



Published in final edited form as:

Phys Chem Chem Phys. 2019 January 23; 21(4): 1841–1851. doi:10.1039/c8cp05537h.

Small molecule binders recognize DNA microstructural variations via an induced fit mechanism

E. Kathleen Carter^a, Sarah Laughlin-Toth^{a,b}, Thomas Dodd^a, W. David Wilson^a, and Ivaylo Ivanov^a

^aDepartment of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA. liivanov@gsu.edu; wdw@gsu.edu

^bDepartment of Pediatrics, Aflac Cancer and Blood Disorders Center, Emory University School of Medicine, Atlanta, GA, USA

Abstract

Regulation of gene-expression by specific targeting of protein-nucleic acid interactions has been a long-standing goal in medicinal chemistry. Transcription factors are considered “undruggable” because they lack binding sites well suited for binding small-molecules. In order to overcome this obstacle, we are interested in designing small molecules that bind to the corresponding promoter sequences and either prevent or modulate transcription factor association via an allosteric mechanism. To achieve this, we must design small molecules that are both sequence-specific and able to target G/C base pair sites. A thorough understanding of the relationship between binding affinity and the structural aspects of the small molecule-DNA complex would greatly aid in rational design of such compounds. Here we present a comprehensive analysis of sequence-specific DNA association of a synthetic minor groove binder using long timescale molecular dynamics. We show how binding selectivity arises from a combination of structural factors. Our results provide a framework for the rational design and optimization of synthetic small molecules in order to improve site-specific targeting of DNA for therapeutic uses in the design of selective DNA binders targeting transcription regulation.

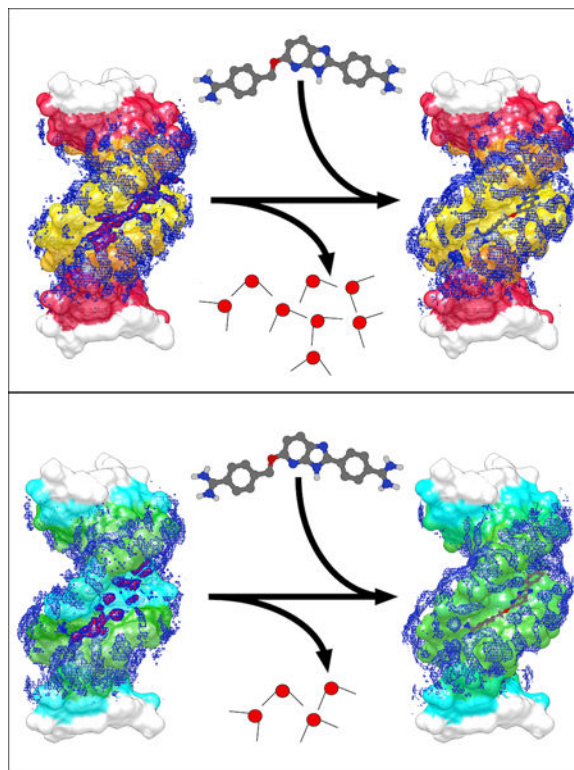
Table of Contents Entry.

Site-specific recognition through contacts, water displacement, and dynamics of a linked azabenzimidazole-diamidine within the DNA minor groove.

Conflicts of interest

There are no conflicts to declare

[†]Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here].
See DOI: [10.1039/x0xx00000x](https://doi.org/10.1039/x0xx00000x)



Introduction

Transcription regulation governs all important aspects of cell biology – cell identity, growth, differentiation and development. Consequently, loss of transcriptional control is a hallmark of many autoimmune disorders, various cancers, neurological, metabolic and cardiovascular diseases.^{1–6} Targeting gene expression with designed small molecules has been a long-standing goal in medicinal chemistry. One major obstacle is that transcription factors generally lack pockets suitable for small-molecule binding and, for this reason, are considered “undruggable”.^{7–12} Transcription factors (TFs) possess DNA-binding domains that recognize DNA in a sequence-specific manner with a footprint of 6 to 12 base pairs. While this recognition length may be too short to ensure specificity, other mechanisms intervene, including interactions with other sequence motifs, cooperative binding and protein-protein interactions outside the DNA binding domains.^{13–16} Instead of targeting transcription factors directly, a more viable approach may be to design small molecules that bind the corresponding promoter sequences and either prevent or modulate transcription factor association.^{17–19} These molecules have to selectively bind DNA at specific sites while also meeting a number of other criteria: water-solubility, cellular and nuclear-uptake, and lacking off-site activities. Diamidines are a class of minor groove DNA binders with just such favorable properties.^{20–22} For instance, DAPI is a commonly used cell stain that binds strongly to A/T base pair rich regions of DNA.^{23, 24} Another small molecule, pentamidine, is an antimicrobial medication used to treat certain types of parasitic diseases.^{25, 26} Nonetheless, 4',6-diamidino-2-phenylindole (DAPI) and pentamidine have respectively been shown to be mutagenic and causing off-target side effects due to their promiscuity and high

affinity for A/T rich regions within DNA (25,26). For this reason, there is a need to design small molecules that are both sequence-specific and able to target G/C base pair sites. Like DAPI and pentamidine, DB2277 is a synthetic, heterocyclic diamidine that binds relatively weakly to A/T-rich base pair sites (Figure 1).^{22, 27} DB2277, however, stands apart in its ability to selectively and strongly bind *in vitro* mixed-site DNA sequences (*i.e.*, G/C and A/T base pair containing).^{22, 27–29} Due to this ability to discriminate against various motifs, DB2277 is an excellent candidate to systematically examine the mechanisms and basic principles of sequence-specific DNA association.^{22, 27–30} In this respect, it is crucial to understand the relationship between the binding affinity and the structural aspects of the small molecule-DNA complex. Shape complementarity, induced fit, water-mediated dynamics, microstructural variation among DNA sequences and DNA deformation effects play an outsized role in determining successful binding. Our previous work showed that the microstructure of unbound DNA sequences correlates strongly with the binding affinity and selectivity for a number of small synthetic molecules.³⁰

In this contribution, we present a detailed molecular modeling study that examines the mechanisms for sequence-specific DNA recognition by DB2277. We used molecular dynamics (MD) to extensively sample the conformational space for six DB2277:DNA complexes with central G/C base pairs and vary A/T flanking sequences and compared the results to the preferred binding sequence, AAAAGTTTT, previously determined experimentally.³⁰ Additionally, we analyzed the changes in DNA hydration upon binding and explained hydration contributions to the overall binding affinities. From the MD trajectories, we also computed persistent DB2277:DNA contacts, DNA minor groove deformation and changes in the ionic distributions upon binding. In this way, we provide a comprehensive picture of the association of a small molecule such as DB2277 with mixed-site DNA sequences. By uncovering the origins of specificity and the structural factors contributing to the binding affinity by using the rapid advances in computational chemistry, we provide a framework for the rational design and optimization of synthetic small molecules to achieve site-specific targeting of DNA for therapeutic uses.

Methods

Molecular Dynamics Simulations

Simulation systems with the following double-stranded DNA sequences were set up: 5'-CGAAAAGTTTTTCG-3', 5'-CGAATTGAATTCG-3', 5'-CGATATGATATCG-3', 5'-CGAAAAGCTTTTCG-3', 5'-CGAATTGCAATTCG-3', 5'-CGATATGCATATCG-3'. The initial models were built in canonical B-form geometry using the Nucleic Acid Builder (NAB) plugin of AMBER16.^{31, 32} To parameterize the DB2277 ligand we first performed geometry optimization in Gaussian09³³ and subsequently electrostatic potential calculations with the B3LYP DFT³⁴ functional and 6–31+G* basis set.³⁵ RESP charges for the ligand were calculated using the Merz-Singh-Kollman (MK) Scheme.^{36, 37} AMBER preparation files and force field parameters were generated with the *GAFF* force field using ANTECHAMBER.^{38, 39} Potential energy scans were performed in Gaussian09 to determine angle and dihedral parameters of DB2277 not defined in *GAFF* force field (Figure S2, Table S1, and Table S2).⁴⁰ The protonation state of DB2277 was chosen based on the results

reported in Harika *et al.*²⁷ and confirmed by calculating the pKa of the azabenzimidazole nitrogens. Details regarding the thermodynamic cycle used to calculate the pKas and the exact values can be found in the electronic supporting information (Figure S3 and Table S3).⁴¹⁻⁴³ DB2277 was manually docked and aligned with the central G•C base pair in the minor groove of the six DNA models. The sequences with only one G•C base pair (5'-CGAAAAGTTTTTCG-3', 5'-CGAATTGAATTCG-3', 5'-CGATATGATATCG-3') are non-palindromic (i.e. 5'-AAAAGTTTT-3' vs. 5'-AAAACTTTT-3') and due to the asymmetry in DB2277, the ligand was oriented in both 5'→3' and 3'→5' directions (Figure S1A-C). The free DNA structures and bound complexes were solvated in 78 Å x 78 Å x 78 Å rectilinear boxes with TIP3P water⁴⁴ and neutralized with Na⁺ ions in TLeap.³² To reach physiological salt concentration of 150 mM, additional Na⁺ and Cl⁻ ions were added to the systems.

