Published in final edited form as:

Chemistry. 2017 December 11; 23(69): 17612–17620. doi:10.1002/chem.201704563.

First Structure of a Designed Minor Groove Binding Heterocyclic Cation that Specifically Recognizes Mixed DNA base pair Sequences

Narinder K. Harika^[a], Markus W. Germann^{*,[a]}, and W. David Wilson^{*,[a]}

^[a]Department of Chemistry, Georgia State University, Atlanta, GA 30303-3083, USA

Abstract

The high-resolution NMR structure of the first heterocyclic, non-amide, organic cation that strongly and selectively recognizes mixed AT/GC bp sequences of DNA in a 1:1 complex is described. Compound designs of this type provide essential methods for control of functional, non-genomic DNA sequences and have broad cell uptake capability, based on studies from animals to humans. The high-resolution structural studies described in this report are essential for understanding the molecular basis for the sequence-specific binding as well as for new ideas for additional compound designs for sequence-specific recognition. The molecular features, in this report, explain the mechanism of recognition of both A•T and G•C bps and are an interesting molecular recognition story. Examination of the experimental structure and the NMR restrained molecular dynamics model suggests that recognition of the G•C base pair involves two specific H-bonds. The structure illustrates a wealth of information on different DNA interactions and illustrates an interfacial water molecule that is a key component of the complex.

Graphical Abstract

Keywords

DNA; minor groove binder; NMR spectroscopy; mixed base pair recognition; solution structures

INTRODUCTION

Most known, non-polyamide minor groove compounds that bind reversibly to DNA are AT sequence specific.^[1,2] These include reference standards such as netropsin, stains such as

*Address correspondence to either author: W. David Wilson Tel: 404-413-5503; Fax: 404-413-5505; wdw@gsu.edu, Markus W. Germann Tel: 404-641-4981; Fax: +1 404 413 5561; mwg@gsu.edu.

Hoechst dyes, DAPI and others, and therapeutics including many heterocyclic cations.^[3–5] While these compounds have been quite successful in modulating specific biological functions, they are also quite limited for many applications by their lack of diversity in the sequence specific DNA recognition.^[6,7] As one approach to widen the sequence recognition of these compounds, we have introduced H-bond accepting modules to A•T recognizing groups to explore G•C (bp) recognition in the context of heterocyclic cation structures.^[8–10]

The first step in this process, adding an H-bond accepting module for recognition of a G•C bp with flanking A•T bps, has been successfully accomplished with several quite different, new compound designs. One new design contains a central pyridine H-bond acceptor, DB2120 (Figure 1 (inset)). The compound covers essentially a full turn of the double helix, and specifically and strongly recognizes a single G•C bp with flanking A•T sequences.^[8,11] An N-methyl benzimidazole (N-MeBI) module has also been used as an H-bond acceptor when paired with a thiophene to create a sigma-hole structured interaction. This module also gives strong and selective recognition of a G•C bp.^[12]

The compound investigated in this report contains an aza-benzimidazole (aza BI) H-bond accepting group, DB2277, (Figure 1). This molecule also recognizes a single G•C bp with flanking AT sequences.^[13–15] In order to extend the compound design efforts for single G•C bp as well as for extending the recognition to two or more G•C bp in more complex sequences, structural information on mixed sequence complexes is essential. Since we have not been able to obtain diffraction-quality crystals of DB2277-DNA complexes, high-resolution NMR experiments have been used in the structural analysis. A number of NMR studies of benzimidazole (BI) derivatives of Hoechst 33258 have been done with A•T binding sites and they show strong binding in the minor groove with much weaker binding at sites with G•C bps.^[16–19] Other minor groove binders, pentamine, berenil and SN6999, also bind strongly to AT minor groove sequences.^[20–22] They all form H-bonds to acceptor groups on the A•T bp edges and fit the shape and width of the minor groove quite well. To this time, however, no previous NMR structural studies have been reported for heterocyclic cations with a G•C bp containing sequence. Since these non-amide compounds have excellent cell uptake and biological activities, there is a compelling need to develop analogs with broad DNA sequence recognition.^[23,24]

A sequence that has been found to bind DB2277 strongly and in a single orientation has an –AAGATA– binding site (Figure 1B). This sequence has been used in the high-resolution NMR investigations of the DB2277–AAGATA complex structure reported here. Both 1D and 2D ¹H and ³¹P spectra were collected in H₂O and D₂O (buffer solution). The NMR results were used in restrained molecular dynamics (rMD) calculations to determine the structure of the DB2277 complex. Examination of the structure indicates that the aza BI group forms two strong H-bonds with the G•C bp. Based on the NMR restrained molecular dynamics model the complex is extensively hydrated and one amidine is linked to the DNA base pairs at the floor of the minor groove through a water interface. The other amidine forms a direct H-bond to the bases. The most dynamic part of the bound DB2277 is the H₂O-linked amidine. The entire hydration layer is also quite dynamic. While a variety of classical AT specific compounds have had the structure of their DNA complexes determined by both NMR and X-ray,^[25–28] this is the first compound in the heterocyclic cation family

whose structure has been determined when bound to a mixed AT/GC sequence. The mechanism of GC binding is unique and illustrates new molecular recognition features that will be very useful for designing compounds to target specific sequences of DNA.

