BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

The Unusual Monomer Recognition of Guanine-Containing Mixed Sequence DNA by a Dithiophene Heterocyclic Diamidine

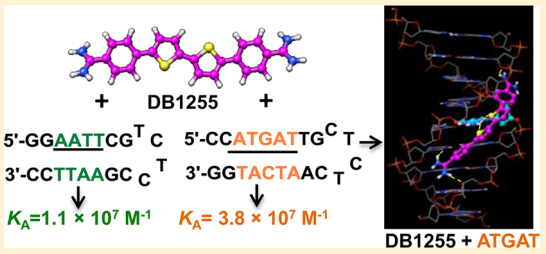
Manoj Munde,[†,§] Arvind Kumar,[†] Paul Peixoto,[‡,∥] Sabine Depauw,[‡] Mohamed A. Ismail,[†,⊥] Abdelbasset A. Farahat,[†,@] Ananya Paul,[†] Martial V. Say,[†] Marie-Hélène David-Cordonnier,*[,‡] David W. Boykin,*[,†] and W. David Wilson*[,†]

[†]Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, Georgia 30303-3083, United States

[‡]INSERM U837-JPARC (Jean-Pierre Aubert Research Center), Team "Molecular and Cellular Targeting for Cancer Treatment", Université Lille Nord de France, IMPRT-IFR-114, Institut pour la Recherche sur le Cancer de Lille, Place de Verdun, F-59045 Lille Cedex, France

Supporting Information

ABSTRACT: DB1255 is a symmetrical diamidinophenyl-dithiophene that exhibits cellular activity by binding to DNA and inhibiting binding of ERG, an ETS family transcription factor that is commonly overexpressed or translocated in leukemia and prostate cancer [Nhili, R., Peixoto, P., Depauw, S., Flajollet, S., Dezitter, X., Munde, M. M., Ismail, M. A., Kumar, A., Farahat, A. A., Stephens, C. E., Duterque-Coquillaud, M., Wilson, W. D., Boykin, D. W., and David-Cordonnier, M. H. (2013) *Nucleic Acids Res. 41*, 125−138]. Because transcription factor inhibition is complex but is an attractive area for anticancer and antiparasitic drug development, we have evaluated the DNA interactions of additional derivatives of DB1255 to gain an improved understanding of the biophysical chemistry of complex function and inhibition. DNase I footprinting, biosensor surface plasmon resonance, and circular dichroism experiments show that DB1255 has an unusual and strong monomer binding mode in minor groove sites that contain a single GC base pair flanked by AT base pairs, for example, 5′-ATGAT-3′. Closely related derivatives, such as compounds with the thiophene replaced with furan or selenophane, bind very weakly to GC-containing sequences and do not have biological activity. DB1255 is selective for the ATGAT site; however, a similar sequence, 5′-ATGAC-3′, binds DB1255 more weakly and does not produce a footprint. Molecular docking studies show that the two thiophene sulfur atoms form strong, bifurcated hydrogen bond-type interactions with the G-N-H sequence that extends into the minor groove while the amidines form hydrogen bonds to the flanking AT base pairs. The central dithiophene unit of DB1255 thus forms an excellent, but unexpected, single-GC base pair recognition module in a monomer minor groove complex.



With our improving understanding of the critical role of functional control sequences in DNA from projects such as ENCODE, it is now clear that the opportunities for selective targeting of specific and important, but nontranscribed, sequences of DNA offer exciting options for gene control.[1] Such control possibilities have potential in enhancing our understanding of gene expression mechanisms as well as therapeutic development that could not be imagined even a short time ago. The recent evidence that G-quadruplex DNA could be selectively targeted,[2,3] as well as the excellent clinical anti-infective and anticancer activity of minor groove binding agents,[4] provides encouraging examples of the therapeutic possibilities in this area. To allow new gene control mechanisms, it is essential to find agents that target a specific DNA sequence and exert a desired biological response, such as inhibition or activation of a critical transcription factor. At the same time, binding the active compounds to other DNA sequences should not cause significant biological effects in the same concentration range. Selective activity can be accomplished with molecules that are not particularly large, so that they maintain cell uptake and are reasonable in cost and effort to synthesize on a large, clinical scale.[5−7] While compounds that meet these requirements are not common, they do exist, but unfortunately, the rules for their design are essentially unknown. Emerging information about local DNA microstructural states[8−10] may help unravel some keys to the selective biological action of effective DNA-targeting small molecules.

Initial steps for success in developing agents that function, as described above, require design and synthesis of new DNA-binding compounds and evaluating their DNA affinity and selectivity. A compound that has recently been shown to exhibit effects quite close to the desired *in vivo* behavior is dithiophene

Received: November 26, 2013
Revised: February 3, 2014
Published: February 4, 2014