All systems were relaxed using steepest-descent minimization for 5,000 steps with positional restraints on the nucleic acid residues. In the canonical ensemble, the systems were heated from 0 K to 310 K over 10 ps with 5 kcal·Å⁻²·mol⁻¹ harmonic restraints imposed on the nucleic acid residues heavy atoms. Five kcal Å⁻²·mol⁻¹ distance restraints were applied to the terminal base pairs to enforce hydrogen bonding to prevent fraying at the ends of the DNA duplexes. The distance between the nitrogen of the DB2277 azabenzimidazole ring and the guanine-NH₂ group was also restrained (k=5 kcal Å⁻²·mol⁻¹) to prevent DB2277 from drifting out of the groove during initial equilibration of the systems. Over eight stages of equilibration (500 ps) in the isothermal-isobaric ensemble (T = 310 K, P = 1 atm), the harmonic restraints were released. The electrostatic interactions were treated using the smooth Particle Mesh Ewald⁴⁵ with a cutoff of 10.0 Å and the SHAKE algorithm⁴⁶ was used to constrain bonds between hydrogen and heavy atoms. After the harmonic restraints were fully released, distance restraints between the azabenzimidazole-N and guanine-NH₂ were removed while gradually decreasing distance restraints on the capping base pairs to 1 kcal·Å⁻²·mol⁻¹ over additional 50 ns of dynamics. During the production runs, the distance restraints on the capping base pairs were maintained at 1 kcal·Å⁻²·mol⁻¹ to prevent base fraying, which could alter the dynamics of the DB2277:DNA complex. For production, simulations were extended to 500 ns and trajectory snapshots were saved every 2 ps. The DB2277:AATTGCAATT simulation was extended an additional 50ns after DB2277 slides away from the targeted GpC site, which was observed in the first 50ns of the MD trajectories (Supplemental Movie 1). All simulations were carried out with resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE)⁴⁷ and were performed using the PMEMDCUDA module⁴⁸ of AMBER16³² with the parm99 force field and the parmbsc0 + ε/ζOL1+χOL4 force field modifications for DNA.⁴⁹⁻⁵¹ Trajectories were postprocessed using CPPTRAJ module of AMBERTOOLS17⁵² to produce 125,000 snapshots for analysis and visualization in VMD.

53

Trajectory analysis with Curves+

Minor groove width and base pair step translational and rotational helical parameters were calculated using the Curves+ and Canal programs⁵⁴ for the unbound and bound complexes. Two-dimensional contoured histograms were produced by defining a spline along the backbone of the DNA using 50 points and measuring the minor groove widths from the two-

dimensional distance matrix produced by Curves+. A smooth line was interpolated along the entire groove for each frame analyzed. Width data generated from Curves+ were then binned into 100 evenly spaced bins and represented as histograms using MATLAB.⁵⁵ To correlate the changes in minor groove width to the experimental steady state dissociation constants for all sequences, we introduced a single metric to represent the histogram data. This metric was derived by subtracting the two-dimensional histograms of the minor groove width for the unbound complexes from their bound states and summing the resulting bin values to produce an integrated overall change in groove width upon binding (Figure S7). Experimental steady-state dissociation constants were converted to effective ΔG in kcal mol⁻¹ using the relation $\Delta G = RT \ln K_D$ to relate the observations from the MD simulations to the energetics of binding.

In addition to conformational analysis done using Curves+, the program Canion⁵⁶ was used to analyze the ion distribution around the DNA duplexes. Using Canion, the location of an ion is described by a distance (D) along the helical axis, a radial distance (R) from the axis and an angle (A) from a reference vector (dashed line) which tracks the helical twist of the nucleic acid (Figure S9). The ion distribution is then plotted in terms of molarity in three dimensions. An angle, $A \approx 90^\circ$, places an ion in the minor groove and an angle $A \approx 270^\circ$ places an ion in the major groove.⁵⁶ An ion is inside the duplex for $R < 10.25 \text{ \AA}$.⁵⁶ The longitudinal distance D tracks the ion displaced along the sequence of the DNA.⁵⁶ A similar metric to the integrated change in minor groove width was used to compare the ion displacement to the effective binding free energy. The molarity of the free structures was subtracted from molarity of the bound structures. The change in molarity was then summed and plotted against the effective free energy (Figure S10D-F).

Grid Inhomogeneous Solvation Theory

The solvent-solute and solute-solute interactions were quantified using grid inhomogeneous solvation theory (GIST).⁵⁷ First, the system is discretized into equal volume voxels (Figure S8A). The energetics of the system are determined by summing over all the voxels on the grid after the individual interactions between water molecules in the voxel and the solute and solvent are computed from the stored frames of the MD trajectory. In addition to per voxel energetics, GIST also calculates the number density of oxygen centers for each voxel. The number densities are referenced to bulk by dividing by bulk density (0.0334 molecules \AA^{-3}) so that an isovalue greater than or equal to one represents waters at or greater than bulk density.⁵⁷ Using the density data, the GIST analysis for the free and bound DNA constructs was restricted to water molecules close to the duplex (Figure S8B). The solvent energies were decomposed into enthalpic and entropic terms

$$\Delta G = \Delta E_{sw} + \Delta E_{ww} - T \Delta S_{trans} - T \Delta S_{orient}$$

where the subscripts s and w on the enthalpic terms denote solute and water, respectively (Figure S8C and S6D).⁵⁷ The entropic term is broken down into $T \Delta S_{orient}$, orientational entropy of the solvent, and $T \Delta S_{trans}$, translational entropy of the solvent.⁵⁷ The enthalpy is

calculated from the force field non-bonded energies. The ΔS_{trans} and ΔS_{orient} terms are given as

$$\Delta S_{trans} = -k_B \rho^o \int g(r) \ln g(r) dr$$

$$\Delta S_{orient} = \rho^o \int g(r) S^o(r) dr$$

with

$$S^o(r) \equiv \frac{-k_B}{8\pi^2} \int g(\omega|r) \ln g(\omega|r) d\omega$$

where \mathbf{r} is the location of the water oxygen and ω is the orientation in the solute frame of reference.⁵⁷ The number density of bulk solvent is given by ρ^o , and $g(r) \equiv \rho(r)/\rho^o$ and $g(\omega|r) \equiv \rho(\omega|r)/\rho_\omega^o$; where $\rho(r)$ is the number density function of the system (53). After determining energetic data on a per voxel basis, the thermodynamic components of solvation can be integrated using the post-processing software suite, gistpp⁵⁸, in specific regions of interest or across the entire system (Figure S8E-G). The isosurfaces can also be visualized in order to determine localized solvation hotspots. In order to quantify the free energy of expelled water upon DB2277 binding, additional functionality of gistpp was used after the initial GIST analysis.⁵⁸ First, a binary density map was created to estimate the volume of water that would be expelled from the DNA duplex upon binding.⁵⁷ Using the most probable binding pose from the bound simulation (chosen as the centroid of the most occupied cluster out of ten clusters), a volume was created that extends 3 Å from any heavy atom (Figure S8H). Then, ΔE_{Total} and $T \Delta S_{Total}$ maps for the entire system were multiplied by the binary density map to produce density maps restricted to the region where waters would be displaced upon binding (Figure S8I-L).⁵⁹ The resulting ΔE_{Total} and $T \Delta S_{Total}$ maps can be subtracted to determine the free energy of the displaced waters (Figure S8M).