RESULTS

NMR:

In a previous 1D NMR study we reported the binding of DB2277 to a wide range of mixed DNA sequences containing a single, central G•C base pair, flanked by A•T bp tracts.^[14] The DNA sequence with an –AAGATA– binding site showed a selective binding orientation with DB2277 at a 1:1 binding ratio and is, therefore, the sequence selected for high resolution NMR structural analysis presented here. The detailed NMR results including 1D and 2D 1H NMR spectra for both exchangeable and non-exchangeable protons, were obtained at 600 and 850 MHz. This complex gave excellent high-resolution spectra for structural insight into DB2277 binding in the minor groove of the mixed base pair DNA sequence. ³¹P 1D and 2D spectra were also obtained for assignment assistance and structural analysis. The detailed methods for these NMR results are given in the supplementary material. The full sequence of the DNA hairpin used and the DB2277 structure are shown in Figure 1 with their numbering schemes.

Imino proton spectra:

The –AAGATA– DNA hairpin complex with DB2277 displays well-resolved imino proton signals for 5 A•T and 4 G•C bp of DNA along with a broad upfield band for the unpaired T imino protons of the loop region (Figure 2). DB2277 was added up to a 1:1.4 DNA:DB2277 binding ratio into the solution containing DNA hairpin. After the addition of 0.4 equivalents of DB2277, free DNA signals start to disappear with the emergence of new peaks for the DB2277-DNA complex. This co-existence of free DNA and DB2277-DNA complex peaks demonstrates slow exchange between free DNA and the complex on the NMR chemical shift timescale. The absence of any free DNA peak at a 1:1 binding ratio also indicates tight binding of DB2277 with the DNA hairpin. Imino proton signals were assigned using 2D NOESY in 90% H₂O/10% D₂O. A single and selective binding orientation of the DB2277-DNA complex is observed in imino proton spectra with single resonance peaks for each base pair (Figure 2).^[14] 1D proton NMR experiments of the DNA hairpin with ¹⁵N-labeled central G in the –AAGATA– binding site were conducted with DB2277 and reported previously.^[14] The findings from this experiment along with 1D NOE experiments conducted on the DB2277-DNA complex allow the assignment of the peak at 11.5 ppm to a downfield shifted aza-benzimidazole-NH (BI -NH) proton signal of DB2277.^[14]

2D NMR:

NOESY spectra in D₂O buffer were collected to assign the free DNA and DB2277-DNA complex signals at 1:1 binding ratio. Assignments were confirmed by NOESY experiments conducted at both 600 and 850 MHz NMR at different mixing times (50 ms and 100 ms). NOESY experiments in 90% H₂O/10% D₂O were used to observe the specific binding of the exchangeable BI -NH proton of DB2277 with DNA.^[14] Standard protocols for B-DNA assignment are used to assign DNA peaks.^[29,30] A sequential walk from the aromatic H6/H8

proton to the anomeric H1' sugar protons of adjacent bases was used for the assignments (Figure 3 and S2). Protons of free DNA, free DB2277 and the DB2277-DNA complex in solution were assigned from the NOESY spectrum in D₂O solution. These proton assignments were further confirmed and supported by employing 2D TOCSY, ³¹P-correlated 2D spectra (HPCOR) and ¹³C-¹H HSQC experiments (see details in the Methods section). A proton-decoupled ³¹P spectrum with phosphorus assignments for free DNA is shown in Figure S3A and chemical shifts of a ³¹P (δP in ppm) signal for both free DNA and DB2277-DNA complex are plotted in Figure S3B.