   dx.doi.org/10.1021/bi401582t | *Biochemistry* 2014, 53, 1218−1227

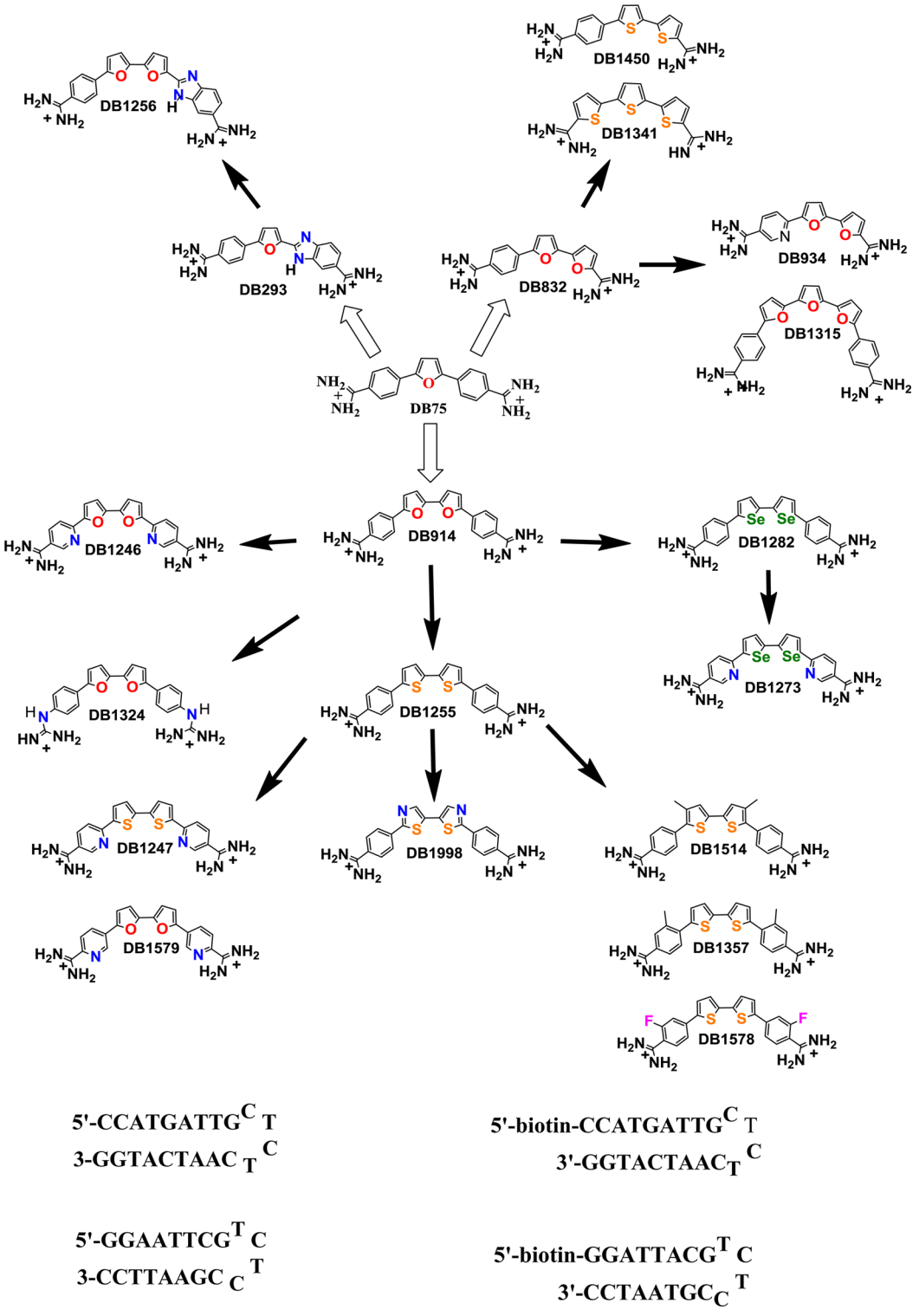


Figure 1. DNA oligomer sequences and compounds used in this study.

DB1255 (Figure 1).[11] The compound targets the DNA binding of the ERG protein, an ETS transcription factor that is commonly overexpressed or translocated in leukemia and prostate carcinoma.[12−16] This is a promising finding because transcription factors have generally been considered undruggable and direct modulation of gene expression by targeting oncogenic transcription factors is an exciting new area of development for cancer treatment. The inhibition of binding of ERG to DNA was discovered in a screening assay of synthetic DNA binding compounds and was verified by electrophoretic mobility shift assays (EMSAs). Additional EMSA analysis defined the precise DNA-binding sequence required for optimal binding of dithiophene DB1255 and thus for efficient inhibition of the ERG−DNA complex. It is very important that biological assays confirmed that DB1255 could modulate transcription factor binding in cells.[11] Thus, relatively small molecules can specifically target DNA-transcription factor sites, such as the ERG−DNA recognition site, both *in vitro* and in cells.

dx.doi.org/10.1021/bi401582t | *Biochemistry* 2014, 53, 1218−1227

Given the importance of the cellular activity and transcription factor inhibition of DB1255, as well as our lack of detailed knowledge about the biophysical chemistry of the interaction of the compound and its derivatives with DNA, the experiments reported here were initiated. Dithiophene DB1255 is unusual for heterocyclic diamidines that target the DNA minor groove in that it can bind to GC-containing sequences, and this is essential for its inhibition of the ERG transcription factor. It was not clear from the initial studies whether DB1255 binds as a monomer, as with the most similar AT specific minor groove binders, or as a dimer as observed with the furan derivative, DB293, in mixed AT- and GC-containing sequences (Figure 1). It was also not clear what structural features and chemical properties of DB1255 are necessary for its recognition of mixed DNA sequences. To address these key points, spectroscopy, thermal melting, DNase I footprinting, biosensor surface plasmon resonance (SPR), and molecular modeling were used to evaluate the interactions of DB1255 and its analogues with both pure AT-containing and GC-containing (Figure 1) DNA sequences. Very surprisingly, we have found that, unlike DB293 and synthetic polyamides that bind to mixed DNA sequences as dimers, DB1255 binds to certain GC-containing sequences in the minor groove as a monomer, and the binding is very dependent on the dithiophene group.

MATERIALS AND METHODS

DNAs, Compounds, and Buffers. The syntheses of DB75, DB832, DB914, DB1255, DB1282, DB1341, DB1450, and DB1998 (Figure 1) have been previously described,[17,18] and the scheme and experimental details for DB2297 are given in the Supporting Information. Their purity was verified by nuclear magnetic resonance and elemental analysis. Concentrated stock solutions (1 mM) of compounds were prepared in water. Solutions of the compounds for biosensor surface plasmon resonance (SPR) and spectroscopic studies were prepared by dilution with 0.01 M cacodylic buffer (pH 6.25) with 0.001 M EDTA and 0.1 M NaCl. SPR binding studies were conducted with 5′-biotinated DNAs as previously described,[19,20] while spectroscopic studies were performed with non-biotin-labeled DNAs (Figure 1). The concentration of the DNA solutions was determined spectrophotometrically at 260 nm using extinction coefficients per nucleotide for 5′-CCATGATTGCTCTCAATCATGG-3′ and 5′-GGAATTCGTCTCCGAATTCC-3′. The extinction coefficients were calculated by the nearest-neighbor method.[21]

DNase I Footprinting. Experiments were performed as previously described[11] using a 265 bp 3′-end-labeled DNA fragment obtained from *Eco*RI and *Pvu*II double digestion of the pBS plasmid (Stratagene, La Jolla, CA) followed by radiolabeling using [γ-^{32}P]dATP (Perkin-Elmer) and the Klenow enzyme for 30 min. The 265 bp radiolabeled DNA fragment was then purified on a 10% polyacrylamide gel under native conditions. Increasing concentrations of the various ligands were incubated with the radiolabeled DNA fragment for 15 min at 37 °C to ensure equilibrium prior to digestion for 3 min upon addition of DNase I (0.01 unit/mL) in 20 mM NaCl, 2 mM $MgCl_2$, and 2 mM $MnCl_2$ (pH 7.3). The DNA samples were then precipitated, heated at 90 °C for 4 min in loading denaturing buffer, and chilled in ice prior to being loaded on an 8% denaturing polyacrylamide gel for 90 min at 65 W in TBE buffer. The data were collected using a Phosphor Imager and analyzed using ImageQuant as previously described.[11] Each resolved band was assigned to a particular base of the DNA fragment by comparison of its position relative to the guanine sequencing standard (G-track) obtained using DMS/piperidine treatment of the 265 bp DNA fragment.

Thermal Melting (T_m). T_m experiments were conducted with a Cary 300 UV−visible spectrophotometer in 1 cm quartz cuvettes. The absorbance of the DNA−compound complex was monitored at 260 nm as a function of temperature, and DNA without the compound was used as a control. Cuvettes were mounted in a thermal block, and the solution temperatures were monitored by a thermistor in a reference cuvette with a computer-controlled heating rate of 0.5 °C/min. Experiments were generally conducted at a hairpin duplex oligomer concentration of 3×10^{-6} M. For experiments with complexes, a ratio of two compounds per oligomer duplex was generally used.

CD Spectroscopy. A 1 cm path length cell was used, and all experiments were conducted at 25 °C. The DNA hairpin duplexes (3×10^{-6} M per hairpin duplex) were titrated with increasing concentrations of compound. The resulting ratios increased from 0.25 to 2.0 (moles of compound to moles of DNA duplex). The experiments were performed in cacodylic acid buffer (pH 6.25). The sensitivity was set at 1 mdeg, and the scan speed was set at 50 nm/min. Four scans were recorded and averaged by the computer for each titration point.