Results and Discussion

DB2277 recognizes preferred DNA sequences via induced fit binding to the DNA minor groove

Here we report molecular dynamics simulations of monomeric complexes of DB227 with single G•C sequences and GpC DNA sequences. The single G•C DNA sequences are non-palindromic about the binding site (i.e., 5'-AAAAGTTTT-3' vs. 5'-AAAAGTTTT-3'), and DB2277 is an asymmetric small molecule. Therefore, DB2277 was simulated with the azabenzimidazole facing in both the 5' → 3' and the 3' → 5' orientations for the single G•C sequences, resulting in six bound complexes (Figure S1A-C). Since the GpC sequences are palindromic around the binding site, a single orientation was simulated for these sequences (Figure S1D-F). Experimentally, DB2277 is known to bind with nanomolar

affinity ($K_D < 50.8 \pm 16.1$ nM) to single G•C sequences as a 1:1 complex.^{22, 27–30} We previously reported substantial variations in DNA microstructure for the same series of DNA constructs with varying flanking sequences around a central G•C base pair.³⁰ Specifically, the minor groove widths were markedly altered by the rearrangement of A•T base pairs flanking the G•C or GpC sites. Local variations in microstructure could influence small molecule binding within the DNA minor groove. To provide a quantitative description of the effect of these microstructural variations, here we analyzed the distributions of the minor groove widths along the DNA backbone for our chosen series of sequence using the Curves+ package. Two-dimensional contour histograms for each of the unbound DNA sequences are presented in Figure 2. The AAAAGTTTT sequence has a narrow minor groove width of 4.5 Å at the central G•C base pair (free DNA structure, Figure 2A). The natural breathing motions of DNA induce fluctuations in the minor groove width ranging from 4.0 Å to 6.5 Å (Figure 2A). In contrast, the ATATGCTAT sequence exhibits a much broader groove (8.0 Å) at the same G•C base pair (free DNA structure, Figure 2C). Corresponding histograms for the bound DB2277:DNA complexes are also presented in Figure 2. When bound to DB2277, minor groove widths for the single G•C base pair complexes are nearly identical (5'→3' and 3'→5' orientations, Figure 2A-C). The minor groove widths of the GpC sequences narrow around the small molecule upon binding, but there is variation among the sequences (bound complex, Figure 2D-F). The reduction of the minor groove width upon binding is indicative of an induced-fit recognition mechanism; the minor groove collapses around the small molecule in order to form specific contacts between the ligand and the target binding site.

Comparing the single G•C sequences to the GpC sequences, it is clear that the single G•C sequences are more uniform in their deformation around the small molecule. There is a marked increase in the probability for each sequence to adopt a constricted, narrow groove width of ~5.0 Å along the entire binding site (5'→3' and 3'→5' orientations, Figure 2A-C). The narrowing is most pronounced for the AATTGAATT sequence which decreases from a broad minor groove width of ~8.0 Å at the central G•C base pair in the free state to a narrow 5.0 Å groove upon binding (Figure 2B). The DNA minor groove is less capable of deforming around a small molecule like DB2277 when a second G•C is incorporated to make the central GpC. The constriction of the minor groove after binding is limited to the region upstream of the GpC and affects fewer base pairs than in the G•C sequences (bound complex, Figure 2D-F). For example, DB2277:AATTGCAATT has a high probability of 5.0 Å minor groove width confined to the 3 base pair steps near the 5' region of the sequence, specifically at 5'-AATT-3'. For the AATTGCAATT sequence, this finding can be attributed to the observed migration (or sliding) of DB2277 away from the targeted GpC region of the sequence towards the 5' end (Supplemental Movie 1). As previously reported, the microstructure of the unbound sequence (5'-AATTGCAATT-3') is intrinsically narrow (~4.5 Å) at the A•T steps, but at the targeted binding site, the minor groove is much wider (~8.0 Å). By sliding toward the 5' end of the sequence, DB2277 locates a structurally preferred narrow groove, similar to the microstructure of AAAAGTTTT, in which to bind. In this location, the DNA groove can collapse around DB2277 while the rest of the duplex undergoes a smaller structural deformation. DB2277 does not slide in the AAAAGCTTTT nor in the ATATGCATAT sequences, but the same isolated constriction in the minor groove upstream of the GpC sites is seen in these sequences (bound complex, Figure 2D and 2F).

The observations of the changes in minor groove width suggest that the modification from single G•C to GpC results in a wider and less adaptable minor groove. This forces the minor groove to undergo a larger deformation about the target site to create a more favorable conformation for DB2277.

To correlate the observed changes in minor groove width to the experimental steady state dissociation constants for all sequences, we introduced a single metric to represent the histogram data from Figure 2. This metric was derived by subtracting the two-dimensional histograms of the minor groove width for the unbound complexes from their bound states and summing the resulting bin values to produce an integrated overall change in groove width upon binding (Figure S7). To relate the groove width metric to the energetics of minor groove insertion, experimental steady-state dissociation constants were converted to effective G in kcal mol⁻¹ using the relation $\Delta G = RT \ln K_D$. In Figure 3, we have plotted integrated groove width change against effective G for all DNA sequences. There is a clear linear relationship between integrated change in minor groove width and effective free energy. Binding free energy decreases with increased groove width alteration, indicating that DNA conformation is critical to small molecule recognition. A binding site that requires less deformation leads to a more energetically favorable induced fit.

The preferred binding sequence, AAAAGTTTT undergoes ~15 Å of change in minor groove width, significantly less than the other sequences (Figure 3). Similarly, the minor groove of AAAAGCTTTT deforms the least (~40 Å) of the GpC sequences. The minimal change in minor groove width is due to the natively narrow minor groove of these sequences, as a result of the propensity for A-tracts to have consistent high angles of propeller twist.³⁰ This allows for the formation of bifurcated hydrogen bonds along the backbone that restrict the breathing of the minor groove. The AAAAGCTTTT sequence does not form these bonds as readily as the AAAAGTTTT sequence due to the incorporation of the second G•C base pair. Therefore, the AAAAGCTTTT sequence has a lower probability of assuming a 4.5 Å width groove width than the AAAAGTTTT sequence (free DNA structure, Figure 2A, 2D). This results in a more substantial integrated change in minor groove width upon binding (red triangles vs. red circle, Figure 3). The consistency in microstructure before and after binding is also evident when comparing the inter-base pair parameters. Sequences AAAAGTTTT and AAAAGCTTTT change less from the free structures after binding compared to the sequences flanked with 5'-AATT-3' or 5'-ATAT-3' (red lines, Figure S5 and S6). The benefit of binding in a preferred intrinsic microstructure is demonstrated by the decreased integrated change in minor groove width of AATTGCAATT compared to that of the correlating single G•C sequence. AATTGCAATT undergoes approximately 100 Å of integrated change in minor groove width compared to approximately 270 Å for the corresponding single G•C sequence (blue circles vs triangles respectively, Figure 3). By shifting away from the broad GpC region, DB2277 induces the constriction from 8.0 Å to 5.0 Å in a much smaller, naturally narrow, part of the sequence and thus decreases the overall deformation (bound complex, Figure 2E and Figure S7E). Although DB2277 has minimized the change in groove width by sliding to a narrower region in the AATTGCAATT sequence, it still deforms the minor groove more than AAAAGCTTTT upon binding (blue circle vs. red circle respectively, Figure 3). Regardless, DB2277 forms a stronger complex to AATTGCAATT

(effective $G = -11.27 \text{ kcal mol}^{-1}$) than to AAAAGCTTTT (effective $G = -9.85 \text{ kcal mol}^{-1}$).
³⁰ This contradicts the notion that GpC sequences cause large deformations upon binding and these deformations are in turn responsible for lower binding affinity.

DB2277 replaces structural water molecules in the minor groove spine of hydration

While all canonical B-form DNA contain organized arrays of water molecules within the minor groove, previous reports indicate that the residence time and the number of structured waters is affected by sequence.^{60–64} The intrinsically rigid, narrow and deep minor groove of A-tract sequences has been shown to contain a higher number of long-lived water molecules.^{65–67} The number of water molecules associated with the groove can also affect the deformability of DNA. Water molecules trapped in narrow, deep minor grooves are enthalpically hindered from exchange with bulk solvent due to steric and electrostatic interactions with the DNA. Water exchange proceeds through a series of metastable intermediates to minimize the free energy barrier for release from the DNA. Displacement of structural waters from the minor groove upon small molecule binding is entropically favored.^{61, 68, 69} Consequently, mixed sequence DNA duplexes are inherently more flexible. The high deformation cost of binding to intrinsically flexible sequences, combined with their broad, shallow grooves makes them less suitable for trapping structured water molecules.
60, 61

The solvent energetics for both the free and bound forms of the six sequences of interest were analyzed using grid inhomogeneous solvation theory (GIST) in order to investigate the influence of sequence-dependent hydration on the binding affinity of DB2277.⁵⁷ With GIST, the analysis of solvent energetics can be restricted to include only specific structural waters important for binding (Figure S8C-G).⁵⁸

The flanking base pair variations in the six sequences had a pronounced effect on the array of waters in the minor groove for the free DNA structures (free DNA structure, Figure 4). The A-tract sequences have the most continuous array of water molecules along the minor groove. Breaking the A-tract DNA at the GpC site disrupts the continuity of the structural water spine (free DNA structure, Figure 4A and 4D). Neither the single G•C nor the GpC sequence flanked with AATT produced a hydration pattern as uniform as the A-tract sequences (free DNA structure, Figure 4B and 4D). In contrast, in both ATATGATAT and ATATGCATAT, the structured water array is completely absent (free DNA structure, Figure 4C and 4F). By incorporating a second G•C base pair to form the GpC sequences, the groove not only becomes wider but also becomes more flexible, as observed by the ~20% increase in RMSF at the core binding site (Figure S6). The alternating sequences, ATATGATAT and ATATGCATAT, in particular, are highly dynamic with RMSF greater than that of the other sequences (Figure S6C and S6F). In contrast, AAAAGTTTT DNA has an intrinsically rigid and narrow minor groove and is, therefore, more capable of trapping structured waters. These results suggest that natively flexible and broad minor grooves are responsible for the observed decrease in structure associated with the mixed sequence DNA.