B1 and B1' proton signals of DB2277 (Figure 1) were observed to resonate at the same chemical shift in the DNA-DB2277 complex spectra at a 1:1 binding ratio (as can be seen in Figure 7 below). Similar degeneracy was observed for B2, B5 and B6 protons with B2', B5' and B6' protons respectively. This suggests that these protons interchange due to the rapid flipping of the phenyl rings in bound DB2277. Despite their degeneracy, assignments of B1 versus B1' overlapped proton pairs could be distinguished based on their location in the DB2277-DNA structure (as can be seen in the structure described below) and their vicinity to the specific DNA protons in the minor groove. For example, careful investigation of the structure of DB2277-DNA complex mentioned in detail below helps in assigning the phenyl proton in proximity to H2 protons of A3 and A4 as B1 whereas the phenyl proton close to the H4'/H5'/H5'' proton of G5 and A6 as B1'. Other phenyl protons of DB2277 i.e. B2 vs. B2', B5 vs. B5' and B6 vs. B6' were assigned similarly. Chemical shifts of NMR signals obtained from the free DB2277 protons and the DB2277 protons bound to DNA are listed in Table S1. Assignments of free DNA and the DB2277-DNA complex protons in the "NOE walk" of H6/H8 base-H1' sugar region are shown in the supplementary Figure S2 and Figure 3 respectively.

Chemical shift perturbations and ring current effects:

The chemical shift differences between free DNA and DB2277-DNA complex for H1', H2' and H4' sugar protons, H2 protons of adenine and methyl protons of thymine are shown in Figure 4. Chemical shift differences of H2'', H3' sugar protons, H6/H8 base protons and H5 proton of cytosine are shown in Figure S4A, and imino proton shift differences are shown in Figure S4B. Significant chemical shift differences are observed for the protons close to the central G in the minor groove of DNA as compared to the protons present in the flanking regions in the DB2277-DNA complex (Figure 4).

The significant upfield and downfield shifts can be explained by the ring current effect induced by the aromatic groups of DB2277 on DNA resonances. Nuclei present above and below the aromatic ring experience significant upfield shifts and nuclei in the plane of the ring experience significant downfield shifts. The significant upfield shift of H1', H2' and H4' sugar protons is due to the mutual interaction of aromatic groups of DB2277 with DNA protons which depends upon their relative orientation in space. The upfield shifts of H1' and H2' sugar protons of G5, A6 and C18 in the binding site results from shielding from aromatic moieties in DB2277. The significant upfield shift (around 1.0 ppm) of H4' sugar protons in the binding site of the DB2277-DNA complex is also evident in Figure 4. The H4' proton of G5 is situated on top of the phenyl ring attached to the OCH₂ group whereas

the H4' proton of its complementary base (C18) is situated below the phenyl ring attached to the aza BI moiety as evident in Figure 5A. The upfield shift of H4' protons of A6 and T7 is also due to their positioning in the plane of the aromatic moieties in DB2277.

The downfield shift of H2 protons of adenine (Figure 4) can be explained in a similar fashion by their position in the de-shielded region of conjugated systems in DB2277. The pronounced downfield shift effect experienced by the H2 proton of A6 in the binding site is due to the de-shielding caused by close positioning of this proton at the edge of a phenyl ring of DB2277 (Figure 4). Small chemical shift differences were observed for protons in the flanking region of the DNA which are not in the vicinity of DB2277 (Figure S4).

A loss of connectivity through the H1' sugar protons of T7 and T20 in the sequential walk from H1' to base protons (Figure 3) is most likely due to the dynamics involved in the rapid flipping of phenyl rings of DB2277. The rapid flipping of phenyl rings of DB2277, as observed from NMR data (as can be seen in Figure 7 below), results in the interchange of phenyl ring protons as mentioned above. The observed loss of signals from these H1' protons is supported by the induced ring current effects due to the relative orientation of the H1' protons of T7 and T20 in relation to the aromatic rings of DB2277. Additionally, positioning of the H1' proton of T7 right below the phenyl ring and the H1' of T20 right above the aza BI group is evident in Figure 5B.

NMR restraints:

Intramolecular NOEs of DNA and DB2277 along with the intermolecular NOEs between DNA and DB2277 were used for the structure calculations of DB2277-DNA complex. Intermolecular NOEs observed between the DB2277 protons and the DNA protons obtained from NOESY data are listed in Table 1. Strong intermolecular NOE contacts are observed for the DB2277 aromatic protons with A4, A6 and A16 protons of DNA and their assignments are listed in Table 1. The interactions of the DB2277 protons with DNA protons are also represented in the schematic model shown in Figure 6.

The BI -NH proton of DB2277 shows strong intermolecular contacts with the imino proton of G5 in the DNA-DB2277 complex at a 1:1 binding ratio.^[14] Similarly, B5 and B6 phenyl protons also show strong intermolecular NOE contacts with the imino proton of G5 in the DNA-DB2277 complex (Table 1). The assignments for the strong intermolecular cross peaks observed between the H2 of adenine and phenyl protons of DB2277 in NOESY spectrum are shown in Figure 7. Strong intermolecular NOEs are also observed between H1' of G5 and B2 phenyl proton of DB2277 (Figure 3). The H1' proton of C18 also shows strong intermolecular NOE contacts with B5 and B6 phenyl protons and NOEs are marked in the NOESY spectrum in Figure 3. These important intermolecular contacts also localize the specific binding of DB2277 with the central G in the minor groove of DNA.

