12-10-2018

Understanding the Recognition of Mixed Sequence DNA through Minor Groove Binding Compounds

Narinder Kaur Harika

UNDERSTANDING THE RECOGNITION OF MIXED SEQUENCE DNA THROUGH MINOR GROOVE BINDING COMPOUNDS

by

NARINDER KAUR HARIKA

Under the Direction of W. David Wilson, PhD

ABSTRACT

The broad range of diseases controlled by transcription factors (TFs) and their potential abilities to modulate gene expression have led to an emerging interest in the development of small molecules that target TF-DNA complexes. Still, there is only a limited number of types of designed small molecules that show strong and sequence-specific binding to DNA along with good cellular uptake properties for therapeutic use. Most of the successful nuclear stains or therapeutic agents that bind non-covalently in the minor groove of DNA are AT specific. Synthesis of novel compounds to better target the mixed AT/GC base pair (bp) sequences with a broad range of applications like targeting TF is a daunting task.

Our novel heterocyclic cation, DB2277, contains the aza-benzimidazole group (aza-BI) that specifically recognizes the single GC bp interspersed between AT bp sequences in the minor groove of DNA. NMR spectroscopy revealed the presence of major and minor binding species in the DB2277 complex with AAAGTTT type of DNA. NMR exchange dynamics have shown that the exchange between major and minor species is much faster than the compound's dissociation from the complex, as determined using surface plasmon resonance (SPR). To understand the molecular basis of recognition of mixed bp sequences and to acquire ideas to design new sequence-specific compounds, structural information of the DB2277-DNA complex is essential. Experimental structure of the unique and selective binding orientation of DB2277 with "AAGATA" binding site of DNA has been obtained using high-resolution NMR and molecular dynamics (MD) simulations, which suggests the involvement of two specific and strong H-bonds in recognition of the central GC bp. Extended MD calculations have shown dynamic water-mediated H-bond contacts between amidine of DB2277 and the bases at the floor of the minor groove and 180° rotations of the phenyl linked to a flexible linker (OCH_2) in a bound compound for the first time. Therefore, designing additional compounds with the ability to recognize a vast array of biologically important DNA sequences is essential for extending the use of new heterocyclic compounds in therapeutic applications in the future.

INDEX WORDS: DNA minor groove, Minor groove binder, Mixed base pair recognition, Nuclear Magnetic Resonance (NMR), Molecular Dynamics (MD) simulations, Solution Structure.

UNDERSTANDING THE RECOGNITION OF MIXED SEQUENCE DNA THROUGH MINOR GROOVE BINDING COMPOUNDS

by

NARINDER KAUR HARIKA

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

UNDERSTANDING THE RECOGNITION OF MIXED SEQUENCE DNA THROUGH MINOR GROOVE BINDING COMPOUNDS

by

NARINDER KAUR HARIKA

Committee Chair: W. David Wilson

Committee: Markus W. Germann

Gregory Poon

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2018

DEDICATION

To God

ੴ ਸਤਿ ਨਾਮੁ

For giving me the courage

&

To my beloved ਮਾਂ – ਬਾਪ (Mom & Dad)

For their unconditional love and support

***If I have seen further than others, it is by standing upon the shoulders of giants*.**

(By Isaac Newton)

ACKNOWLEDGEMENTS

Gratitude and Humility prevails when we embrace the fact that our lives have been buoyed by the generosity of others.

I owe my deepest gratitude towards my supervisor, Dr. W. David Wilson, who created this invaluable space for me to do my research, for which I will be truly indebted. I humbly appreciate his guidance, involvement, support, and constant encouragement which helped me learn a lot. I also express my warmest gratitude to my other supervisor Dr. Markus Germann, and it was an honor for me to learn NMR under his supervision. Dr. Wilson and Dr. Germann have made their support available to me in many ways, and I have gained invaluable knowledge and experience from all the discussions and advice from them throughout my research work.

I am grateful to Dr. David W. Boykin for providing all the compounds, without which this project could not be completed. I owe a great debt of gratitude to Dr. Gregory Poon for all the useful discussions and suggestion. His keen insight helped me to overcome many difficulties in the research work. I highly appreciate Mrs. Carol Wilson, who showed me love and support at every step of this journey. I am also thankful to Dr. Arvind Kumar and Dr. Jyotsna Thota for their generous advice and encouragement.

I sincerely thank our current Department Chair, Dr. Donald Hamelberg, for providing all the necessary facilities and guidance throughout my course. I also thank the technical and administrative staff of the Department of Chemistry.

I consider myself very fortunate to join this wonderful lab and colleagues who have become like family to me. I enjoyed working with Beibei Liu, Pu Guo, and Tam Vo in the lab. Thanks to Alex Spring and Marina Evich for their useful discussions and Ekaterina Stroeva for her help with the NMR instrumentation. I am indebted to many of my colleagues for their support in the past and at

present. I thank my colleague, Ananya Paul for helping me stay the course in both my personal and professional journeys.

I owe a great debt of gratitude to my siblings, Manpreet Kaur, Damanjeet Singh, and my brother-in-law Harbans Singh for their unfailing love and support which has always been a source of strength. I am thankful to my nephew and niece, Sehaj and Jap for cheering me up with their love through all phases of my life.

A special thanks to my very close friend, Zach Goldberg for his unconditional support and encouragement during my final years of the program. I can't thank enough to my favorite buddies Namrata, Khushi, Arpana, and Arya for filling my life with different colors and helping me to keep my spirits up.