The binding site of DB2277 overlaps with the structured water array along the minor groove that must be displaced upon binding. In Figure 5, the G of the displaced water is plotted against the effective free energy of binding. The resulting linear trend suggests that

sequences with tightly bound waters in the minor groove at the central G•C or GpC also have the best binding affinity. The free energy of the structured waters is dominated by a negative enthalpic term, suggesting a strong hydrogen bond network within the minor groove, which outweighs the associated unfavorable entropic term (Table S1). Because a free energetic penalty is incurred by displacing structured water, DB2277's ability to outcompete structured water indicates the existence of a microstructural environment that is preferable for binding polar or charged ligands.

DB2277 binding of the preferred sequence, AAAAGCTTTT, incurs the highest energetic penalty of structure water displacement with $G = \sim -9 \text{ kcal mol}^{-1}$. However, this sequence requires the least deformation, leading to a very favorable DB2277 binding free energy. The next preferred sequence, AATTGCAATT, loses significantly less-structured water ($G = \sim -3.5 \text{ kcal mol}^{-1}$). In the free structure, AATTGCAATT has more structured water at the 5' end of the sequence than at the GpC region, indicating that DB2277 prefers the same microstructure that traps structural water (bound complex, Figure 4E). Although the minor groove of AAAAGCTTTT deforms less than that of the AATTGCAATT sequence, it incurs a high energetic penalty ($G = \sim -6.0 \text{ kcal mol}^{-1}$) when displacing the water and a higher entropic penalty from DB2277 association (Figure S6D and S6E). This likely explains why DB2277 does not bind as favorably to AAAAGCTTTT as AATTGCAATT. Unlike the corresponding GpC sequence, DB2277 does not slide in the AATTGAATT duplex. This causes greater structured water energy loss ($G = \sim -6 \text{ kcal mol}^{-1}$) and minor groove deformation. However, the change in RMSF from free to bound for both the 5'→3' and 3'→5' orientations of DB2277 (red, blue respectively, Figure S4B) is less than that for the AAAAGCTTTT sequence, suggesting there would be less entropic penalty due to increased rigidity. Although ATATGATAT and ATATGCATAT displace fewer favorable waters, they still do not provide a suitable minor groove geometry for DB2277 and significantly reduce the flexibility of the target DNA.

To provide a complete view of minor groove desolvation upon binding, the ion displacement was studied using the Canion package.⁵³ Upon binding, there is a decrease in ion molarity in the minor groove ($A \approx 90^\circ$) for all 6 sequences (gray region, Figure S10A). To correlate the observed changes in ion molarity to the variation in sequence and to the binding affinity, the molarity of the unbound structures was subtracted from their corresponding bound complexes and summed to produce an integrated change in molarity. The ion displacement (M) from the minor groove does not correlate as strongly to effective G of binding as G of water displacement (Figure S10D). All of the palindromic sequences displace more ions than the nonpalindromic sequences but the sequence itself has little effect on the change in ion molarity (Figure S10E). Similar consistent change in molarity can be seen in the longitudinal direction (Figure S10F). This indicates that the variation in binding affinity is less dependent on ion displacement than on other factors such as minor groove deformation and change in hydration.

From results of the GIST hydration data and the Canion ion density analysis, it is clear that DB2277 prefers a binding site that is geometrically suited to its structure. Generally, such a site is also amenable to trapping structured water. This trade-off between geometric

suitability and structured water stability largely determines DB2277's binding affinity to a given site, with geometric suitability tending to outweigh desolvation energy.

Direct and water-mediated contacts stabilize DB2277 within the target site

Previous studies have shown that heterocyclic diamidines are not selective toward G•C base pairs.^{70, 71} To improve selectivity, the core azabenzimidazole of DB2277 (Figure 1) was added by design to increase the probability of forming a hydrogen bond to the NH₂ functional group of the central G base.²² This is the primary mode for sequence discrimination by DB2277 that we observe in the MD simulations. In the single G•C complexes, the hydrogen bond between the azabenzimidazole and the NH₂ functional group of guanine is persistent for > 80% of the simulation length. In addition, when bound to AAAAGTTTT in the 5' → 3' direction (5'→3' orientation, Figure S1A), the azabenzimidazole forms a bifurcated hydrogen bond from the -NH group of the imidazole to the guanine and to the thymine base on the opposing DNA strand, one step above the central G•C (Figure 6A and 6C). Support for this hydrogen bond pattern is provided by previous NMR studies done by Harika *et al.* who observed similar binding patterns for DB2277 bound to single G•C base pairs in A-tract DNA.²⁷ In the AATTGAATT and ATATGATAT sequences, the conserved hydrogen bond from the -NH group of the imidazole to guanine-N3 is present, but the bifurcation has been ablated by the removal of the second acceptor (Figure S11I and S11O). When DB2277 is positioned the 3'→5' orientation in the AAAAGTTTT sequence the azabenzimidazole again forms a bifurcated hydrogen bond from the -NH group of the imidazole. However, in this case, it binds to the cytosine of the G•C pair and thymine on the reference strand one step below the central G•C (Figure 6D and 6F).²⁷ The same hydrogen bond from the -NH group of the imidazole to central G•C cytosine is present in the 3' → 5' orientation AATTGAATT and ATATGATAT sequences (Figure S11L and S11R) but again the bifurcation is not present due to the change in sequence.

Non-equivalent binding is observed for the two amidine groups due to the asymmetry of DB2277. The phenylamidine directly attached to the core azabenzimidazole (amidine-1, Figure 1) is rigid because it is not connected by a linker (Figure S14). When DB2277 is oriented in the 5'→3' direction in the AAAAGTTTT sequence, amidine-1 forms a hydrogen bond with the thymine two steps above the central G•C on the complementary strand of the sequence (Figure 6A and 6C). When in the opposite direction (3'→5'), it forms a hydrogen bond with the thymine two steps below the central G•C on the reference strand (Figure 6D and 6F). The same hydrogen bonds are made to the ATATGATAT sequence, which has thymine nucleotides at the same positions (Figure S11O and S11R). In the AATTGAATT sequence, amidine-1 forms hydrogen bonds to adenine nucleotides instead of thymine nucleotides because of the change in sequence at these positions (Figure S11I and S11L).

The phenylamidine attached to the flexible linker -OCH₂- (amidine-2, Figure 1) does not directly contact the DNA in any of the single G•C sequences, regardless of the orientation of DB2277. It is highly dynamic and freely rotates in solution, as shown by Harika and Wilson (Figures S16 and S17).⁷² Instead of forming persistent hydrogen bonds to the DNA duplex, amidine-2 is stabilized by several water-mediated contacts to the bases and the backbone of

the DNA (Figure 6B, C, E, and F). This results in preferred torsional angles as shown in Figure S16A-C and S17A-C.

Due to the designed isohelical shape of DB2277, contacts made between DB2277 in the 5' → 3' orientation are more persistent than in the 3' → 5' orientation. For example, bound in the 5' → 3' orientation in AAAAGTTTT, DB2277 curves around the minor groove of the DNA sequence, aligning amidine-1 with the thymine on the complementary strand, thereby improving the hydrogen bond formation resulting in a persistence of 91.83%. In the opposite orientation, the hydrogen bond acceptor for amidine-1 is on the reference strand, which does not fit the shape of DB2277 and decreases the overall persistence to 71.75%. This is true for the AATTGAATT and ATATGATAT sequences as well, which have decreased amidine-1 hydrogen bond persistence when comparing the 5' → 3' orientation to the 3' → 5' orientation (~4% and ~7% respectively, Table S5).

The hydrogen bond patterns for AAAAGCTTTT and ATATGCATAT are very similar to their single G•C counterparts. The hydrogen bond between the azabenzimidazole of DB2277 to the guanine-NH₂ functional group is present for > 90% of the simulation in both cases. In the AAAAGCTTTT sequence, DB2277 forms a bifurcated hydrogen bond from the -NH group of the imidazole to the central guanine and to the thymine across the strand one step above as seen in the single G•C sequence (Figure S12A and S12C). The hydrogen bond from the -NH group of the imidazole to guanine is present but no bifurcation is formed in the ATATGCATAT sequence (Table S6 and Figure S12H and S11J). Unlike the single G•C sequences, both amidines of DB2277 directly contact the DNA. In both the AAAAGCTTTT and ATATGCATAT sequences, amidine-1 forms a hydrogen bond with the thymine two steps above the central GpC on the complementary. Amidine-2 forms transient hydrogen bonds to the backbone atoms of the cytosine in the first G•C pair and to the base above it. However, its primary mode of association to the sequences is via water mediated contacts (Figure S12A, C, H, and J). It is clear from Figures S16 and S17 that amidine-2 freely rotates in solution while the core is tightly bound even more so than the single G•C sequences. This could be due to the broad minor groove in this region, which imposes less geometrical restraint on this part of the molecule.