The B1 and B2 phenyl protons of DB2277 show strong intermolecular contacts with the H2 protons of A3 and A4 and with the H1' proton of G5. These NOEs reveal that the B1 and B2 protons of the phenyl ring in DB2277 are pointed towards the floor of the minor groove (Figure 7). The B1' and B2' phenyl protons show NOE with H5' of A6 which reveals that these phenyl protons are facing more towards the top of the minor groove (Table 1).

Intermolecular cross peaks observed between the B5' and B6' phenyl protons of DB2277 and the H4'/H5' of C18 and T19 of DNA verify that these protons also face towards the top of the minor groove of DNA (Table 1). The intermolecular contacts of the B5 and B6 protons with H1' of C18 and H2 of A6 show that these protons are pointing towards the floor (bases) of the minor groove of DNA (Figure 7 and Table 1).

Restrained Molecular Dynamics Simulations (rMD):

MD simulations with NMR restraints were performed on both free DNA and the DB2277-DNA complex. An initial B form DNA model was generated in AMBER. The detailed methods for rMD calculations are given in the supplementary material.

The parametrization of DB2277 is described in the supplementary section. The optimized geometry of DB2277 is used in the initial structure for the DB2277-DNA complex. Visual inspection of various docked conformations from autodock-vina 4.0 helped to select the orientation of DB2277 in relation to DNA based on NMR restraints. The strong intermolecular contacts of H2 protons of adenine with phenyl protons of DB2277 obtained from NOESY, assisted in defining the initial orientation of DB2277 in relation to DNA (Figure 7). For example, H2 protons of A3 and A4 are located in proximity to the phenyl ring attached to the OCH₂ group and H2 protons of A6 and A12 are located in the vicinity of the other phenyl ring which is attached to the aza BI.

Distance restraints from NMR data were calculated and incorporated along with other restraints as described in the Methods section. rMD simulations were performed on the free DNA and DB2277-DNA complex and a rmsd plot for free DNA in the –AAGATA– binding site from 4 ns of rMD simulation is shown in Figure S5A. The heavy atom rmsd value for free DNA in the binding site is 1.7 Å (Figure S5A). The average structure from the 4 ns rMD is used as a reference frame for the rmsd calculation.

An rmsd plot for the DB2277-DNA complex is shown in Figure S5B. A heavy atoms rmsd (root-mean-square deviation) value for the DB2277-DNA complex from 2 ns of rMD simulation is 0.6 Å (Figure S5B). The rmsd value for DB2277-DNA is in good agreement with the values obtained from previous DNA minor groove complex studies.^[16,31] The average structure from the 2 ns rMD is used as a reference for rmsd fitting. Snapshots were obtained every 10 ps for the last 100 ps of rMD calculations for both free DNA and the DB2277-DNA complex. Energy minimization with NMR restraints was subsequently performed on these converged conformers in explicit solvation to remove any random thermal fluctuations. Superimposition of these final ten converged structures is shown in Figure S6 for free DNA and in Figure 8 for the DB2277-DNA complex. More fluctuations were observed for the terminal G•C bp for both free DNA and the DB2277-DNA complex. The increase in flexibility at the ends of the double helix is a common observation in MD simulations of oligonucleotides.^[32] The NMR refinement statistics are provided in Table 2. The free -AAGATA- DNA structure has been deposited with accession numbers 6ASF (PDB) and 30335 (BMRB). DB2277-DNA structure has been deposited with accession numbers 6AST (PDB) and 30336 (BMRB).

The final structure for analysis is obtained by averaging all the production run snapshots for the last 100 ps rMD simulation for both free DNA and the DB2277-DNA complex. This structure was subsequently refined with energy minimization. A minor groove view of the minimized average structure of the DB2277-DNA complex is shown in Figure S7. Furthermore, the DB2277-DNA equilibrated structure from the rMD simulation was also used for 200 ns of MD simulations without NMR restraints. Findings from the free MD calculations suggest that the DB2277-DNA complex remains stable for the simulation time period and DB2277 stays tightly bound into the minor groove of DNA in a single selective binding orientation.

Analysis of helicoidal parameters of DNA from rMD simulations:

The minor groove widths for free DNA and the DB2277-DNA complex were calculated from trajectories for the last 100 ps of rMD simulations. A comparison of minor groove widths for free DNA and the DB2277-DNA complex is shown in Figure S8. The minor groove widths are similar for the free and complex DNA but there is a slight clamping down of the complex at the GC bp where the aza BI has two H-bonds for strong binding and specific recognition. Plots for the roll and twist angles for free DNA and DB2277-DNA complex are shown in Figure S9 (A) and (B) respectively. These angles are similar to the ones for the free DNA and do not show any systematic changes.

