soc.*, 2003, **125**, 1759.

18 L. Su, N. Bryan, S. Battista, J. Freitas, A. Garabedian, F. D'Alessio, M. Romano, F. Falanga, A. Fusco, L. Kos, J. Chambers, F. Fernandez-Lima, P. P. Chapagain, S. Vasile, L. Smith and F. Leng, *Sci. Rep.*, 2020, **10**, 18850.

19 Y. Miao, T. Cui, F. Leng and W. D. Wilson, *Anal. Biochem.*, 2008, **374**, 7.