AATTGCAATT is unlike any of the other sequences because it breaks the conserved hydrogen bond between the azabenzimidazole and guanine-NH₂ and slides one step in the 5' direction (Figure 6H and 6I). In this position, the -NH group of the imidazole forms a hydrogen bond to the thymine one step above the central GpC. In the AAAAGCTTTT sequence, the thymine in this position is part of the bifurcation between the imidazole of DB2277 and the central guanine. Amidine-1 forms hydrogen bonds with the thymine two steps above and on the complementary strand (Figure 6I). Amidine-2 forms a weak hydrogen bond (persistence 17.62%) with the cytosine on the reference strand at the central GpC site. Like the other sequences, amidine-2 forms water several mediated contacts to the DNA and is the most freely rotatable (Figure S16E and S17E). By sliding one step above the GpC site, DB2277 can form hydrogen bonds with both amidines to O2 atoms on the pyrimidines (amidine-1 binds to O2 on thymine, amidine-2 binds to O2 on cytosine, Figure 6H and Table S6). These contacts are preferred over the less favorable hydrogen bonds formed in the AATTGAATT sequence between amidine-1 the N3 atom of adenine. The

intrinsic, slightly wider minor groove of AATTGCAATT versus AATTGAATT facilitated the displacement of DB2277 towards the 5' end indicating that once the preferred microstructure is found, DB2277 can stabilize and form direct contacts to the DNA.

The designed azabenzimidazole core of DB2277 selectively recognizes the central G•C in all sequences except for in AATTGCAATT. The preferred binding sequence, AAAAGTTTT, is the only sequence that forms a bifurcated hydrogen with DB2277. First, DNA microstructure for this sequence is intrinsically favorable for small molecule binding. Second, once bound, DB2277 has the ability to form an additional hydrogen bond that stabilizes its binding position. These factors contribute to the observed low K_D .³⁰ In the second most preferred sequence, AATTGCAATT, binding is optimized by sliding away from the central GpC to decrease minor groove deformation and the water displacement penalty. This demonstrates that sequence-specific DNA targeting requires consideration of both the DNA microstructure and interactions with specific functional groups.

Conclusions

Here we present a comprehensive analysis of sequence-specific DNA association of a synthetic minor groove binder. We show that binding occurs via an induced fit mechanism. DNA structural deformation, water displacement and contact formation at the target site all contribute to the observed binding affinities and collectively establish preference for certain DNA sequences. Our findings imply that that planar, isohelical small molecules, like DB2277, seek structurally compatible sites on DNA, which are dependent upon the flanking sequence of the core binding motif. These sites are generally deep and narrow regions of the minor groove that are more rigid than the surrounding DNA. Structured water has to be displaced from these sites in order for the binding to occur. Our analysis shows that the balance between water displacement and minor groove deformation is critical for determining binding affinity and dictates the preferred binding site within the DNA sequence. To stabilize the complex after recognition and association, the amidine groups of DB2277 form direct and water-mediated contacts to the DNA groove, the persistence of which depends on the flanking sequence. Our results clarify certain design principles for achieving selective binding with small molecules. Clearly, it is necessary, but not sufficient, to match the curvature of the minor groove and provide electrostatically complementary functional groups such as amidines. However, it is also critical to take into consideration the local shape and microstructural variation of the target DNA. In addition, the exact positioning of structural water in the minor groove with respect to suitably placed polar functional groups is an important indicator of preferred binding sites. Our study provides a framework for the rational design and optimization of fragment-based drugs designed for site-specific targeting of the DNA minor groove.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Lily Vasileva for her helpful preliminary work on this project and Professor David Boykin for originally supplying DB2277 and for helpful discussions.

This work was supported by National Institutes of Health grant GM110387 [to II] and GM111749 [to WDW].

Computational resources were provided in part by an allocation (CHE110042) from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 and in part by and allocation (m1254) from the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

Notes and references

1. Inukai S, Kock KH and Bulyk ML, *Curr Opin Genet Dev*, 2017, 43, 110–119. [PubMed: 28359978]
2. Jiang X and Yang Z, *Oncotargets Ther*, 2018, 11, 3533–3539. [PubMed: 29950858]
3. Jin Y, Messmer-Blust AF and Li J, *Trends Cardiovasc Med*, 2011, 21, 1–5. [PubMed: 22498013]
4. Rogers JM and Bulyk ML, *Wiley Interdiscip Rev Syst Biol Med*, 2018, DOI: 10.1002/wsbm.1423, e1423.
5. Simone R, Fratta P, Neidle S, Parkinson GN and Isaacs AM, *FEBS Lett*, 2015, 589, 1653–1668. [PubMed: 25979174]
6. Wu R, Zhang QH, Lu YJ, Ren K and Yi GH, *DNA Cell Biol*, 2015, 34, 6–18. [PubMed: 25216212]
7. Bouhrel MA, Lambert M and David-Cordonnier MH, *Curr Top Med Chem*, 2015, 15, 1323–1358. [PubMed: 25866275]
8. Chen BJ, Wu YL, Tanaka Y and Zhang W, *Int J Biol Sci*, 2014, 10, 1084–1096. [PubMed: 25332683]
9. Moretti R and Ansari AZ, *Biochimie*, 2008, 90, 1015–1025. [PubMed: 18343232]
10. Munde M, Kumar A, Peixoto P, Depauw S, Ismail MA, Farahat AA, Paul A, Say MV, David-Cordonnier MH, Boykin DW and Wilson WD, *Biochemistry*, 2014, 53, 1218–1227. [PubMed: 24495039]
11. Raskatov JA, Nickols NG, Hargrove AE, Marinov GK, Wold B and Dervan PB, *Proc Natl Acad Sci U S A*, 2012, 109, 16041–16045. [PubMed: 22988074]
12. Rodriguez J, Mosquera J, Couceiro JR, Vazquez ME and Mascarenas JL, *Chem Sci*, 2015, 6, 4767–4771. [PubMed: 26290687]
13. Hollenhorst PC, McIntosh LP and Graves BJ, *Annu Rev Biochem*, 2011, 80, 437–471. [PubMed: 21548782]
14. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR and Weirauch MT, *Cell*, 2018, 172, 650–665. [PubMed: 29425488]
15. Marr MT 2nd Isogai Y, Wright KJ and Tjian R, *Genes Dev*, 2006, 20, 1458–1469. [PubMed: 16751183]
16. Spitz F and Furlong EE, *Nat Rev Genet*, 2012, 13, 613–626. [PubMed: 22868264]
17. Chen A and Koehler AN, *Science*, 2015, 347, 713–714. [PubMed: 25678644]
18. Darnell JE Jr., *Nat Rev Cancer*, 2002, 2, 740–749. [PubMed: 12360277]
19. Koehler AN, *Curr Opin Chem Biol*, 2010, 14, 331–340. [PubMed: 20395165]
20. Liu Y, Kumar A, Boykin DW and Wilson WD, *Biophys Chem*, 2007, 131, 1–14. [PubMed: 17889984]
21. Majmudar CY and Mapp AK, *Curr Opin Chem Biol*, 2005, 9, 467–474. [PubMed: 16122970]
22. Paul A, Chai Y, Boykin DW and Wilson WD, *Biochemistry*, 2015, 54, 577–587. [PubMed: 25495885]
23. Larsen TA, Goodsell DS, Cascio D, Grzeskowiak K and Dickerson RE, *J Biomol Struct Dyn*, 1989, 7, 477–491. [PubMed: 2627296]
24. Wilson WD, Tanious FA, Barton HJ, Jones RL, Fox K, Wydra RL and Strekowski L, *Biochemistry*, 1990, 29, 8452–8461. [PubMed: 2252904]