Surface Plasmon Resonance (SPR). SPR measurements were performed with a four-channel Biacore 2000 optical biosensor system. The 5′-biotin-labeled DNA samples were immobilized onto streptavidin-coated sensor chips (Biacore SA) as previously described.[19,20] Three flow cells were used to immobilize the DNA oligomer samples, while a fourth cell was left blank as a control. The SPR experiments were performed at 25 °C in filtered, degassed, 10 mM cacodylic acid buffer (pH 6.25) containing 100 mM NaCl and 1 mM EDTA. Steady-state binding analysis was performed with multiple injections of different compound concentrations over the immobilized DNA surface at a flow rate of 25 μL/min and 25 °C. Solutions with known ligand concentrations were injected through the flow cells until a constant steady-state response was obtained. Compound solution flow was then replaced by buffer flow, resulting in dissociation of the complex. The reference response from the blank cell was subtracted from the response in each cell containing DNA to give a signal (RU, response units) that is directly proportional to the amount of bound compound. The predicted maximal response per bound compound in the steady-state region (RU_{max}) was determined from the DNA molecular weight, the amount of DNA on the flow cell, the compound molecular weight, and the refractive index gradient ratio of the compound and DNA, as previously described.[22] The reference response from the blank cell was subtracted from the response in each cell containing DNA to give a signal (RU, resonance units) that is directly proportional to the amount of bound compound. The stoichiometry of the reaction can be calculated as follows:

$$r = \text{moles of bound ligand/moles of DNA or where } r = RU/RU_{max}$$

where RU is the observed (experimental) response in the plateau region and RU_{max} is the predicted maximal response for a single small molecule binding to a nucleic acid site. Dividing the observed steady-state response RU by RU_{max} yields a stoichiometry-normalized binding isotherm. The calculated

dx.doi.org/10.1021/bi401582t | *Biochemistry* 2014, 53, 1218−1227

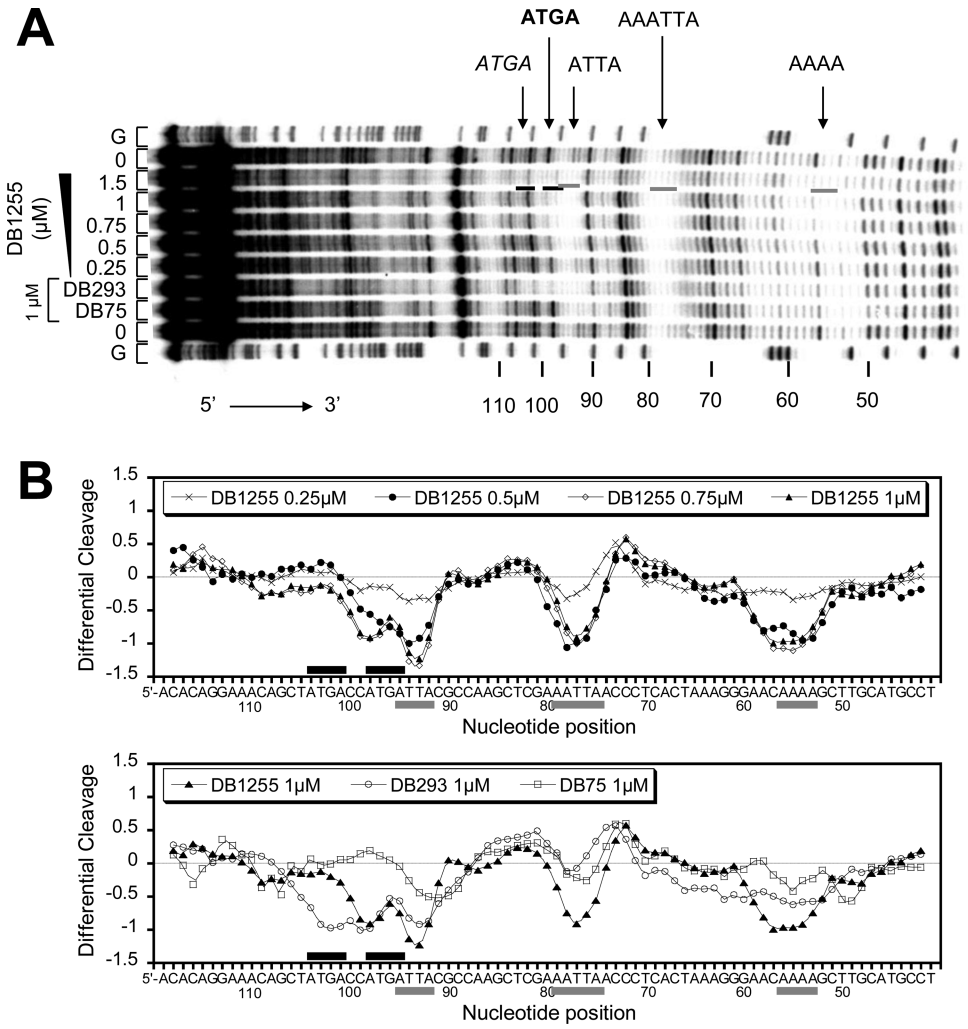


Figure 2. DNase I footprinting titration experiments that aimed to examine the binding of DB1255 to DNA. (A) Denaturing polyacrylamide gel of the 265 bp 3′-end-radiolabeled DNA fragments incubated with the various tested compounds and (B) the corresponding densitometric analysis. Gray boxes indicate AT-rich tracks of the indicated sequences. Black boxes indicate the ATGA site specific for DB293 (italic) or specifically recognized by DB1255 (bold). G indicates a G-track was performed to localize guanines in the DNA fragment and thus deduce the position of each base within the known 265 bp sequence.

value (r) was fit to an appropriate binding model, either a single-site model ($K_2 = 0$) or a two-site model:

$$\mathrm{RU}/\mathrm{RU}_{\text{pred-max}} = r = (K_1C_{\text{free}} + 2K_1K_2C_{\text{free}}) / (1 + K_1C_{\text{free}} + 2K_1K_2C_{\text{free}} \ldots \quad (1)$$

where K_1 and K_2 are macroscopic binding constants and C_{free} is the free compound concentration in equilibrium with the complex (the concentration in the flow solution).[20]

Molecular Docking. Dithiophene DB1255 and difuran DB914 were optimized at the B3LYP/6-31G* level of theory using Spartan 10.[23] All minimized ligands were assigned Gasteiger−Hückel charges by using Autodock version 4.02.[24] A DNA duplex sequence of d[(5′-CCATGATCT-3′)(5′-AGATCATGG-3′)], based on footprinting results, was generated from the Biopolymer-build DNA double helix from the Tripos SYBYL-X1.2 software package.[25] The modeled double-stranded DNA (ds-DNA) was then docked with the optimized ligand using Autodock version 4.02.[26]

The center of the macromolecule is the grid center with a grid size of 20 Å × 25 Å × 33 Å and a grid spacing of 0.375 Å. Docking runs were performed using the Lamarckian genetic algorithm (LGA) with no modifications of docking parameters. LGA was used because of the existence of rotatable bonds in the ligands and to evaluate the correct conjugate DNA conformation, as it is known to reproduce various experimental ligand−DNA complex structures. Initially, we used a population of random individuals (population size of 150), a maximal number of 2500000 energy evaluations, a maximal number of evaluations of 2700, and a mutation rate of 0.02 fs. Fifty independent flexible docking runs were conducted for each ligand, and then the lowest-energy dock conformation obtained from the flexible docking was resubmitted for rigid docking to remove the internal energy of the ligand (steric clashes) and retain the hydrogen bonding interaction with ds-DNA bases.