Epsilon (ϵ) and zeta (ζ) torsional angles differentiate BII backbone conformations ($\epsilon\text{-}\zeta > 0^\circ$) from the canonical BI conformations ($\epsilon\text{-}\zeta < 0^\circ$) of DNA.^[33] The equilibrium ratios of BI and BII conformational states of DNA in solution by NMR have been extensively investigated.^[34,35] To evaluate possible BII states in the free –AAGATA– and complex sequence $J_{H3',P}$ coupling constants were determined experimentally from NMR to calculate ϵ torsional angles (Table S2).^[36] Calculated ϵ values were used for rMD simulations for both free DNA and the DB2277-DNA complex. $\epsilon\text{-}\zeta$ values calculated for selected dinucleotide steps towards the 5' end side of the DNA from the last 100 ps of rMD simulations for free DNA and DB2277-DNA complex are shown in Figure S10. Other bases also show little or no variation in their $\epsilon\text{-}\zeta$ values from the standard BI state. Values for all dinucleotide steps fluctuate around -90° except the C2pA3 step that shows more variation in $\epsilon\text{-}\zeta$ values (Figure S10) but remains in the BI region.

DISCUSSION

Our initial results with DB2277 demonstrated strong and specific binding of the compound to single G•C bp DNA sequences with flanking A•T bps.^[13–15] The results also illustrated the dependence of the binding on the compound structure. Isomers and close analogs of DB2277 bound much more weakly and with low specificity. What these results did not reveal was the molecular basis of the specific recognition of single G•C bp DNA sequences by DB2277, which is addressed by the NMR and MD structural analysis reported here.

The –AAGATA– DNA binding site gives a unique complex with DB2277 that is amenable for high-resolution NMR experiments.^[14] Significant chemical shift changes are observed for both DB2277 and DNA on complex formation. For DNA the largest shifts are in the –AAGATA– sequence in agreement with that sequence as the DB2277 binding site. The 2D

NMR results provided distances in both the free DNA and DB2277 as well as interatomic distances in the complex that were incorporated into restrained MD calculations to provide a structure for the complex. The results from the rMD calculations clearly show that DB2277 fits tightly into the minor groove at the G•C bp in the –AAGATA– sequence (Figure 9A and S7).

The first question that we wish to address from the structure model is how does DB2277 selectively recognize a G•C bp in a DNA sequence? The experimental structure models shown in Figures 5, 9 and the schematic in Figure 6 provide a clear answer. In the NMR restrained molecular dynamics model the G5-NH in the minor groove of –AAGATA– forms a strong H-bond with the aza-N of the aza BI group with the H-bond distance of 2.1 Å. Since this interaction was part of the original design this observation is quite informative.^[13] The DB2277-G•C interaction is locked down by formation of an –NH to C=O 2.1 Å H-bond between the BI –NH and the keto group of C18 in the minor groove based on the model shown in Figure 5A. The downfield shift of the BI –NH proton due to the strong H-bond with O2 of C18 is also supported by an NMR study conducted by Leupin and co-workers on Hoechst 33258 with A•T sequences.^[14,17] These two H-bonds form strong and specific interactions between DB2277 and the G•C bp that accounts for the specific GC binding. The second question is how are the flanking AT sequences recognized? As can be seen in the rMD model in Figure 9C, the amidine on the phenyl-amidine attached to the aza BI group forms a direct H-bond to a T keto group (H-bond distance of 2.2 Å) in the minor groove (T=O ●●●●● H-N amidine). The model also suggests that the amidine-DNA interaction is stabilized by an ensemble of dynamic H-bonds to water molecules in the minor groove. The amidine-phenyl-aza benzimidazole module, thus, forms very strong, specific and favorable interactions with the target DNA sequence. The amidine at the other end of DB2277 interacts much differently with the bases at the floor of the minor groove. The amidine –NH groups are too far from the bases to form direct H-bonds. Instead, there is an interfacial water molecule as suggested by the model in Figure 9B that links an amidine-NH to a T=O at the floor of the minor groove (–NH ●●●●● O-H ●●●●● O=T). This amidine group is also stabilized by a dynamic extended water network in the minor groove (Figure 9B). Interfacial water molecules are a key component of many protein-DNA complexes (see for example Poon, G.M.K. et al., 2017 and references therein) but are quite rare in the minor groove complexes of small organic cations.^[37,38]

Two other features of the DB2277 molecule provide significant assistance to the strong binding in the minor groove complex. The –O-CH₂– linkage provides appropriate spacing and flexibility to the system to track along the shape of the minor groove and appropriately index the functional groups of DB2277 with those of DNA. Removal of this linker gives a molecule with lower affinity for single GC sequences.^[13] Another stabilizing feature of the complex is from the phenyl C-H protons that are near the floor of the minor groove. These aromatic protons carry a small positive charge and can form a stabilizing interaction with A-N3 and T=O groups in the minor groove (Figure 9C).^[21] These DB2277-DNA interactions coupled to the dynamic stabilizing hydration network provide a very favorable complex for specific DNA recognition.