25. Berger BJ, Henry L, Hall JE and Tidwell RR, Clin Pharmacokinet, 1992, 22, 163–168. [PubMed: 1559310]
26. Del Poeta M, Schell WA, Dykstra CC, Jones S, Tidwell RR, Czarny A, Bajic M, Kumar A, Boykin D and Perfect JR, Antimicrob Agents Chemother, 1998, 42, 2495–2502. [PubMed: 9756747]
27. Harika NK, Paul A, Stroeve E, Chai Y, Boykin DW, Germann MW and Wilson WD, Nucleic Acids Res, 2016, 44, 4519–4527. [PubMed: 27131382]
28. Laughlin S, Wang S, Kumar A, Farahat AA, Boykin DW and Wilson WD, Chemistry, 2015, 21, 5528–5539. [PubMed: 25703690]
29. Laughlin S and Wilson WD, Int J Mol Sci, 2015, 16, 24506–24531. [PubMed: 26501262]
30. Laughlin-Toth S, Carter EK, Ivanov I and Wilson WD, Nucleic Acids Res, 2017, 45, 1297–1306. [PubMed: 28180310]
31. T. a. C. Macke DA, In molecular modeling of nucleic acids, American Chemical Society, Washington, D.C., USA, 1998.
32. D. A. Case, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, et al., Journal, 2017.
33. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, et al., Journal, 2016.
34. Becke AD, J Chem Phys, 1993, 98, 5648–5652.
35. Davidson ER and Feller D, Chemical Reviews, 1986, 86, 681–696.
36. Singh U. C. a. K., A. P, J Comput Chem, 1984, 5, 129–145.
37. Besler BH, Merz KM Jr., and Kollman PA, J Comput Chem, 1990, 11, 431–439.
38. Wang JM, Wang W, Kollman PA and Case DA, J Mol Graph Model, 2006, 25, 247–260. [PubMed: 16458552]
39. Wang JM, Wolf RM, Caldwell JW, Kollman PA and D. A. Case, Journal of Computational Chemistry, 2004, 25, 1157–1174. [PubMed: 15116359]
40. Spackova N, Cheatham TE, Ryjacek F, Lankas F, van Meervelt L, Hobza P and Sponer J, J Am Chem Soc, 2003, 125, 1759–1769. [PubMed: 12580601]
41. Silva CO, da Silva EC and Nascimento MAC, The Journal of Physical Chemistry A, 2000, 104, 2402–2409.
42. Cammi R and Tomasi J, J Comput Chem, 1995, 16, 1449–1458.
43. Cossi M, Barone V, Cammi R and Tomasi J, Chemical Physics Letters, 1996, 255, 327–335.
44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW and Klein ML, J Chem Phys, 1983, 79, 926–935.
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG, J Chem Phys, 1995, 103, 8577–8593.
46. Ryckaert J, Ciccotti G and Berendsen HJC, J Comput Chem, 1977, 23, 327–341.
47. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR and Wilkins-Diehr N, Comput Sci Eng, 2014, 16, 62–74.
48. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S and Walker RC, J Chem Theory Comput, 2013, 9, 3878–3888. [PubMed: 26592383]
49. Krepl M, Zgarbova M, Stadlbauer P, Otyepka M, Banas P, Koca J, Cheatham TE, Jurecka P and Sponer J, J Chem Theory Comput, 2012, 8, 2506–2520. [PubMed: 23197943]
50. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE and Simmerling C, J Chem Theory Comput, 2015, 11, 3696–3713. [PubMed: 26574453]
51. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Loughton CA and Orozco M, Biophys J, 2007, 92, 3817–3829. [PubMed: 17351000]
52. Roe DR and Cheatham TE 3rd, J Chem Theory Comput, 2013, 9, 3084–3095. [PubMed: 26583988]
53. Humphrey W, Dalke A and Schulten K, J Mol Graph, 1996, 14, 33–38, 27–38. [PubMed: 8744570]
54. Lavery R, Moakher M, Maddocks JH, Petkeviciute D and Zakrzewska K, Nucleic Acids Research, 2009, 37, 5917–5929. [PubMed: 19625494]
55. MATLAB (R2017b), The MathWorks Inc, Natick, MA, 2017.

56. Lavery R, Maddocks JH, Pasi M and Zakrzewska K, *Nucleic Acids Research*, 2014, 42, 8138–8149. [PubMed: 24906882]
57. Nguyen CN, Young TK and Gilson MK, *J Chem Phys*, 2012, 137, 044101.
58. Ramsey S, Nguyen C, Salomon R, Walker R, Gilson M and Kurtzman T, *Abstr Pap Am Chem S*, 2017, 254.
59. Nguyen CN, Cruz A, Gilson MK and Kurtzman T, *J Chem Theory Comput*, 2014, 10, 2769–2780. [PubMed: 25018673]
60. Yonetani Y and Kono H, *Biophys J*, 2009, 97, 1138–1147. [PubMed: 19686662]
61. Johannesson H and Halle B, *J Am Chem Soc*, 1998, 120, 6859–6870.
62. Privalov PL and Crane-Robinson C, *Eur Biophys J*, 2017, 46, 203–224. [PubMed: 27457765]
63. Haq I, *Archives of Biochemistry and Biophysics*, 2002, 403, 1–15. [PubMed: 12061796]
64. Halle B and Denisov VP, *Biopolymers*, 1998, 48, 210–233. [PubMed: 10699841]
65. Woods KK, Lan T, McLaughlin LW and Williams LD, *Nucleic Acids Research*, 2003, 31, 1536–1540. [PubMed: 12595562]
66. Woods KK, Maehigashi T, Howerton SB, Sines CC, Tannenbaum S and Williams LD, *J Am Chem Soc*, 2004, 126, 15330–15331. [PubMed: 15563130]
67. Kopka ML, Fratini AV, Drew HR and Dickerson RE, *Journal of Molecular Biology*, 1983, 163, 129–146. [PubMed: 6834428]
68. Wei D, Wilson WD and Neidle S, *J Am Chem Soc*, 2013, 135, 1369–1377. [PubMed: 23276263]
69. Mazur S, Tanious FA, Ding D, Kumar A, Boykin DW, Simpson IJ, Neidle S and Wilson WD, *Journal of Molecular Biology*, 2000, 300, 321–337. [PubMed: 10873468]
70. Nguyen B, Tanious FA and Wilson WD, *Methods*, 2007, 42, 150–161. [PubMed: 17472897]
71. Nanjunda R and Wilson WD, *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]*, 2012, Chapter 8: Unit 8.
72. Harika NK and Wilson WD, *Biochemistry*, 2018, 57, 5050–5057. [PubMed: 30048590]

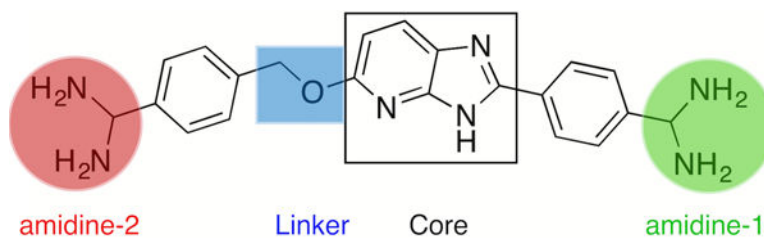


Figure 1.

Chemical structure of DB2277. DB2277 is a heterocyclic compound designed to target G•C in poly AT sequences via a single hydrogen bond between the central azabenzimidazole (black box, core) and the G-NH₂ group. A phenylamidine is bound to the imidazole (green circle, amidine-1) on one side. Attached to the left of the azabenzimidazole is a flexible linker, -OCH₂-, which joins a second phenylamidine to the azabenzimidazole (red circle, amidine-2).