I do not have enough words to express my gratitude towards my parents, Bhupinder Singh and Hardeep Kaur, and grandpa "*nana*", who stood by me through thick and thin. Their belief in me gave me the liberty to choose what I desired. I salute my mother for her selfless love, care, pain and sacrifice that profoundly shaped my life.

I would like to applaud all the people without whom this project would not have been completed and who supported me in large and small. At last, I would like to thank the Molecular Basis of Diseases and Georgia State University for funding my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Overview

DNA (Deoxyribonucleic Acid) was used as a target for the discovery of anticancer drugs even before the structure of DNA was discovered (1). For many years, scientists have shown empirically derived compounds with anticancer, antibacterial and anti-parasitic activities which either target DNA directly or inhibit the enzymes involved in gene regulation (2-4). DNA, as a central component of cellular machinery, is involved in a broad range of biological functions inside the cell. Therefore, such compounds with the ability to target DNA and their role in regulating the gene expression, have the potential for far-reaching applications in therapeutic development (5). DNA is a stable storage unit of all the genetic information and transfer of the genetic information in biological systems is explained through the central dogma (a term coined by Francis Crick in 1958) in molecular biology (6, 7). It describes the one-way flow of genetic information from DNA to RNA to proteins (**Figure 1.1**). In order to decode the DNA, the transcription process is used to copy DNA to information carrying RNA intermediary molecules which leads to the production of the protein molecules (8) (**Figure 1.1**). Transcription and translation processes serve as a major molecular "on/off" switch for the genetic expression (9).

1.2 Structural Characteristics of DNA

DNA, a secret of life that contains all the information required by any cell in the body to thrive, was first identified by Friedrich Miescher in the late 1860s. Watson and Crick concluded in 1953 that the DNA molecule exists in the form of a three-dimensional double helix (6). Erwin Chargaff explained the additional details of the structure of DNA (10). Two major conclusions found by Chargaff in 1950 are: 1) nucleotide composition of DNA varies among species. 2) All DNA follows the Chargaff's rule, in spite of its varied composition, the total amount of purines (A + G)

are essentially equal to the total amount of pyrimidines (C + T). Watson and Crick model represents the following features of DNA, also outlined in **Figure 1.2**: 1) DNA is a double-stranded helix, with the two strands connected by hydrogen bonds. Adenine base is always paired with thymine, and cytosine is always paired with guanine. 2) Most DNA double helices are right-handed. Only one known type of DNA, called Z-DNA, is left-handed. 3) The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand and vice versa (**Figure 1.2**). Nucleotides on the same strand are linked to each other by their phosphate groups, which bind the 3' end of one sugar to the 5' end of the next sugar. This antiparallel arrangement of the two backbone strands of DNA gives rise to another structural feature of the molecule i.e. the major and minor grooves, opposite to each other, also shown in the **Figure 1.2.** 4) The outer edges of the bases are exposed and available for potential hydrogen bonding as well. Through these H-bonds, other molecules including proteins can have access to DNA.

The minor groove of B-form DNA is narrow and shallow with the width of only about 10 Å whereas the major groove is deeper and wider, 24 Å width approximately. The major groove serves to provide more specific hydrogen bond donor and acceptor sites compared to the minor groove, therefore most of the DNA-interactive proteins bind in the major groove (3, 11). However, most of the small molecules less than 1000 Da with antibiotic and anticancer properties bind in the minor groove of DNA primarily.

1.3 Recognition of the minor groove of DNA by small molecules

Targeting the minor groove of nucleic acids through binding of small molecules helps in regulation of a wide variety of biological processes, therefore is an emerging field of interest. Typical small molecules that bind the DNA minor groove are heterocyclic dications and polyamides, for

example, netropsin, berenil, pentamidine and distamycin as shown in **Figure 1.3**. Natural Minor Groove Binders (MGB) include netropsin and distamycin. These compounds preferentially bind to the AT rich regions in the minor groove (12). Netropsin binds in the minor groove as a monomer whereas distamycin and its analogues, usually bind as a dimer in a widened minor groove, face to face and antiparallel (13-15). Synthetic compounds need proper curvature and positive charge to bind to the minor groove of DNA (16). In order to expand the sequence recognition capabilities and surpass the classical AT specific agents to include more mixed base pair (bp) sequences, re-designing of the small molecules is essential.

Designing of such sequence-specific small molecules may provide a powerful tool that can be directed towards production of valuable therapeutic agents. Some of these designed synthetic dicationic compounds are clinically used as therapeutic agents against a number of diseases (17-20). For example, pentamidine (bis-amidine) shown in **Figure 1.3** has been used for antiparasitic therapy since the 1940s (21, 22) but like many others among these compounds, it has several toxicity issues including provocation of hypoglycemia and renal and hepatotoxicity. The potent DNA binding capability of DAPI, a fluorescent compound, makes it an important stain in cellular microscopy and shows antitrypanosomal activity (23). Hoechst dyes which belong to the bis-benzimidazoles class of compounds are also strong minor groove binders with fluorescent properties depending upon the attached substituents (24). For example, Hoechst 33258 is very effective as a human topoisomerase I poison, yet it failed in phase II trials after passing human phase I clinical trials for pancreatic cancer (25, 26). Several analogues of Hoechst 33258 and a large number of hybrid molecules were synthesized based on netropsin and distamycin (**Figure 1.3**) with improved selectivity and specificity and with reduced side side-effects (27, 28). Study of the small molecule-DNA complexes also provides insight into the structural and thermodynamic

characteristics of the complexes (29, 30). Based on wide-ranging experience in design of minor-groove agents for AT recognition, an approach to rationally design the compounds which can bind strongly to mixed base pair DNA sequences is quite attractive (31-34).