CONCLUSIONS

Our previous studies have shown specific recognition of a single G•C bp flanked by AT sequences by DB2277 in the minor groove of DNA. The DB2277 molecule shows a single and selective binding orientation with the –AAGATA– binding site of DNA. The molecular basis of the specific recognition of a single G•C bp in the –AAGATA– binding site of DNA by DB2277 is clearly defined by high-resolution NMR. Molecular dynamics calculations using NOE restraints from NMR provide structural insight into the unique DB2277-AAGATA complex. This is the first NMR structure for the DNA complex of the heterocyclic cation family compounds with a mixed AT/GC sequence. The NMR restrained MD results indicate that the strong interactions of G5-NH with the aza-N of the aza BI group and C18-O2 with BI –NH of the aza BI group are responsible for the specific recognition and tight fit of DB2277 at the G•C bp in the minor groove of DNA. The results also indicates that the DB2277-AAGATA complex is further stabilized by dynamic hydration network and an interfacial water molecule that plays a crucial role in mediating the amidinium proton-DNA base pair interactions. This information will be very useful in future design efforts of sequence specific minor groove binders.

EXPERIMENTAL SECTION

Sample preparation, 1D and 2D NMR experiments, Distance restraints calculations, DB2277 force field parametrization and Molecular Dynamics (MD) Simulations used in this article can be found in the Supporting Information. Supporting information for this article is available on the WWW.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We are grateful to the National Institutes of Health for support [grant number NIH R01 GM111749 to W.D.W and D.W.B] and to the Molecular Basis of Disease Fellowship. We thank Professor David W. Boykin for many helpful discussions and DB2277. Alex Spring and Marina Evich are thanked for useful discussions. We thank Carol Wilson for manuscript assistance.

REFERENCES

- (1). Trent JO, Clark GR, Kumar A, Wilson WD, Boykin DW, Hall JE, Tidwell RR, Blagburn BL and Neidle S, Journal of Medicinal Chemistry 1996, 39, 4554–4562. [PubMed: 8917643]
- (2). Abu-Daya A, Brown PM and Fox KR, Nucleic Acids Research 1995, 23, 3385–3392. [PubMed: 7567447]
- (3). Wemmer DE, Annual Review of Biophysics & Biomolecular Structure 2000, 29, 439.
- (4). Khalaf AI, Al-Kadhimi AAH and Ali JH, Acta Chimica Slovenica 2016, 63, 689–704. [PubMed: 28004090]
- (5). Conte MR, Jenkins TC and Lane AN, European Journal of Biochemistry 1995, 229, 433–444. [PubMed: 7744066]
- (6). Scott FJ, Nichol RJO, Khalaf AI, Giordani F, Gillingwater K, Ramu S, Elliott A, Zuegg J, Duffy P, Rosslee M-J, Hlaka L, Kumar S, Ozturk M, Brombacher F, Barrett M, Guler R and Suckling CJ, European Journal of Medicinal Chemistry 2017, 136, 561–572. [PubMed: 28544982]