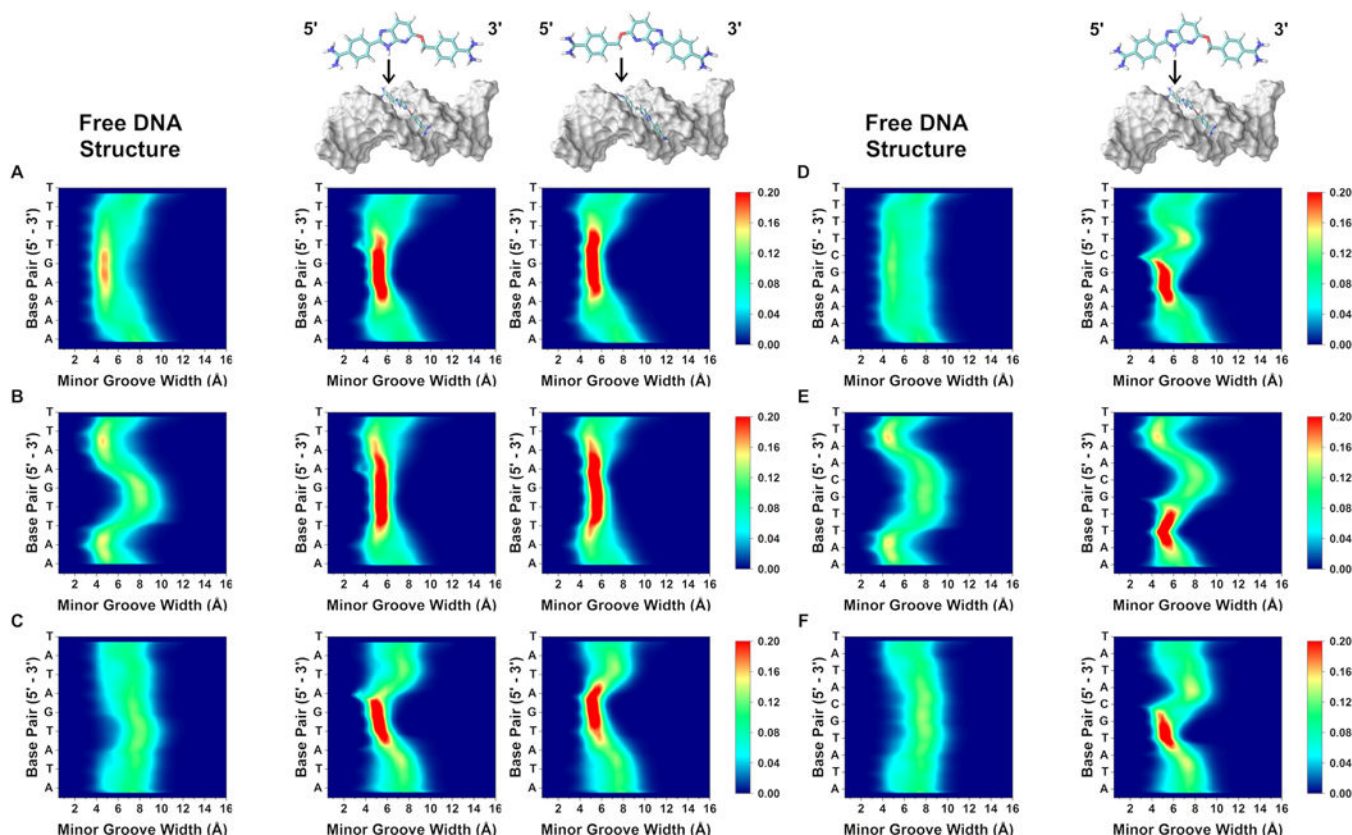


Figure 2.

Minor groove width of the free and bound structures. 2D contour histograms of minor groove width per base pair for the free and bound simulations. (A) AAAAGTTTT, (B) AATGAATT, (C) ATATGATAT, (D) AAAAGCTTTT, (E) AATGCAATT, (F) ATATAGCATAT. The color gradient indicates increasing probability (blue to red) of width in Angstroms (Å). (A, B, C) The single G•C sequences are non-palindromic and the asymmetric small molecule DB2277 was positioned in two orientations in the minor groove. In the 5' → 3' orientation, amidine-1 is closest to the 5' end of the target sequence. In the 3' → 5' orientation, amidine-1 is closest to the 3' end of the target sequence. (D, E, F) Only one bound complex was simulated for the GpC sequences because they are palindromic. DB2277 binds via an induced fit binding mechanism

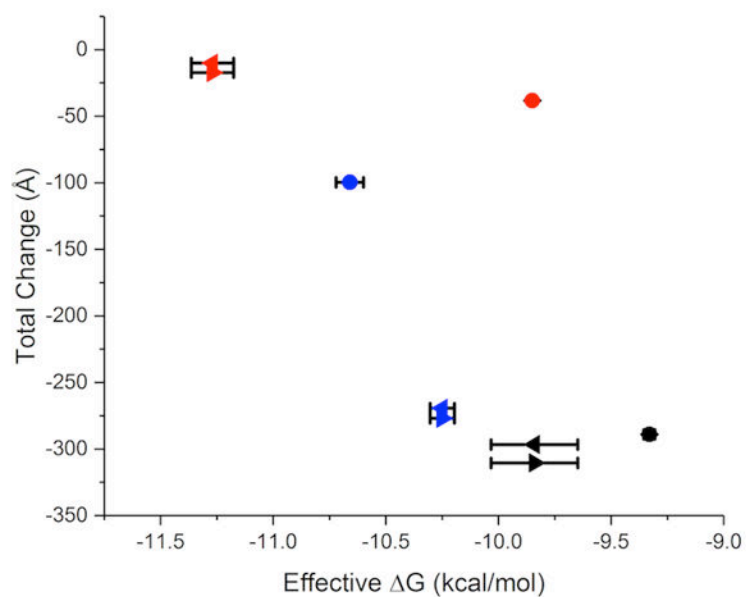


Figure 3.

Minor groove width deformation correlates with effective binding free energy. Effective binding free energy decreases with increased integrated change in minor groove width. Sequences AAAAGTTTT and AAAAGCTTTT are in red, AATTGAATT and AATTGCAATT are in blue and ATATGATAT and ATATGCATAT are in black. Circles represent palindromic (GpC) sequences. Triangles indicate the two orientations of the nonpalindromic single G•C sequences (5'→3' ►) (3'→5' ◄).

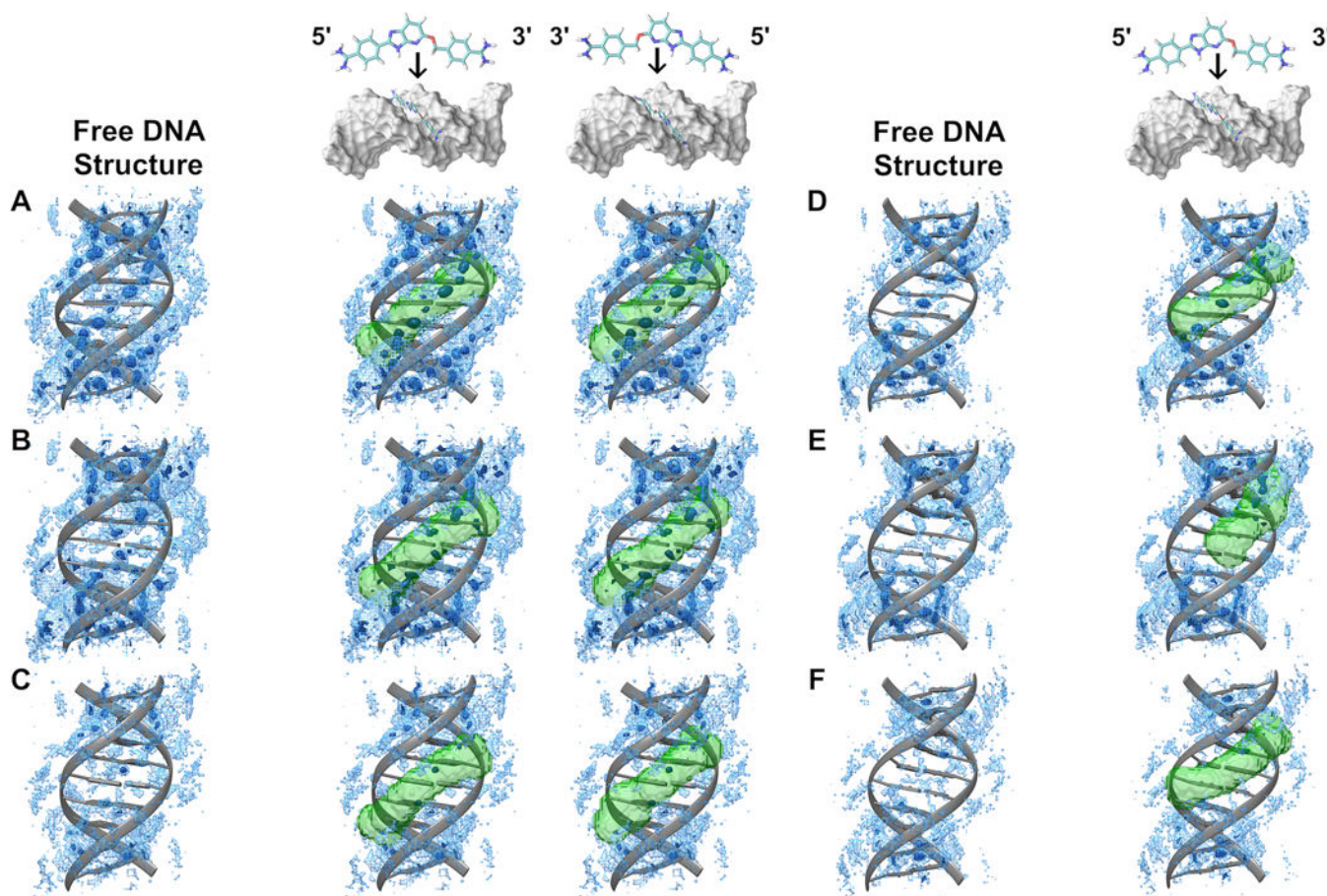


Figure 4.

Structured water associated with free and bound DNA complexes. The effect of sequence on structured water within the minor groove is shown above. (A) AAAAGTTTT, (B) AATTGAATT, (C) ATATGATAT, (D) AAAAGCTTTT, (E) AATTGCAATT, (F) ATATAGCATAT. Favorable waters ($\Delta G < 0 \text{ kcal mol}^{-1}$) are visualized as isodensity (isovalues blue mesh: -0.005 , blue surface: -0.012) for each sequence of interest. The binding position of the asymmetric small molecule, DB2277, in the non-palindromic single G•C sequences (A, B, C) is defined by the green transparent surface for both orientations. In the $5' \rightarrow 3'$ orientation, amidine-1 is closest to the $5'$ end of the target sequence. In the $3' \rightarrow 5'$ orientation, amidine-1 is closest to the $3'$ end of the target sequence. Only one bound orientation, defined by green transparent surface, was simulated for the palindromic GpC sequences (D, E, F).