RESULTS

DNase I Footprinting. To evaluate the optimal binding sites for DB1255 and analogues on DNA and to gain insight into the precise bases that are crucial for DNA complex formation, a DNase I footprint assay was conducted. The DNA restriction fragment used is one that has previously been tested with several heterocyclic diamidines.[6] Footprinting gel results are shown in Figures 2A and 3A−C and Figure S1A of the

dx.doi.org/10.1021/bi401582t | Biochemistry 2014, 53, 1218−1227

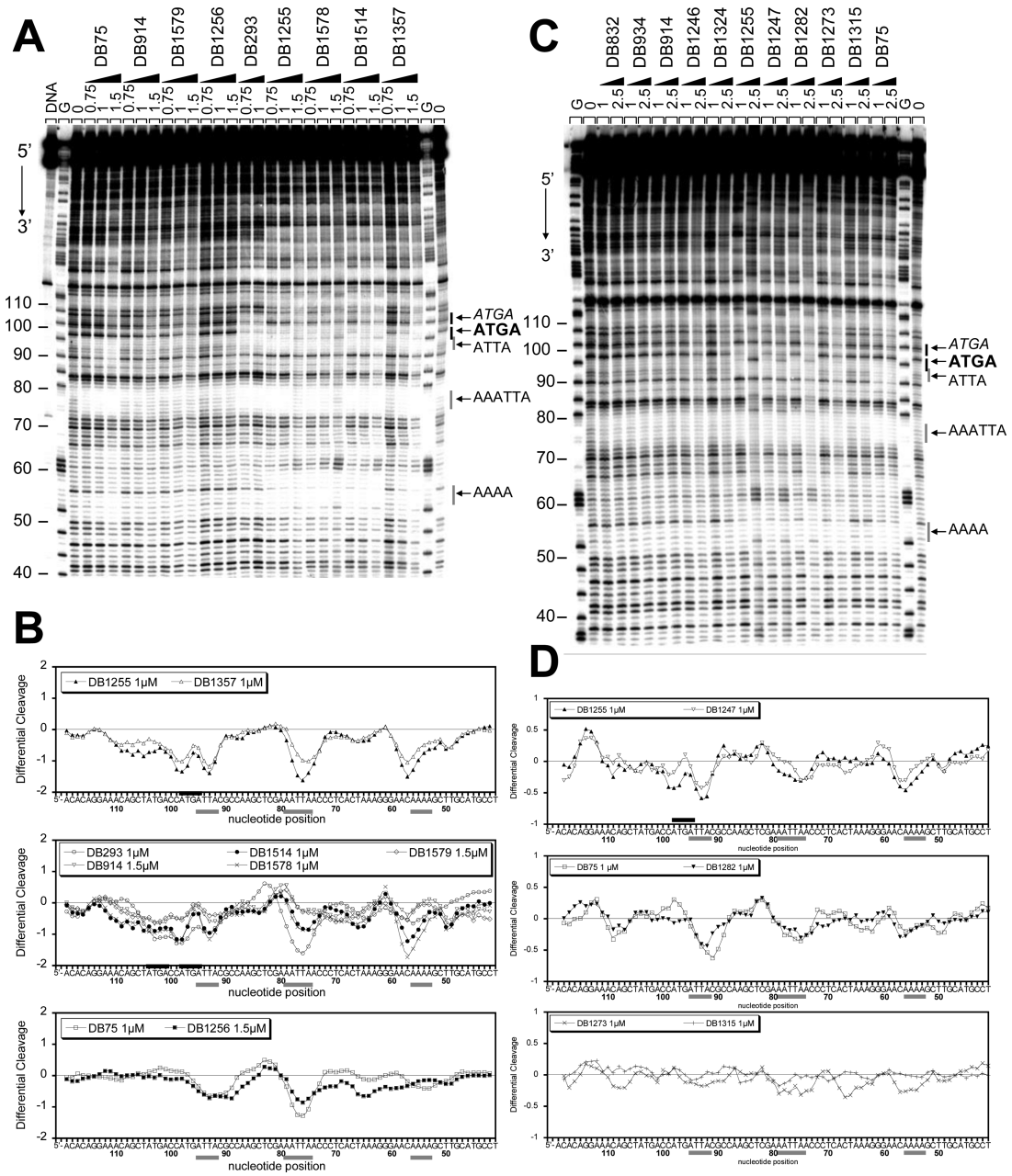


Figure 3. DNase I footprinting titration experiments with derivatives of DB1255. (A and C) Denaturing polyacrylamide gel and (B and D) the corresponding densitometric analysis. G, gray, and black boxes are as defined in the legend of Figure 2.

Supporting Information, and densitometer traces are shown in Figures 2B and 3B–D and Figure S1B of the Supporting Information. The DNA sequence used contains three AT sites that give strong footprints with AT specific minor groove binders: 5′-AAAA-3′ between positions 50 and 60, 5′-AAATTA-3′ between positions 70 and 80, and 5′-ATTA-3′ between positions 90 and 100. The sequence also has two 5′-ATGA-3′ sites within different contexts (5′-ATGAC-3′ and 5′-ATGAT-3′) on both sides of position 100, and both bind DB293 as a dimer. DB1255 behaves in a different manner with strong footprints at only one 5′-ATGA-3′ site (5′-ATGAT-3′ and not 5′-ATGAC-3′) and the 5′-ATTA-3′ site between positions 90 and 100 (Figure 2). It also has footprints at the more remote 5′-AAAA-3′ and 5′-AAATTA-3′ sites described above. The results for the site with a single GC base pair in an AT sequence agree with the optimal footprinting site observed around the ERG protein binding site for DB1255, 5′-AAGTT-3′.[11] Clearly, DB1255 has an overall recognition capability different from that of classical AT specific minor groove binders or the DB293 dimer at ATGA.

Related derivatives were also evaluated for sequence selectivity on the same DNA sequence (Figure 3) to establish structure–binding selectivity relationships. Among the dithiophenes, the addition of a methyl group to the phenyls (DB1357) or the thiophenes (DB1514) and addition of a fluoro substituent to the phenyls (DB1578) maintain the binding on the 5′-ATGA-3′ site that is recognized by DB1255 (Figure 2). Similar binding to ATGA also occurs, but to a lesser extent, when phenyls are replaced with pyridines (DB1247). However, the ATGA binding specificity is lost when the two thiophenes are replaced with two furans (DB914), and related compounds DB832, DB934, DB1246, DB1324, DB1579,