Hydrogen bonding, electrostatic interactions along with the strong van der Waals interactions helps to stabilize small molecule-DNA complexes. Highly structured water molecules also play an important role in mediating the interactions of ligands in the minor groove of DNA without any loss in binding affinity (35-37). Targeting the nucleic acids through small molecules is still a challenge due to low specificity of DNA binders and high rate of their failure in clinical trials. Therefore, progression in the discovery of new small molecules requires further development of structural biology and other related technologies.

1.4 NMR Spectroscopy in small molecule-DNA complexes

To understand the biological importance of nucleic acid complexes, various biophysical, biochemical and spectroscopic methods have been used to provide atomic level descriptions of such recognition. NMR spectroscopy is a well proven method for elucidation of high-resolution structures and dynamics of the biomolecular complexes including many other systems which cannot be solved by X-ray diffraction studies (38, 39). In addition, NMR can also directly provide valuable experimental data on small molecule-DNA complexes, for example, quantitative and qualitative binding information can be obtained through changes in NMR parameters. Binding site of small molecule in the complex can also be obtained through change in chemical shift of the assigned NMR signals of the complex.

The development of [^{1}H, ^{1}H] nuclear Overhauser effect spectroscopy (NOESY) has made it possible for NMR spectroscopy to evolve as an alternative approach to X-ray crystallography in the determination of protein and nucleic acid conformation in solution (40). In NOESY

experiment, short (< 5 Å) inter-nuclear distances between hydrogen nuclei can be measured and directly used to probe the conformation of a molecule in solution. NMR techniques have been widely used to study the interactions of small molecules in the minor groove of DNA in solution (41). NMR has also been able to provide high-resolution structures of small molecule-DNA complexes in solution in conjunction with molecular dynamics simulations. In the early days, structures were solved for quinoxaline antibiotics nogalamycin, actinomycin, and the anthracyclines and they were in close agreement with the structures solved by X-ray diffraction (42-47). NMR has also provided a wealth of information about various minor groove binding compounds including oligopyrrole ligands such as netropsin, distamycin and bis-benzimidazoles related to Hoechst 33258, bisquaternary ligands and bisamidine ligands related to pentamidine (48-50). Moreover, NMR studies on these complexes have also revealed internal binding kinetics such as exchange rate calculations (51-53) and the dynamic information provided by NMR also helps to further clarify intermolecular interactions. Information obtained from NMR on small molecule-DNA complexes also depends upon the small molecule concentration in the solution, which is often limited by its solubility and aggregation, i.e. around 0.5-2.0 mM. Several computational methods and force fields have also been significantly improved which have helped in designing novel small molecules targeting nucleic acids.

1.5 Outline of Chapters and Goal of the Research

This dissertation is primarily focused on DNA and its complexes with small molecules. Much of the research work described in this dissertation elaborates the recognition of mixed bp sequences by novel heterocyclic diamidine molecules. The approach used here also provides substantial information on the highly efficient, sequence-specific, and dynamic nature of minor groove binding agents. The following three chapters are mainly focused on determining the structural and

dynamic aspects of DB2277, a novel heterocyclic diamidine, binding in the minor groove of DNA. Chapter 2 explores the binding affinity and dynamic aspects of DB2277 binding in the minor groove using over-thirty mixed base pair DNA sequences with the help of NMR, thermal melting and SPR techniques. Chapter 3 is focused on the determination of the first high-resolution NMR structure of the binding of heterocyclic cation compound, DB2277 that selectively recognizes mixed AT/GC bp DNA sequence in a 1:1 complex. Molecular features describing the recognition mechanism of both A/T bp and G/C bp are represented in this chapter in detail. Chapter 4 is aimed in looking at the phenyl ring rotation of the DB2277 molecule bound to the DNA by conducting multi-microsecond molecular dynamics simulations. This chapter also illustrates how a rotation of a phenyl group can occur in the bound state as well as how mixed base pair recognition can occur for monomer compounds in the bound state. The fundamental goal of all the research work shown here, is to expand the recognition capabilities of the newly designed sequence-specific heterocyclic compounds to recognize the vast array of biologically important and mixed bp DNA sequences. Modification in heterocyclic compounds that belong to the classical type minor groove binders is an approach to expand the applications of minor groove binders.