- (7). Boykin DW, Journal of the Brazilian Chemical Society 2002, 13, 763–771.
- (8). Paul A, Nanjunda R, Kumar A, Laughlin S, Nhili R, Depauw S, Deuser SS, Chai Y, Chaudhary AS, David-Cordonnier M-H, Boykin DW and Wilson WD, Bioorganic & Medicinal Chemistry Letters 2015, 25, 4927–4932. [PubMed: 26051649]
- (9). Nanjunda R and Wilson WD, Current protocols in nucleic acid chemistry / edited by Beaucage Serge L. ... [et al.] 2012,
- (10). Miao Y, Lee MPH, Parkinson GN, Batista-Parra A, Ismail MA, Neidle S, Boykin DW and Wilson WD, Biochemistry 2005, 44, 14701–14708. [PubMed: 16274217]
- (11). Paul A, Kumar A, Nanjunda R, Farahat AA, Boykin DW and Wilson WD, Organic & Biomolecular Chemistry 2017, 15, 827–835. [PubMed: 27995240]
- (12). Guo P, Paul A, Kumar A, Farahat AA, Kumar D, Wang S, Boykin DW and Wilson WD, Chemistry – A European Journal 2016, 22, 15404–15412.
- (13). Chai Y, Paul A, Rettig M, Wilson WD and Boykin DW, The Journal of Organic Chemistry 2014, 79, 852–866. [PubMed: 24422528]
- (14). Harika NK, Paul A, Stroeve E, Chai Y, Boykin DW, Germann MW and Wilson WD, Nucleic Acids Research 2016, 44, 4519–4527. [PubMed: 27131382]
- (15). Paul A, Chai Y, Boykin DW and Wilson WD, Biochemistry 2015, 54, 577–587. [PubMed: 25495885]
- (16). Bostock-Smith CE, Laughton CA and Searle MS, Nucleic Acids Research 1998, 26, 1660–1667. [PubMed: 9512536]
- (17). Fede A, Labhardt A, Bannwarth W and Leupin W, Biochemistry 1991, 30, 11377–11388. [PubMed: 1720660]
- (18). Bostock-Smith CE, Embrey KJ and Searle MS, Chemical Communications 1997, 121.
- (19). Bostock-Smith CE, Harris SA, Laughton CA and Searle MS, Nucleic Acids Research 2001, 29, 693–702. [PubMed: 11160891]
- (20). Chen SM, Leupin W, Rance M and Chazin WJ, Biochemistry 1992, 31, 4406–4413. [PubMed: 1581295]
- (21). Jenkins TC, Lane AN, Neidle S and Brown DG, European Journal of Biochemistry 1993, 213, 1175–1184. [PubMed: 8504811]
- (22). Yoshida M, Banville DL and Shafer RH, Biochemistry 1990, 29, 6585–6592. [PubMed: 2397201]
- (23). Antony-Debre I, Paul A, Leite J, Mitchell K, Kim HM, Carvajal LA, Todorova TI, Huang K, Kumar A, Farahat AA, Bartholdy B, Narayanagari S-R, Chen J, Ambesi-Impiombato A, Ferrando AA, Mantzaris I, Gavathiotis E, Verma A, Will B, Boykin DW, Wilson WD, Poon GMK and Steidl U, Journal of Clinical Investigation 2017,
- (24). Wilson WD, Tanious FA, Mathis A, Tevis D, Hall JE and Boykin DW, Biochimie 2008, 90, 999–1014. [PubMed: 18343228]
- (25). Jenkins TC and Lane AN, Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1997, 1350, 189–204. [PubMed: 9048889]
- (26). Fede A, Billeter M, Leupin W and Wüthrich K, Structure 1993, 1, 177–186. [PubMed: 16100952]
- (27). Clark GR, Boykin DW, Czarny A and Neidle S, Nucleic Acids Research 1997, 25, 1510–1515. [PubMed: 9162901]
- (28). Acosta-Reyes FJ, Dardonville C, de Koning HP, Natto M, Subirana JA and Campos JL, Acta Crystallographica Section D: Biological Crystallography 2014, 70, 1614–1621. [PubMed: 24914972]
- (29). Wüthrich K, NMR of Proteins and Nucleic Acids, New York, 1986, pp. 203–233.
- (30). JNS. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, Oxford, 1995, pp 350.
- (31). Rettig M, Germann MW, Wang S and Wilson WD, ChemBioChem 2013, 14, 323–331. [PubMed: 23355266]
- (32). Cheatham TE and Young MA, Biopolymers 2000, 56, 232–256. [PubMed: 11754338]
- (33). Fratini AV, Kopka ML, Drew HR and Dickerson RE, Journal of Biological Chemistry 1982, 257, 14686–14707. [PubMed: 7174662]

- (34). Gorenstein DG, Phosphorus-31 NMR: Principles and Applications, Academic Press, New York, 1984.
- (35). Lefebvre A, Mauffret O, Lescot E, Hartmann B and Femandjian S, Biochemistry 1996, 35, 12560–12569. [PubMed: 8823193]
- (36). Wu Z, Tjandra N and Bax A, Journal of Biomolecular NMR, 2001, 19, 367–370. [PubMed: 11370783]
- (37). Xhani S, Esaki S, Huang K, Erlitzki N and Poon GMK, The Journal of Physical Chemistry B 2017, 121, 2748–2758. [PubMed: 28296403]
- (38). Liu Y, Kumar A, Depauw S, Nhili R, David-Cordonnier M-H, Lee MP, Ismail MA, Farahat AA, Say M, Chackal-Catoen S, Batista-Parra A, Neidle S, Boykin DW and Wilson WD, Journal of the American Chemical Society 2011, 133, 10171–10183. [PubMed: 21627167]

Figure 1.

A) The molecular structure of DB2277, with numbering scheme for aromatic protons. B) DNA hairpin sequence used in the study. The structure of DB2120 is shown in the inset.