dx.doi.org/10.1021/bi401582t | Biochemistry 2014, 53, 1218–1227

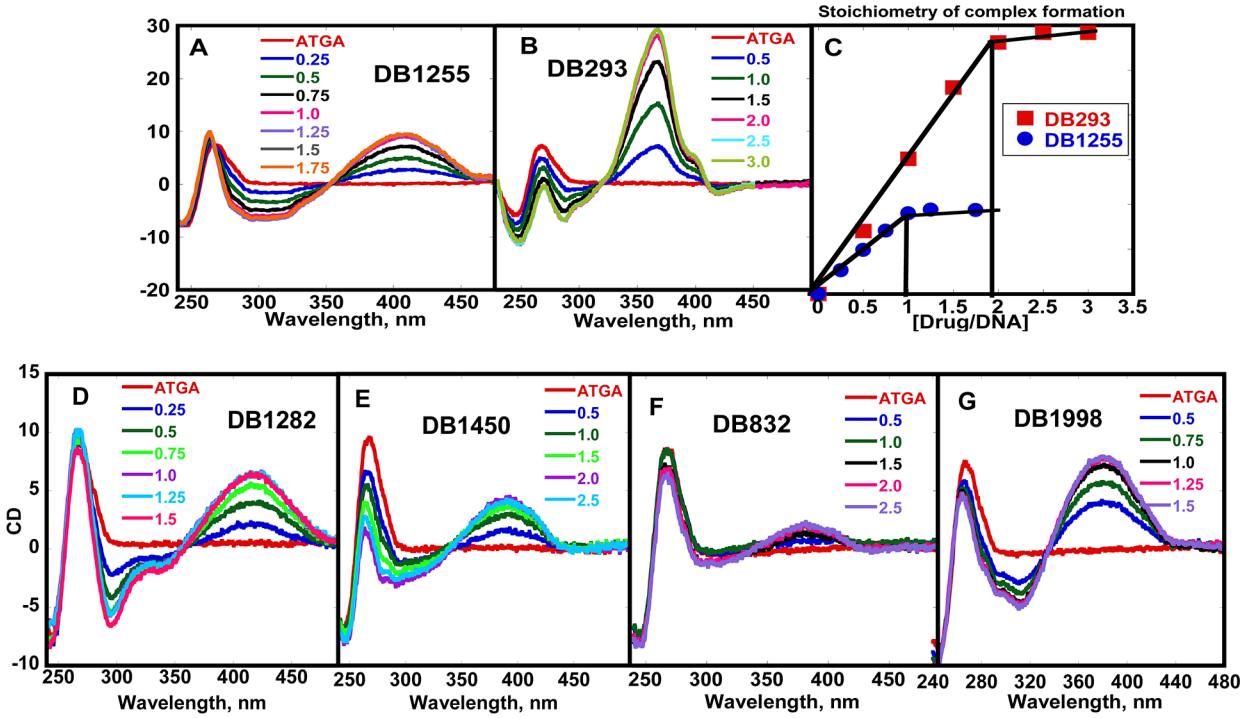


Figure 4. CD spectra of compounds in ATGA: (A) DB1255, (B) DB293, (D) DB1282, (E) DB1450, (F) DB832, and (G) DB1998. The ratios of the compound to DNA are given. (C) Plot of the induced CD signal vs the compound to DNA concentration ratio for DB293 and DB1255.

DB1315, and DB1256, or by two selenophanes (DB1282 and DB1273 with phenyls or pyridines, respectively). Replacing dithiophenes in DB1255 with dithiazoles in DB1998 results in a loss of binding to the 5′-ATGAT-3′ footprint site but retention of the recognition of its closest AT-rich site at positions 92−95 as well as to the AT-rich sequences around positions 75 and 55 (Figure S1 of the Supporting Information). DB1998 is known to have nonspecific binding properties,[11] and because the footprinting ATGA sequence overlaps significantly with AT sites, it does not discriminate between these sites. Therefore, for other biophysical studies such as T_m, SPR, and CD, we have used hairpin sequences that do not have overlapping sites to verify the binding strength of DB1998 and other compounds. Interestingly, the addition of a second furan ring (DB1256) to DB293 results in a longer molecule that totally abolishes the binding to both ATGA sites and strongly reduces the level of binding to AT-rich sites. Binding could be seen only using 1.5 μM DB1256 (Figure 3A,B) and not lower concentrations (Figure S1 of the Supporting Information). These results illustrate that the increased curvature of DB1256 is not optimal for minor groove binding.

Relative Binding Affinity and Specificity Determined by Thermal Melting (T_m). To obtain initial affinity values for the DNA complexes of different compounds, thermal melting experiments were performed with hairpin oligomer duplexes containing a pure AT sequence (5′-AATT-3′) and a mixed sequence (5′-ATGA-3′ in the preferred 5′-ATGAT-3′ context). The specific binding sites in each oligomer duplex were based on DNase I footprinting studies (Figures 2 and 3). The compounds were added to the target DNAs at 1:1 and 2:1 compound:DNA ratios, and the maximal T_m value is near 1:1 for the strong binding compounds. These results indicate that a maximum of 1 mol of compound is bound per mole of DNA sites at saturation. Figure S2 of the Supporting Information shows representative T_m plots of selected compounds in the presence of the 5′-ATGA-3′ sequence, and T_m values are listed in Table S1 of the Supporting Information. Dithiophene DB1255 binds strongly to the 5′-ATGA-3′ hairpin duplex (ΔT_m = 12.2 °C) compared to furan DB914 (2.1 °C) and selenophane DB1282 (5.3 °C) derivatives, in agreement with DNase I footprinting experiments. Dithiazole DB1998 binds to ATGA strongly with a ΔT_m of 14.1 °C but also binds equally strongly to AATT with a ΔT_m of 15.1 °C and thus has poor sequence selectivity, while dithiophene DB1255 and closely related compounds bind more weakly to 5′-AATT-3′ (AATT ΔT_m = 7.5 °C). The fused dithiophene, DB2297, shows a low ΔT_m of 1.0 °C with 5′-ATGA-3′, and this indicates its shape does not match that of the minor groove. From the T_m studies, it is clear that DB1255 has better DNA sequence selectivity and stabilizes the ATGA sequence better than other related compounds (Figure 1).

Binding Modes Determined by Circular Dichroism. Although these compounds are expected to target the DNA minor groove, CD studies are important to establish the binding mode, especially for the ATGA sequence. Strong positive induced CD signals upon titration of the compound to DNA sequences are characteristic of binding in the minor groove of DNA.[27] DB1255 showed a strong positive induced CD (Figure 4A) upon titration into the ATGA sequence. The CD results as a function of ratio are compared for ATGA in Figure 4C, and the plots clearly illustrate that DB293 binds at a 2:1 ratio as opposed to DB1255, which exhibits binding site saturation and CD signal magnitudes characteristic of a 1:1 complex with ATGA. The monomer binding mode is surprising for a sequence with GC base pairs and is unlike that of DB293 and most polyamides of a similar size.[5−7,28] DB1255 also showed positive induced CD in the presence of AATT, but the signal is significantly weaker than in the presence of ATGA (not shown). DB1998 also induces strong CD signals when it is bound to ATGA (Figure 4G), again in agreement with T_m

dx.doi.org/10.1021/bi401582t | Biochemistry 2014, 53, 1218−1227

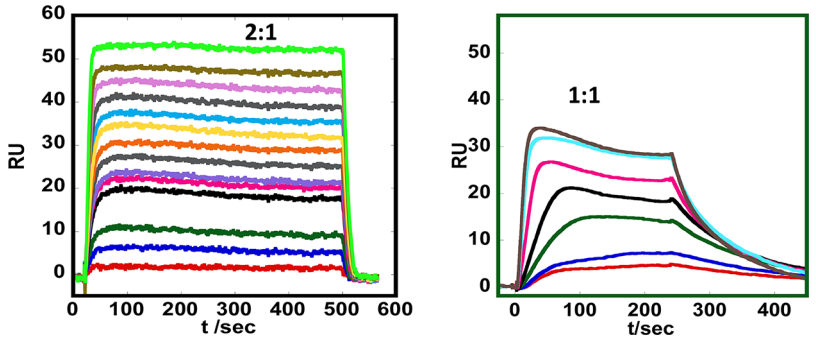


Figure 5. Surface plasmon resonance study. Representative sensorgrams of DB293 and DB1255 in the presence of ATGA. The compound concentration for DB293 ranges from 0.01 to 1 μM and for DB1255 from 0.001 to 0.3 μM (from bottom to top, respectively).