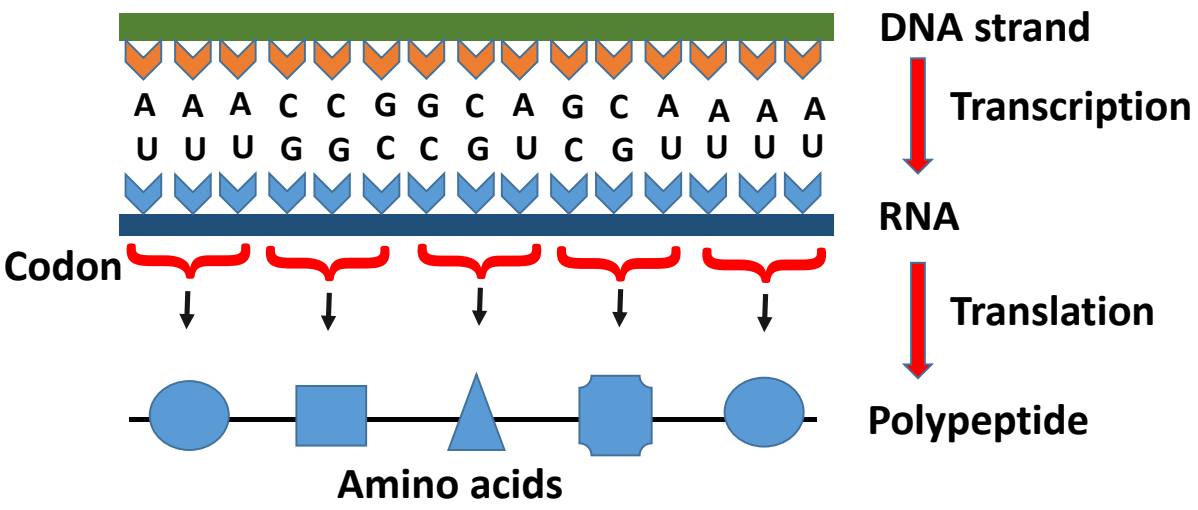


Figure 1.1 Central dogma.
Central dogma includes two processes, transcription and translation which provide the information about genetic flow in the cell.

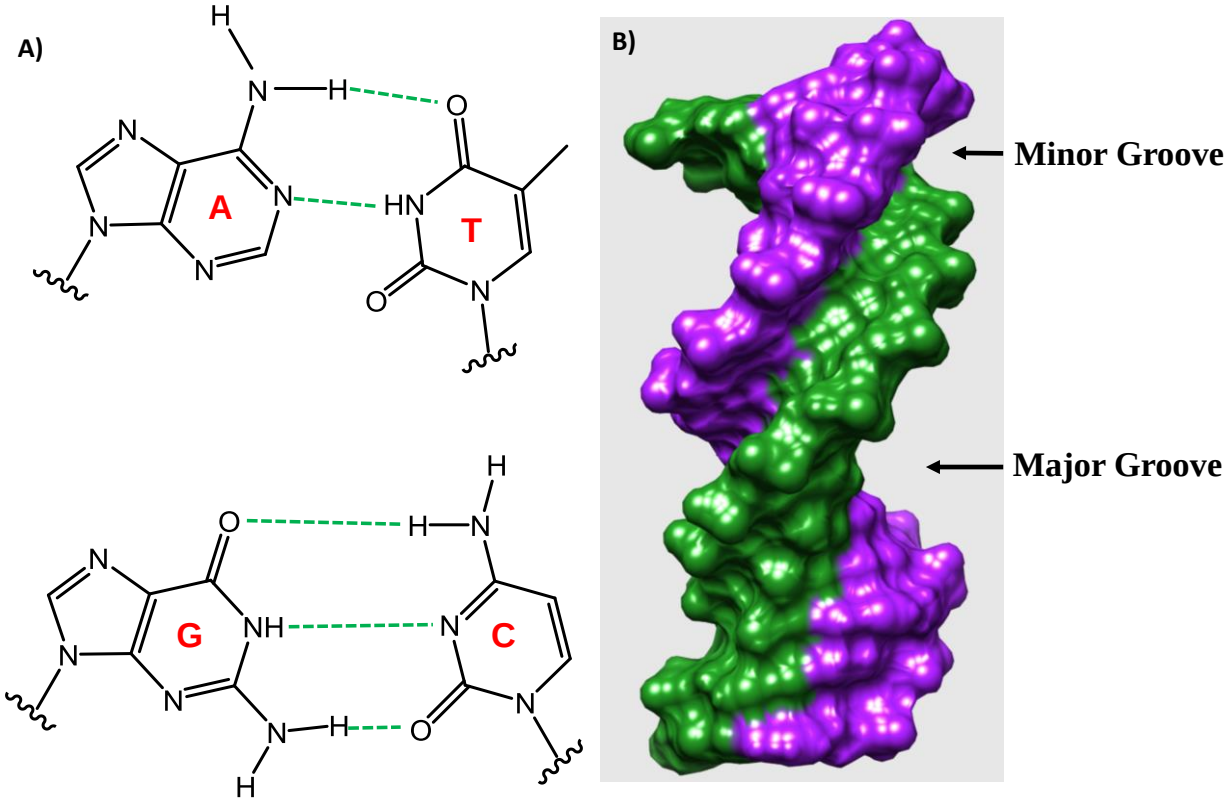


Figure 1.2 A) Base pairing in DNA. B) Major and minor grooves in DNA.
Two hydrogen bonds connect T to A; three hydrogen bonds connect G to C.