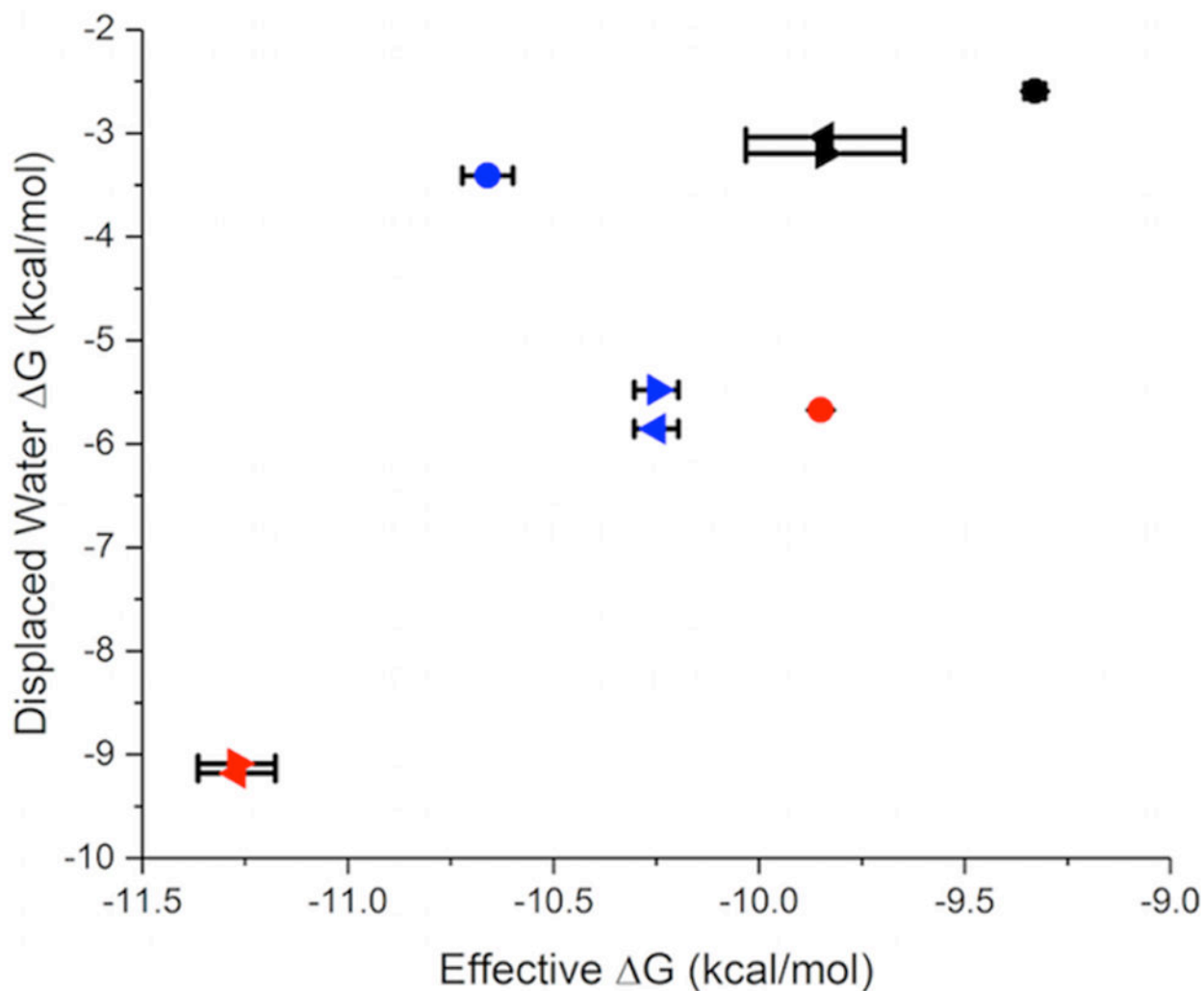


Figure 5.

GIST displaced water free energies correlate with effective binding free energies. Effective binding free energy increases with increased free energy of displaced water. AAAAGTTTT and AAAAGC sequences are in red, AATTGAATT and AATTGCAATT are in blue, and ATATGATAT and ATATGCATAT are in black. Circles indicate palindromic (GpC) sequences. Triangles indicate single G•C sequences with two binding orientations (5'→3' ►) (3'→5' ◄).

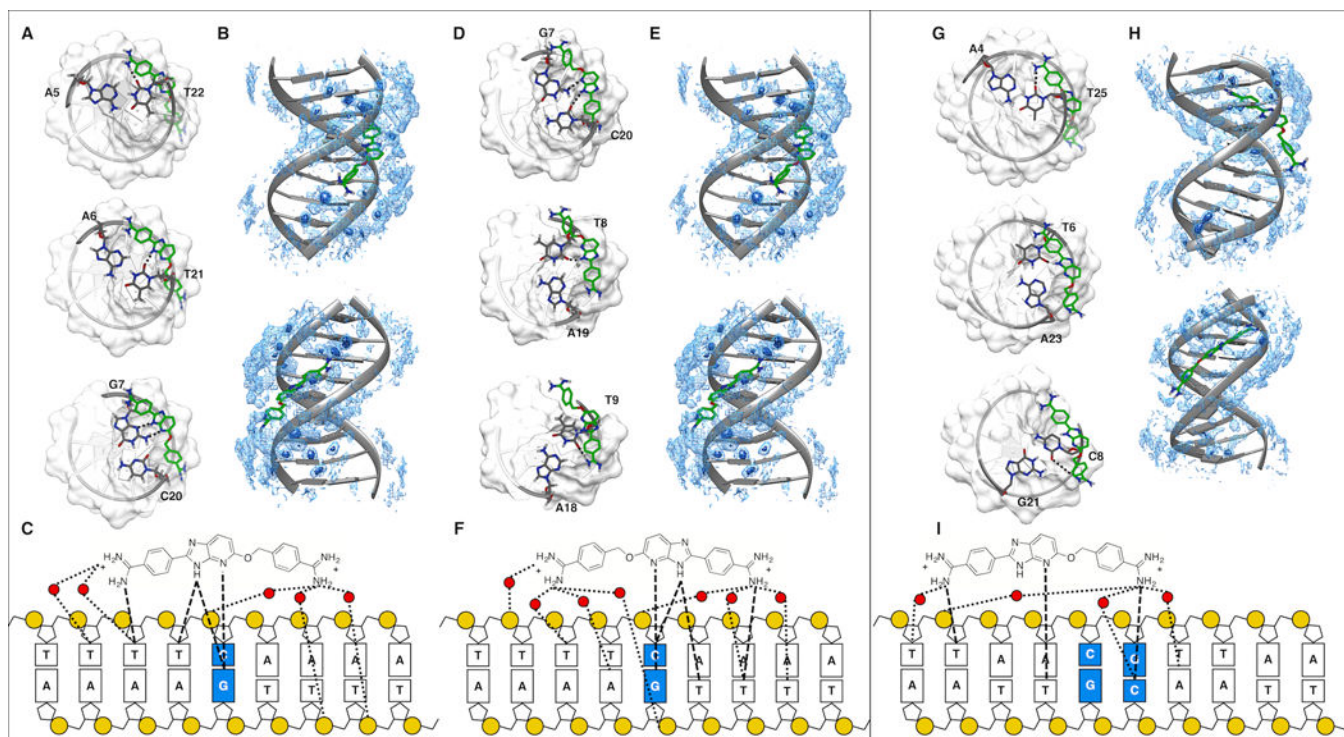


Figure 6.

Hydrogen bonds of DB2277 with DNA. DB2277 contacts the DNA via direct hydrogen bonds and water-mediated contacts. (A, B, C) DB2277 bound to AAAAGTTTTT in the 5' → 3' orientation. (D, E, F) DB2277 bound to AAAAGTTTTT in the 3' → 5' direction. (G-I) DB2277 bound to AATTGCAATT (A, D, G) DB2277 forms persistent contacts to three base pairs of the sequence with amidine-1 and the core azabenzimidazole. (B, E, H) The association of amidine-2 is stabilized by several water-mediated contacts visualized as isodensity (isovalues blue mesh: -0.009, blue surface: -0.02). (C, F) In the AAAAGTTTTT sequence, the core azabenzimidazole of DB2277 forms multiple persistent contacts at the binding site in both the 5' → 3' and the 3' → 5' orientations. (I) To optimize binding, DB2277 shifts one step upstream in the AATTGCAATT sequence and does not make contacts with the target guanine.