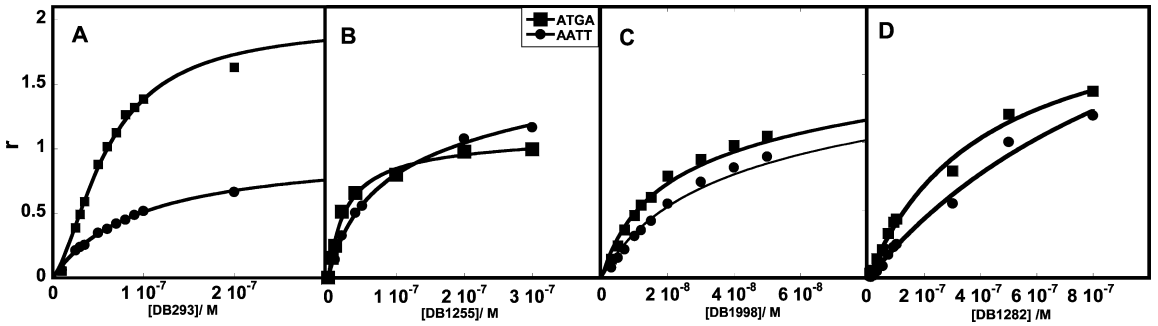


Figure 6. Binding affinity study using SPR. Binding plots for ATGA and AATT for (A) DB293, (B) DB1255, (C) DB1998, and (D) DB1282. The data were fit to a steady-state binding function using appropriate 1:1 or 2:1 models to determine equilibrium binding constants.

values. The tricyclic compounds DB1450, DB832, and DB1341 bind very weakly to DNA (Figure 4E,F) and did not induce a strong signal upon forming a complex with the ATGA sequence, in agreement with footprint and T_m results.

Binding Affinity, Specificity, and Stoichiometry Determined by Biosensor Surface Plasmon Resonance (SPR). To quantitatively evaluate the DNA affinity and stoichiometry of DB1255, biosensor SPR experiments were conducted with immobilized DNA hairpin duplexes containing ATGA and AATT binding sequences (Figure 1). Because the SPR approach responds to mass, it is an excellent method for comparative studies of dications that have very large differences in properties and equilibrium binding constants, *K*. In the initial studies, it was observed that DB1255 sticks to the injection needle and tubing of the instrument, and in the beginning injection time range, the flow solution is depleted of DB1255. Therefore, the experiments for DB1255 were performed in the presence of DMSO in samples and running buffer. We have tested such conditions previously and found that the presence of ≤10% DMSO helped to reduce the loss of compound to tubes and surfaces in SPR without significantly affecting the binding affinity or the DNA properties on the surface.[11] At a DMSO concentration of 10%, we were able to successfully obtain SPR sensorgrams of binding of DB1255 to DNA.

SPR sensorgrams (Figure 5) were subjected to steady-state analyses and then fit to appropriate binding models to determine *K* values for all compounds (Materials and Methods). On the basis of T_m, CD, and footprinting studies, only four compounds, dithiophene DB1255, diselenophane DB1282, dithiazole DB1998, and furan DB293, were selected for SPR binding studies. DB293, which was shown earlier to bind as a cooperative dimer to ATGA, was used to provide a reference for the stoichiometry of these complexes.[1,5] In Figure 5, the SPR signal for DB1255−ATGA complex formation is half that produced by DB293 and shows that DB1255 binds as a monomer compared to the stacked dimer formed by DB293 in the ATGA minor groove. In agreement with the results produced by other methods, the binding to AATT and ATGA in SPR was very dependent on compound structure. SPR results in Figure 6B show that DB1255 binds strongly to ATGA (K_D = 26 nM) compared to AATT (K_D = 86 nM). The Se derivative, DB1282 (Figure 6D and Table 1), binds 15 times weaker to ATGA than DB1255 with significant nonspecific interactions. The differences in binding affinity for these two compounds that differ only at the central heterocycles are striking. DB1255 is approaching saturation of the ATGA binding sites at 0.1 μM, while DB1282 is just beginning to bind at that point.

Table 1. Equilibrium Binding Constants for DB1255 and Its Analogues in the Presence of ATGA and AATT[a]

	K_D (nM)	
compound	ATGA	AATT
DB293[b]	114; 32	45
DB1255	26	86
DB1282	416	1219
DB1998	24	33

[a]The experiments were conducted in cacodylic acid buffer (pH 6.25) at 25 °C. [b]K_1 and K_2 are the first and second equilibrium binding constants, respectively, obtained from a 2:1 cooperative binding model for DB293 with ATGA. Other compounds bind to both DNAs as 1:1 complexes. Dissociation constants were calculated from SPR experiments.

Dithiazole analogue DB1998 binds as strongly to ATGA [K_D = 30 nM (Figure 6C)] as dithiophene DB1255 but shows little

dx.doi.org/10.1021/bi401582t | *Biochemistry* 2014, 53, 1218−1227

specificity as it also binds strongly to AATT [K_D = 24 nM (Table 1)]. The thiazole SPR result is different from the footprinting sequence in which no footprint is seen at ATGA (Figure 3 and Figure S2 of the Supporting Information). The difference is likely due to the isolated ATGA site in the DNA oligomer used in SPR, while there is an AT site almost overlapping the ATGA site in the footprint sequence. The strong binding of DB1998 at AT sequences, also observed in SPR, probably explains the inhibition of its binding to ATGA in the footprint sequence. The fused ring compound, DB2297, binds very weakly to ATGA, and the binding constants are not in the instrument's range for binding analysis. Interestingly, sensorgrams for DB293 (Figure 5) and other compounds show off rates much faster than those of DB1255. Clearly, the features of the DB1255 molecule allow it to bind as a monomer to ATGA with a higher affinity compared to other diamidines, and with much slower binding kinetics. These binding trends are supported by the T_m results listed in Table S1 of the Supporting Information and by DNase I footprinting experiments (Figures 2 and 3 and Figure S1 of the Supporting Information) for these and other compounds.

DISCUSSION

Our previous studies with dithiophene derivative DB1255 showed that the compound could bind to an AT sequence with one GC base pair[11] but provided limited quantitative information about the complex of DB1255 or its analogues with mixed base pair sites. Given the excellent transcription factor inhibition and biological results with DB1255 in cells, we have undertaken a more detailed biophysical analysis of DB1255 and some analogue complexes with DNA. The intrinsically narrow AT DNA minor groove that is recognized by monomer heterocyclic compounds is a well-established phenomenon.[29,30] Mixed guanine-containing DNA sequences typically have wider grooves and are, therefore, difficult to recognize using a monomer complex.[8] On the basis of work with polyamides and heterocyclic diamidines, such as DB293,[5,31] dimer-forming compounds have been used to enhance specificity and affinity for wider DNA minor grooves with GC base pairs. Even after these initial successes, however, we understand relatively little about the general compound structural features necessary for mixed DNA recognition by most heterocyclic cations. Because DB1255 has been found to bind strongly to a mixed AT and GC sequence in both cellular and *in vitro* contexts,[11] the compounds of Figure 1 were evaluated in an effort to understand the compound structural features necessary to recognize mixed sequences of AT and GC base pairs in DNA. The compounds (Figure 1) have variable lengths, symmetries, and ring heteroatoms for testing of their effects on the mixed sequence of DNA.

In our initial investigation of DNA binding by DB1255, it was a surprise to find that it has a high T_m value, very strong CD signals with ATGA, and 4-fold better binding affinity for ATGA versus AATT. Compounds that specifically recognize GC base pairs usually have strong H-bond acceptors such as the furan in DB293. The second major surprise with DB1255 was that it clearly binds strongly as a monomer, not as a dimer, to ATGA. SPR results show better specificity for that sequence with DB1255 than for the DB293 dimer (Table 1). The results in this paper, thus, clearly show that DB1255 has a novel GC recognition mode and is able to induce narrowing in the minor groove width of GC-containing sequences for optimal interactions.