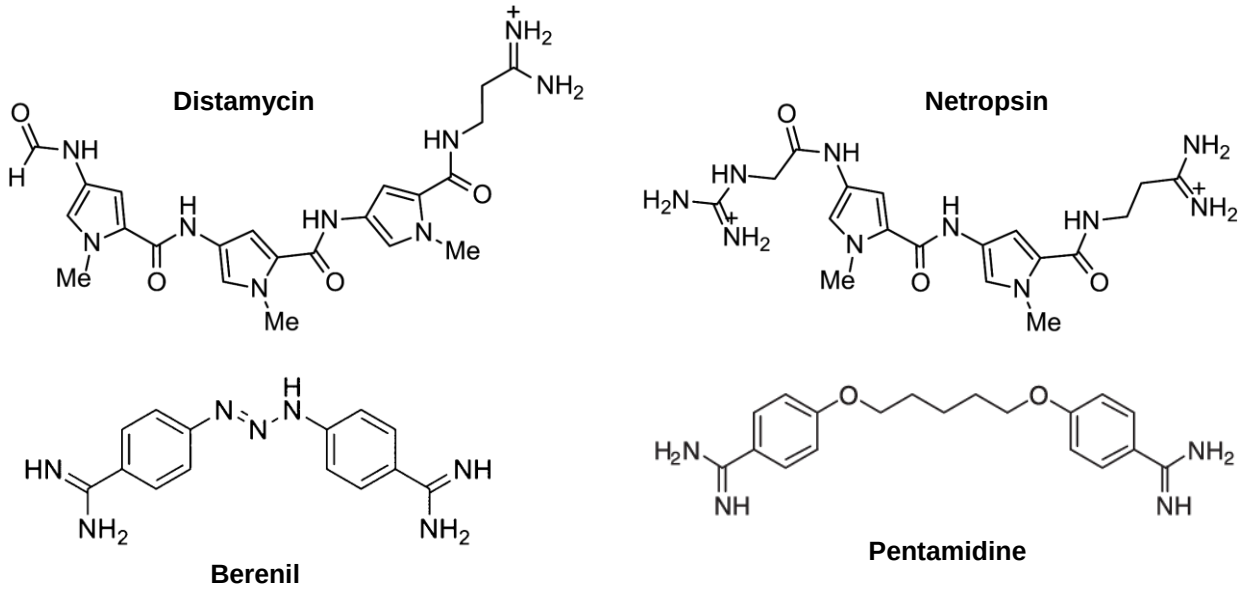


Figure 1.3 Typical DNA minor groove binding small molecules.

1.6 Reference:

1) Gurova, K. (2009) New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents, *Future oncol.* (London, England) 5, 1685.
2) Foye, William O. (1995) Cancer chemotherapeutic agents. American Chemical Society, Washington, DC.
3) Cai, X., Gray, P. J., and Von Hoff, D. D. (2009) DNA minor groove binders: Back in the groove, *Cancer Treat. Rev.* 35, 437-450.
4) Reinhold, W. C., Thomas, A., and Pommier, Y. (2017) DNA-Targeted Precision Medicine; Have we Been Caught Sleeping?, *Trends in cancer* 3, 2-6.
5) Koutsodontis, G., and Kardassis, D. (2004) Inhibition of p53-mediated transcriptional responses by mithramycin A, *Oncogene* 23, 9190.
6) Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature 171*, 737-738.
7) Crick, F. (1970) Central Dogma of Molecular Biology, Nature 227, 561.
8) Latchman, D. (2004) Eukaryotic Transcription Factors, Elsevier Academic Press, London, 4th edn, p. 1.
9) Mitchell, P. J., and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins, *Science* (New York, N.Y.) 245, 371-378.
10) Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation, *Experientia 6*, 201-209.
11) Neidle, S. (2001) DNA minor-groove recognition by small molecules, *Natural Product Reports 18*, 291-309.

12) Abu-Daya, A., Brown, P. M., and Fox, K. R. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands, *Nucleic acids research 23*, 3385-3392.
13) Lewis, E. A., Munde, M., Wang, S., Rettig, M., Le, V., Machha, V., and Wilson, W. D. (2011) Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site, *Nucleic acids research 39*, 9649-9658.
14) Pelton JG, Wemmer DE. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *P. Natl. Acad. Sci. USA* **1989**, 86: 5723-5727.
15) Wartell, R. M., Larson, J. E., and Wells, R. D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid, *The Journal of biological chemistry 249*, 6719-6731.
16) Hunt, R. A., Munde, M., Kumar, A., Ismail, M. A., Farahat, A. A., Arafa, R. K., Say, M., Batista-Parra, A., Tevis, D., Boykin, D. W., and Wilson, W. D. (2011) Induced topological changes in DNA complexes: influence of DNA sequences and small molecule structures, *Nucleic acids research 39*, 4265-4274.
17) Daliry, A., Pires, M. Q., Silva, C. F., Pacheco, R. S., Munde, M., Stephens, C. E., Kumar, A., Ismail, M. A., Liu, Z., Farahat, A. A., Akay, S., Som, P., Hu, Q., Boykin, D. W., Wilson, W. D., De Castro, S. L., and Soeiro, M. N. (2011) The trypanocidal activity of amidine compounds does not correlate with their binding affinity to Trypanosoma cruzi kinetoplast DNA, *Antimicrobial agents and chemotherapy 55*, 4765-4773.
18) Antony-Debre, I., Paul, A., Leite, J., Mitchell, K., Kim, H. M., Carvajal, L. A., Todorova, T. I., Huang, K., Kumar, A., Farahat, A. A., Bartholdy, B., Narayanagari, S. R., Chen, J., Ambesi-Impiombato, A., Ferrando, A. A., Mantzaris, I., Gavathiotis, E., Verma, A., Will, B., Boykin, D. W., Wilson, W. D., Poon, G. M., and Steidl, U. (2017) Pharmacological inhibition of the transcription factor PU.1 in leukemia, *J Clin Invest 127*, 4297-4313.
19) Soeiro, M. N., Werbovetz, K., Boykin, D. W., Wilson, W. D., Wang, M. Z., and Hemphill, A. (2013) Novel amidines and analogues as promising agents against intracellular parasites: a systematic review, *Parasitology 140*, 929-951.
20) Wilson, W. D., Tanious, F. A., Mathis, A., Tevis, D., Hall, J. E., and Boykin, D. W. (2008) Antiparasitic compounds that target DNA, *Biochimie 90*, 999-1014.
21) Stead, A. M., Bray, P. G., Edwards, I. G., DeKoning, H. P., Elford, B. C., Stocks, P. A., and Ward, S. A. (2001) Diamidine compounds: selective uptake and targeting in Plasmodium falciparum, *Molecular pharmacology 59*, 1298-1306.
22) Bray, P. G., Barrett, M. P., Ward, S. A., and de Koning, H. P. (2003) Pentamidine uptake and resistance in pathogenic protozoa: past, present and future, *Trends in parasitology 19*, 232-239.
23) Williamson, J., and McLaren, D. J. (1978) Effects of DAPI a new diamidine trypanocide, on the ultrastructure of Trypanosoma rhodesiense, *Transactions of the Royal Society of Tropical Medicine and Hygiene 72*, 660-661.