Figure 2. Imino proton spectra of DB2277 binding with DNA hairpin at several DNA:DB2277 binding ratios at 285 K. Assignment of the imino proton resonances was done using 2D NOESY.

Figure 3.

NOESY spectrum of DB2277:DNA complex at 1:1 binding ratio at 285 K showing sequential walk through base proton to H1' sugar protons (solid lines). Intermolecular NOEs are marked in green boxes and assigned as follows; a) C18 (H1') - B5, b) C18 (H1') - B6, c) G5 (H1') - B2 and d) intramolecular DB2277 NOE, CH₂O - B2. Missing peak (*) is connected through broken line. Additional peaks are assigned further in Figure S11.

Figure 4. Chemical shift differences ($\Delta\delta$ in ppm) between free DNA and the DB2277-DNA complex for H1', H2' and H4' sugar protons, H2 protons of adenine and methyl protons of thymine.

Figure 5.

A) Specific interactions of G5 -NH₂ with aza BI -N and C18-O2 with aza BI -NH moiety of DB2277. Ring current effects on H4' protons of G5 bp in relation to phenyl rings of DB2277 (black circle). B) Ring current effects due to the mutual interaction of H1' protons of T7 and T20 with phenyl rings of DB2277 (black circle). DB2277 is in ball and stick (C in magenta) and DNA is shown in stick representation (C in green).

Figure 6. Schematic model of interactions of DB2277 protons with DNA protons at the binding site. Strong/Medium interactions are shown in solid lines whereas weak interactions are shown in broken lines. ● represents the -NH_2 group of amidine moieties. * NOE contacts observed in 90% $\text{H}_2\text{O}/10\%$ D_2O NOESY for exchangeable protons.

Figure 7. NOESY spectrum of the DB2277:DNA complex at 1:1 binding ratio at 285 K. Assignments of intramolecular DB2277 contacts are marked with green arrows whereas intermolecular NOEs of DB2277 with adenine of DNA are assigned and marked with red arrows. DB2277 protons are shown in green and H2 protons of adenine are shown in red.

Figure 8.

NMR structure ensemble of the DB2277-DNA complex. Snapshots were taken every 10 ps for the last 100 ps of rMD simulation and subsequently minimized. DNA is represented in new ribbon style (ochre) and DB2277 is shown in licorice (magenta). Hydrogen atoms, terminal bp and water molecules are omitted for the sake of clarity.

Figure 9.

The rMD model with water molecules: A) Minor groove view of the DB2277-DNA complex with proximal water molecules. Inset: B) Dynamic water network with amidine-NH and an interfacial water molecule that forms an H-bond bridge between amidine-NH and T20-O2 at the floor of the minor groove (Upper circle), C) The strong H-bond interaction of amidine-NH with T17-O2 along with an ensemble of dynamic H-bonds to water molecules in the minor groove of DNA. DNA backbone is shown in tube (tan) whereas sugar and base region are shown in stick representation. DB2277 is shown in ball and stick (yellow for H, magenta for C). Hydrogen atoms are omitted from DNA for clarity. H-bonds are reported in Å.

Table 1.

Intermolecular NOEs of DB2277 protons with DNA are listed.

DB2277 Proton	H2	H1'	H4'	H5'/H5''	Imino proton ^a
B1	A4, A3 ^b			G5	
B1'				A6 ^b	
B2	A4, A3 ^b	G5, T19 ^b		G5 ^b	
B2'				A6	
B3			T19 ^b		
B4			T19 ^b		
B5	A6	C18, A6 ^b		T7 ^b	G5
B5'			C18, T19 ^b		
B6	A6, A16	T17, C18			G5
B6'			C18	C18 ^b	
BI -NH					G5
OCH ₂		T19 ^b			

^a Intermolecular NOEs observed in NOESY spectrum collected in 90% H₂O/10% D₂O.

^b Weak NOEs or NOEs visible only at long mixing times.

Table 2.

NMR refinement statistics for free DNA and the DB2277-DNA complex. Structure statistics calculated from final converged 10 snapshots taken every 10 ps for last 100 ps of rMD simulation.

RESTRAINTS	FREE DNA	DB2277-DNA COMPLEX
Total NOEs	254	276
Intra-residue	175	179
Inter-residue (sequential)	78	64
DB2277 intramolecular	-	6
DB2277-DNA intermolecular	-	25
Backbone	88	88
Base pair	35	35
Sugar pucker	18	18
Structure statistics		
Total distance violation (kcal/mol)	21.47 ± 0.96	31.73 ± 2.35
Total bond length penalty (kcal/mol)	0.023	0.023
Total angle penalty (kcal/mol)	2.52 ± 0.06	2.53 ± 0.06

Experimental details are provided in supplementary materials and methods.