To evaluate ideas for how DB1255 is able to effectively bind to the ATGA sequence, a molecular docking study was conducted with the duplex d[(5′-CCATGATCT-3′)(5′-AGATCATGG-3′)].[26] The molecule was docked into the 5′-cATGAt-3′ sequence, found in footprinting and used in the other experiments, and energy minimized. The low-energy form (Figure 7) has bifurcated H-bond-type interactions from

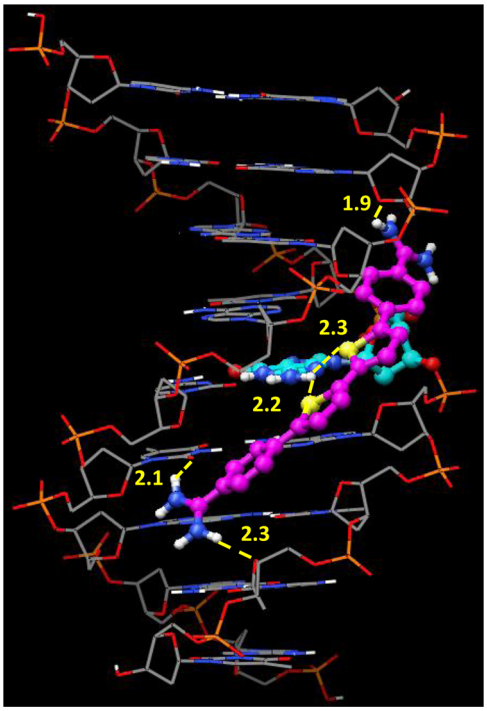


Figure 7. Docking of DB1255 to d[(5′-CCATGATCT-3′)(5′-AGATCATGG-3′)] by using Autodock version 4.02. Two thiophene S atoms (yellow) form bifurcated hydrogen bonds with the G-NH in the minor groove (shown with a yellow dashed line). The lower amidine group forms H-bonds with the thymine (T) carbonyl of T4, which is at the 5′ side of G on the opposite strand), and to a deoxyribose oxygen of the DNA backbone (yellow dashed lines) on the same strand as G. The top amidine group is adjacent to the flanking 3′-T in cATGAt and forms an H-bond with the DNA backbone (yellow dashed line). The DNA backbone is shown as a stick model, colored by atom type (gray for carbon, red for oxygen, blue for nitrogen, and orange for phosphorus). The G5 base is displayed as a ball and stick and colored by atom type (cyan for carbon, white for hydrogen, and blue for nitrogen), and the ligand is displayed as a ball and stick also and colored by atom type (magenta for carbon, white for hydrogen, and blue for nitrogen).

the G-NH_2 group, which projects into the minor groove, to the two thiophene S atoms that are symmetrically located on both sides of the G-NH_2 group. The G-NH_2 group−S distances are 2.3 and 2.2 Å (Figure 7). One amidine (the bottom one in Figure 7) forms an H-bond with a thymine (T) carbonyl group on the complementary strand. The other amidine (the top one in Figure 7) is adjacent to the flanking 3′-T in cATGAt, so that the actual binding site is 5 bp and should be written as 5′-ATGAT-3′. This latter amidine can alternately form H-bonds with the thymine carbonyl (weaker in this calculation) or to an adjacent backbone deoxyribose O (stronger). Interaction with the 3′-T can explain why DB1255 binds more strongly to 5′-ATGAT-3′ than to 5′-ATGAC-3′, which has the 3′-terminal GC base pair H-bond in the minor groove. DB293 binds well to

dx.doi.org/10.1021/bi401582t | *Biochemistry* 2014, 53, 1218−1227

both sequences, but DB1255 binds to only 5′-ATGAT-3′. The thiophenes have the best shape to give the optimal minor groove binding conformation to DB1255 with the thiophenes and terminal amidines in position to strongly recognize the ATGAT sequence.

A similar modeling study was conducted with DB914, the difuran analogue of DB1255, which does not have significant binding with the ATGA sequence. No satisfactory minor groove-bound conformation could be found for DB914 (Figure S3 of the Supporting Information). The difuran system is significantly more curved than the dithiophene of DB1255, and as a result, the difuran is pushed away from the floor of the minor groove and only the amidines interact with the bases. Frequently, a more linear conformation was observed with one furan O pointed into the groove and one pointed out, but all of the observed DNA complexes have a higher energy than the DB1255 complex. A second relatively low-energy model for DB1255 was also found in the modeling experiments (Figure S4 of the Supporting Information) but with a frequency significantly lower than that of the structure in Figure 7. The higher-energy model has a single thiophene S−G NH group interaction such that DB1255 can slide back and forth first to have one thiophene and then the other interacting with the G NH group. It seems likely that the single-interaction models are in a dynamic minor groove equilibrium with the structure that has two interactions. The low-energy structure probably serves as a bridge between the two single-interaction conformations.

Dithiazole DB1998 binds to ATGA in the isolated sequence in SPR and CD but does not bind to it in the more complex footprinting sequence. The compound also failed to show specificity in SPR as it binds equally strongly to pure AT sequences (Table 1). Such a strong propensity to interact at an AT-rich site may have oriented the binding of DB1998 to the AT-rich site adjacent to the ATGA site rather than to the ATGA site in the footprinting experiments. This is a very important observation because in our recent findings,[11] it was concluded that DB1255 has better biological activity than DB1998. DB1255 recognizes mixed sequence DNA strongly and specifically to inhibit DNA binding by the ERG transcription factor and is an excellent lead compound for anticancer therapeutic development.[11] DB2297, the fused ring analogue of DB1255, binds very weakly to ATGA. The presence of fused thiophenes changes the structure and properties of DB2297 completely, giving it a more linear shape, which in this case is not appropriate for minor groove binding. This compound modification shows that the presence of two thiophenes is not sufficient for strong binding if the shape of the compound is not optimal for the ATGA groove. The shorter dithiophene, DB1450, showed nonspecific binding to ATGA via SPR and is not comparable to DB1255. The loss of a phenyl in DB1450 and the nonoptimal radius of curvature cause poor binding of that compound. In DB1341, the presence of three thiophenes did not help to improve the DNA binding affinity for ATGA. In conclusion, these results show that in DB1255 the two thiophenes are essential in making key contacts with the G NH group, and the two terminal phenylamidine modules play an extremely important role in the fit of the compound into the ATGAT minor groove. When this structure is significantly modified, for example, in the difuran and diselenophane compounds, the affinity for ATGA decreases dramatically.

■ ASSOCIATED CONTENT

Supporting Information

Synthetic scheme of DB2297, supporting figures (Figures S1−S4), and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: wdw@gsu.edu. Telephone: (404) 413-5503. Fax: (404) 413-5505.
*E-mail: dwb@gsu.edu. Telephone: (404) 413-5498. Fax: (404) 413-5505.
*E-mail: marie-helene.david@inserm.fr. Telephone: +33 3 20 16 92 23. Fax: +33 3 20 16 92 29.

Present Addresses

[§]M.M.: Central University of Rajasthan, NH-8, Bandar Sindri, Ajmer, Rajasthan, India.
[∥]P.P.: Metastases Research Laboratory, GIGA-CANCER, University of Liege, 4000 Liege Sart Tilman, Belgium.
[⊥]M.A.I.: Department of Chemistry, College of Science, King Faisal University, Hofuf 31982, Saudi Arabia.
[@]A.A.F.: Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt.

Funding

The work at Georgia State University was supported by National Institute of Allergy and Infectious Diseases Grant AI064200 (W.D.W. and D.W.B.). M.-H.D.-C. thanks the Fonds Européen de Développement Régional (FEDER, European Community) and the Région Nord-Pas de Calais, the Ligue Nationale Contre le Cancer (Comité du Nord, Septentrion), the Association Laurette Fugain, and the Association pour la Recherche sur le Cancer for grants. M.-H.D.-C. is also grateful to the Institut pour la Recherche sur le Cancer de Lille (IRCL) for financial support.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

SPR, surface plasmon resonance; RU, response unit of Biacore SPR instruments; CD, circular dichroism; T_m, thermal melting.