24) Umetskaia, V. N. (1992) [Photochemical transformation of Hoescht 33258 dye molecules complexed with cell nucleus DNA under the effect of UV-irradiation], *Biofizika 37*, 39-42.
25) Chen, A. Y., Yu, C., Bodley, A., Peng, L. F., and Liu, L. F. (1993) A New Mammalian DNA Topoisomerase I Poison Hoechst 33342: Cytotoxicity and Drug Resistance in Human Cell Cultures, *Cancer Res.* 53, 1332-1337.
26) Baraldi, P. G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M. A., Pavani, M. G., and Romagnoli, R. (2004) DNA minor groove binders as potential antitumor and antimicrobial agents, *Med. Res. Rev.* 24, 475-528.
27) Nelson, S. M., Ferguson, L. R., and Denny, W. A. (2007) Non-covalent ligand/DNA interactions: Minor groove binding agents, *Mutat. Res. /Fundamental and Molecular Mechanisms of Mutagenesis* 623, 24-40.
28) Khan, G. S., Shah, A., Zia ur, R., and Barker, D. (2012) Chemistry of DNA minor groove binding agents, *Journal of photochemistry and photobiology. B, Biology 115*, 105-118.
29) Rettig, M., Germann, M. W., Ismail, M. A., Batista-Parra, A., Munde, M., Boykin, D. W., and Wilson, W. D. (2012) Microscopic rearrangement of bound minor groove binders detected by NMR, *The journal of physical chemistry. B 116*, 5620-5627.
30) Paul, A., Chai, Y., Boykin, D. W., and Wilson, W. D. (2015) Understanding mixed sequence DNA recognition by novel designed compounds: the kinetic and thermodynamic behavior of azabenzimidazole diamidines, *Biochemistry 54*, 577-587.
31) Paul, A., Kumar, A., Nanjunda, R., Farahat, A. A., Boykin, D. W., and Wilson, W. D. (2017) Systematic synthetic and biophysical development of mixed sequence DNA binding agents, *Organic & biomolecular chemistry 15*, 827-835.
32) Munde, M., Kumar, A., Peixoto, P., Depauw, S., Ismail, M. A., Farahat, A. A., Paul, A., Say, M. V., David-Cordonnier, M. H., Boykin, D. W., and Wilson, W. D. (2014) The unusual monomer recognition of guanine-containing mixed sequence DNA by a dithiophene heterocyclic diamidine, *Biochemistry 53*, 1218-1227.
33) Chai, Y., Paul, A., Rettig, M., Wilson, W. D., and Boykin, D. W. (2014) Design and synthesis of heterocyclic cations for specific DNA recognition: from AT-rich to mixed-base-pair DNA sequences, *J Org Chem 79*, 852-866.
34) Nanjunda, R., and Wilson, W. D. (2012) Binding to the DNA minor groove by heterocyclic dications: from AT-specific monomers to GC recognition with dimers, *Current protocols in nucleic acid chemistry Chapter 8*, Unit8.8.
35) Liu, Y., Kumar, A., Depauw, S., Nhili, R., David-Cordonnier, M. H., Lee, M. P., Ismail, M. A., Farahat, A. A., Say, M., Chackal-Catoen, S., Batista-Parra, A., Neidle, S., Boykin, D. W., and Wilson, W. D. (2011) Water-mediated binding of agents that target the DNA minor groove, *J Am Chem Soc 133*, 10171-10183.
36) Athri, P., and Wilson, W. D. (2009) Molecular dynamics of water-mediated interactions of a linear benzimidazole-biphenyl diamidine with the DNA minor groove, *J Am Chem Soc 131*, 7618-7625.

37) Nguyen, B., Neidle, S., and Wilson, W. D. (2009) A role for water molecules in DNA-ligand minor groove recognition, *Acc Chem Res 42*, 11-21.
38) Rahman, A., and Choudhary, M. I. (2011) *Structure-activity Relationship Studies in Drug Development by NMR Spectroscopy: Volume 1*, Bentham Science Publishers.
39) Patel, D. J., Pardi, A., and Itakura, K. (1982) DNA conformation, dynamics, and interactions in solution, Science (New York, N.Y.) 216, 581-590.
40) Pellecchia, M. (2005) Solution Nuclear Magnetic Resonance Spectroscopy Techniques for Probing Intermolecular Interactions, Chemistry & Biology 12, 961-971.
41) Oliver, S. (2003) DNA and RNA Binders. From Small Molecules to Drugs. Vols. 1 & 2. Edited by Martine Demeunynck, Christian Bailly and W. David Wilson, *Angewandte Chemie International Edition 42*, 4994-4994.
42) Searle, M. S., Hall, J. G., Denny, W. A., and Wakelin, L. P. (1988) NMR studies of the interaction of the antibiotic nogalamycin with the hexadeoxyribonucleotide duplex d(5'-GCATGC)2, *Biochemistry 27*, 4340-4349.
43) Liu, X., Chen, H., and Patel, D. J. (1991) Solution structure of actinomycin-DNA complexes: drug intercalation at isolated G-C sites, *Journal of biomolecular NMR 1*, 323-347.
44) Odefey, C., Westendorf, J., Dieckmann, T., and Oschkinat, H. (1992) Two-dimensional nuclear magnetic resonance studies of an intercalation complex between the novel semisynthetic anthracycline 3'-deamino-3'-(2-methoxy-4-morpholinyl)-doxorubicin and the hexanucleotide duplex d(CGTACG), Chemico-biological interactions 85, 117-126.
45) Gao, X. L., and Patel, D. J. (1989) Antitumour drug-DNA interactions: NMR studies of echinomycin and chromomycin complexes, *Quarterly reviews of biophysics 22*, 93-138.
46) Chen, H., and Patel, D. J. (1995) Solution structure of a quinomycin bisintercalator-DNA complex, J Mol Biol 246, 164-179.
47) Gao, X. L., and Patel, D. J. (1989) Solution structure of the chromomycin-DNA complex, *Biochemistry 28*, 751-762.
48) Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin, *Proceedings of the National Academy of Sciences of the United States of America 82*, 1376-1380.
49) Pjura, P. E., Grzeskowiak, K., and Dickerson, R. E. (1987) Binding of Hoechst 33258 to the minor groove of B-DNA, *J Mol Biol 197*, 257-271.
50) Larsen, T. A., Goodsell, D. S., Cascio, D., Grzeskowiak, K., and Dickerson, R. E. (1989) The structure of DAPI bound to DNA, *Journal of biomolecular structure & dynamics 7*, 477-491.
51) Pardi, A., Morden, K. M., Patel, D. J., and Tinoco, I., Jr. (1983) Kinetics for exchange of the imino protons of the d(C-G-C-G-A-A-T-T-C-G-C-G) double helix in complexes with the antibiotics netropsin and/or actinomycin, *Biochemistry 22*, 1107-1113.

52) Patel, D. J. (1982) Antibiotic-DNA interactions: intermolecular nuclear Overhauser effects in the netropsin-d(C-G-C-G-A-A-T-T-C-G-C-G) complex in solution, *Proceedings of the National Academy of Sciences of the United States of America 79*, 6424-6428.
53) Leroy, J. L., Gao, X. L., Gueron, M., and Patel, D. J. (1991) Proton exchange and internal motions in two chromomycin dimer-DNA oligomer complexes, *Biochemistry 30*, 5653-5661.

2 IMINO PROTON NMR GUIDES THE REPROGRAMMING OF A•T SPECIFIC MINOR GROOVE BINDERS FOR MIXED BASE PAIR RECOGNITION

Narinder K. Harika, Ananya Paul, Ekaterina Stroeva, Yun Chai, David W. Boykin, Markus W. Germann, and W. David Wilson.

Harika, N. K., et al. *Nucleic acids research*. **2016** 44(10):4519-27.

My contribution to this chapter was writing, sample preparation, data acquisition and analysis.

2.1 Abstract

Sequence-specific binding to DNA is crucial for targeting transcription factor-DNA complexes to modulate gene expression. The heterocyclic diamidine, DB2277, specifically recognizes a single G•C base pair in the minor groove of mixed base pair sequences of the type AAAGTTT. NMR spectroscopy reveals the presence of major and minor species of the bound compound. To understand the principles that determine the binding affinity and orientation in mixed sequences of DNA, over thirty DNA hairpin substrates were examined by NMR and thermal melting. The

NMR exchange dynamics between major and minor species shows that the exchange is much faster than compound dissociation determined from biosensor-SPR. Extensive modifications of DNA sequences resulted in a unique DNA sequence with binding site AAGATA that binds DB2277 in a single orientation. A molecular docking result agrees with the model representing rapid flipping of DB2277 between major and minor species. Imino spectral analysis of a ^{15}N-labeled central G clearly shows the crucial role of the exocyclic amino group of G in sequence-specific recognition. Our results suggest that this approach can be expanded to additional modules for recognition of more sequence specific DNA complexes. This approach provides substantial information about the sequence specific, highly efficient, dynamic nature of minor groove binding agents.

2.2 Key words

Nuclear Magnetic Resonance; Minor groove binders; Minor groove recognition; Mixed DNA sequences, Selectivity

2.3 Abbreviations

NMR, nuclear magnetic resonance; bp, base pair; hp, hairpin; SPR, surface plasmon resonance; RU, response unit from SPR; Tm, thermal melting; EXSY, exchange spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; FPLC, fast protein liquid chromatography; B3LYP, Becke three-parameter hybrid functional combined with Lee-Yang-Parr correlation functional; DFT, density functional theory.

2.4 Introduction

Targeting transcription factor-DNA complexes for regulation of gene expression, is very attractive for the treatment of a number of diseases (1-3). This could be done by targeting the transcription factor (TF) or the DNA binding site (4). It has proven difficult to target TFs directly and they are

frequently considered “undruggable” (5). The alternative of targeting the DNA binding site of the TF with designed small molecules is, thus, gaining increasing attention (6-8). A barrier to this approach for control of gene expression in cells is that only a limited variety of small molecules bind strongly and sequence-specifically to DNA (9,10). The modification of heterocycles in classical type minor groove binders is an approach to expand the applications of minor groove binders (11,12).