■ REFERENCES

(1) Ecker, J. R., Bickmore, W. A., Barroso, I., Pritchard, J. K., Gilad, Y., and Segal, E. (2012) Genomics: ENCODE explained. *Nature 489*, 52−55.

(2) Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem. 5*, 182−186.

(3) Lam, E. Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun. 4*, 1796−1800.

(4) Vidal, A., Munoz, C., Guillen, M. J., Moreto, J., Puertas, S., Martinez-Iniesta, M., Figueras, A., Padulles, L., Garcia-Rodriguez, F. J., Berdiel-Acer, M., Pujana, M. A., Salazar, R., Gil-Martin, M., Marti, L., Ponce, J., Mollevi, D. G., Capella, G., Condom, E., Vinals, F., Huertas, D., Cuevas, C., Esteller, M., Aviles, P., and Villanueva, A. (2012) Lurbinectedin (PM01183), a new DNA minor groove binder, inhibits growth of orthotopic primary graft of cisplatin-resistant epithelial ovarian cancer. *Clin. Cancer Res. 18*, 5399−5411.

(5) Wang, L., Bailly, C., Kumar, A., Ding, D., Bajic, M., Boykin, D. W., and Wilson, W. D. (2000) Specific molecular recognition of mixed

nucleic acid sequences: An aromatic dication that binds in the DNA minor groove as a dimer. *Proc. Natl. Acad. Sci. U.S.A. 97*, 12−16.
(6) Wang, L., Carrasco, C., Kumar, A., Stephens, C. E., Bailly, C., Boykin, D. W., and Wilson, W. D. (2001) Evaluation of the influence of compound structure on stacked-dimer formation in the DNA minor groove. *Biochemistry 40*, 2511−2521.
(7) Wang, L., Kumar, A., Boykin, D. W., Bailly, C., and Wilson, W. D. (2002) Comparative thermodynamics for monomer and dimer sequence-dependent binding of a heterocyclic dication in the DNA minor groove. *J. Mol. Biol. 317*, 361−374.
(8) Hancock, S. P., Ghane, T., Cascio, D., Rohs, R., Di Felice, R., and Johnson, R. C. (2013) Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res. 41*, 6750−6760.
(9) Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature 461*, 1248−1253.
(10) Bishop, E. P., Rohs, R., Parker, S. C., West, S. M., Liu, P., Mann, R. S., Honig, B., and Tullius, T. D. (2011) A map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. *ACS Chem. Biol. 6*, 1314−1320.
(11) Nhili, R., Peixoto, P., Depauw, S., Flajollet, S., Dezitter, X., Munde, M. M., Ismail, M. A., Kumar, A., Farahat, A. A., Stephens, C. E., Duterque-Coquillaud, M., Wilson, W. D., Boykin, D. W., and David-Cordonnier, M. H. (2013) Targeting the DNA-binding activity of the human ERG transcription factor using new heterocyclic dithiophene diamidines. *Nucleic Acids Res. 41*, 125−138.
(12) Baldus, C. D., Burmeister, T., Martus, P., Schwartz, S., Gokbuget, N., Bloomfield, C. D., Hoelzer, D., Thiel, E., and Hofmann, W. K. (2006) High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J. Clin. Oncol. 24*, 4714−4720.
(13) Demichelis, F., Fall, K., Perner, S., Andren, O., Schmidt, F., Setlur, S. R., Hoshida, Y., Mosquera, J. M., Pawitan, Y., Lee, C., Adami, H. O., Mucci, L. A., Kantoff, P. W., Andersson, S. O., Chinnaiyan, A. M., Johansson, J. E., and Rubin, M. A. (2007) TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene 26*, 4596−4599.
(14) Metzeler, K. H., Dufour, A., Benthaus, T., Hummel, M., Sauerland, M. C., Heinecke, A., Berdel, W. E., Buchner, T., Wormann, B., Mansmann, U., Braess, J., Spiekermann, K., Hiddemann, W., Buske, C., and Bohlander, S. K. (2009) ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: A comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J. Clin. Oncol. 27*, 5031−5038.
(15) Salek-Ardakani, S., Smooha, G., de Boer, J., Sebire, N. J., Morrow, M., Rainis, L., Lee, S., Williams, O., Izraeli, S., and Brady, H. J. (2009) ERG is a megakaryocytic oncogene. *Cancer Res. 69*, 4665−4673.
(16) Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X. W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science 310*, 644−648.
(17) Ismail, M. A., Bialy, S. A., Brun, R., Wenzler, T., Nanjunda, R., Wilson, W. D., and Boykin, D. W. (2011) Dicationic phenyl-2,2′-bichalcophenes and analogues as antiprotozoal agents. *Bioorg. Med. Chem. 19*, 978−984.
(18) Ismail, M. A., Boykin, D. W., and Stephens, C. E. (2006) An efficient synthesis of 5,5′-diaryl-2,2′-bichalcophenes. *Tetrahedron Lett. 47*, 795−797.
(19) Nanjunda, R., Munde, M., Liu, Y., and Wilson, W. D. (2011) *Real-time monitoring of nucleic acid interactions with biosensor plasmon resonance*, p 392, Taylor and Francis, Boca Raton, FL.
(20) Nguyen, B., Tanious, F. A., and Wilson, W. D. (2007) Biosensor-surface plasmon resonance: Quantitative analysis of small molecule-nucleic acid interactions. *Methods 42*, 150−161.
(21) Fasman, G. D. (1975) *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, 3rd ed., Vol. *1*, CRC Press, Cleveland, OH.
(22) Davis, T. M., and Wilson, W. D. (2000) Determination of the refractive index increments of small molecules for correction of surface plasmon resonance data. *Anal. Biochem. 284*, 348−353.
(23) *Spartan'10 Tutorial and User's Guide* (2010) Wavefunction, Inc., Irvine, CA.
(24) Sanner, M. F. (1999) Python: A programming language for software integration and development. *J. Mol. Graphics Modell. 17*, 57−61.
(25) Liu, Y., Chai, Y., Kumar, A., Tidwell, R. R., Boykin, D. W., and Wilson, W. D. (2012) Designed compounds for recognition of 10 base pairs of DNA with two at binding sites. *J. Am. Chem. Soc. 134*, 5290−5299.
(26) Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem. 19*, 1639−1662.
(27) Rodger, A., and Nordén, B. (1997) *Circular Dichroism and Linear Dichroism*, Oxford University Press, New York.
(28) Wang, S., Kumar, A., Aston, K., Nguyen, B., Bashkin, J. K., Boykin, D. W., and Wilson, W. D. (2013) Different thermodynamic signatures for DNA minor groove binding with changes in salt concentration and temperature. *Chem. Commun. 49*, 8543−8545.
(29) Nguyen, B., Neidle, S., and Wilson, W. D. (2009) A role for water molecules in DNA-ligand minor groove recognition. *Acc. Chem. Res. 42*, 11−21.
(30) Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Nat. Prod. Rep. 18*, 291−309.
(31) Munde, M., Ismail, M. A., Arafa, R., Peixoto, P., Collar, C. J., Liu, Y., Hu, L., David-Cordonnier, M. H., Lansiaux, A., Bailly, C., Boykin, D. W., and Wilson, W. D. (2007) Design of DNA minor groove binding diamidines that recognize GC base pair sequences: A dimeric-hinge interaction motif. *J. Am. Chem. Soc. 129*, 13732−13743.