The greatest variety of A•T specific minor groove binding compounds such as netropsin, DAPI, pentamidine, furamidine and Hoechst 33258 and its analogs are a complement to the highly sequence-specific hairpin polyamides (13). Many of the A•T specific compounds are taken up by cells and have a unique variety of therapeutic as well as biotechnology applications (14,15). Non-covalent binding minor groove agents have been used in therapeutic targeting of DNA in various types of cells with successful application all the way to clinical trials (16,17). As described above, many minor groove binding compounds are highly A•T specific (18) but it is essential to be able to better target mixed, A•T and G•C base pair (bp) sequences (19,20). Such novel compounds would be useful for a new class of potential drug molecules that can target promoter DNA sequences to modulate transcription factor binding (21-23). In order to expand past the pure A•T specific compounds, new types of compounds must be designed (24-26). These modified compounds still require the correct curvature and charge to match the minor groove but they will also need functional groups that have H-bond acceptors for favorable interactions with the G-NH_2 group that projects into the minor groove and presents a steric block to strong binding by A•T specific minor groove binders (27,28).

Our understanding of G•C bp recognition by small molecule minor groove binders is very limited, and for use of the DNA minor groove for new drug design and development we need a

more detailed understanding of compound-DNA complexes (29-31). Compounds with aza-benzimidazole H-bond acceptors have initially been designed for G•C recognition (32). DB2277 (**Figure 2.1**) is a novel example with strong binding to an AAAGTTT binding site (32,33). Evaluation of an analog of DB2277 without the CH_2O linker, for example, indicated the importance of flexibility in strong, sequence-dependent binding at mixed DNA sequence (32). To fully appreciate the binding of DB2277 to its target DNA sequence, structural information is required. NMR, in conjunction with molecular modeling, is uniquely suited to study the molecular interactions and dynamic behavior of DNA-ligand complexes (34-38). An initial NMR experiment for a DB2277 complex with an AAAGTTT sequence, however, revealed the surprising result that two species are present, even in a 1:1 complex. Interestingly, these two species were not observed in other high-resolution biophysical experiments (32,33). To understand these multiple binding species and their importance for compound design, we have investigated the DB2277 complex with numerous hairpin DNA duplexes.

2.5 Results

2.5.1 *Proton NMR Spectra*

The chemical shifts of DNA protons of the A•T and G•C bp have been used to monitor the binding of DB2277 with DNA. Numerous changes in the chemical shifts for DNA imino protons have been observed upon complexation with DB2277, particularly in the central region of the hairpin DNAs. We have investigated over thirty DNA sequences and results for six G-hp DNAs that

illustrate different modes of sequence-specific binding of DB2277 to DNA are shown here (**Table 2.1**).

2.5.2 *G-hp1 (AAAGTTT Binding Site)*

The G-hp1 DNA sequence gives well-resolved spectra with six A•T and four G•C bp signals and a broad upfield band for the unpaired T loop bases (**Figure 2.1**). Assignment of the imino proton signals of free DNA shown in the **Figure 2.1** was done using 1D-NOE experiments. Titration of G-hp1 with DB2277 revealed both upfield and downfield chemical shift changes. The addition of DB2277 to G-hp1 at the ratio of 1:0.5 (DNA:DB2277) shows both free DNA and complex resonance peaks in the spectra indicating slow exchange on the NMR timescale. No significant chemical shift changes were observed for DNA imino peaks of the terminal bp of the duplex indicating their minimal contribution to binding with DB2277. The appearance of an unexpected additional peak at 11.3 ppm was observed with the addition of DB2277. This additional peak was not observed in 1D NMR spectrum of free DB2277 and the assignment of this peak which belongs to the bound DB2277 is discussed in detail below. No significant changes were observed after reaching a 1:1 ratio of DNA:DB2277. At ratios above 1:1, slightly broadened peaks are probably due to some non-specific, electrostatic interaction of the excess DB2277 cation with the anionic DNA backbone. No unbound G-hp1 peaks are observed at the binding ratio of 1:1 (DNA:DB2277) suggesting a tight binding of DB2277 with the G-hp1, in agreement with SPR results (33).

For DB2277 binding with G-hp1 as a single species, six imino proton resonance peaks are expected for A•T bp at a 1:1 binding ratio. The imino proton spectra (**Figure 2.1**), however, reveal that all complex signals are doubled, even though the stoichiometry of binding is 1:1. The two sets of signals clearly indicate two binding modes for DB2277 with G-hp1, a major and a minor binding species based on signal intensity (39,40). Interestingly, other biophysical methods such as surface

plasmon resonance, mass spectrometry, circular dichroism and isothermal titration calorimetry also indicated a 1:1 complex between DB2277 and G-hp1 (33) but these methods were unable to show any evidence of two binding orientations of DB2277. The local symmetry (similar A•T bp sequences on both sides of the central G•C bp) of G-hp1 may result in the two orientations of DB2277 with similar energy. The two modes suggest complexes with reverse orientations in the AAAGTTT sequence (**Figure 2.1**) as previously observed for the polyamide minor groove binder, netropsin, in all A•T sequences (40).

In order to understand the origin of the two bound species and to facilitate study of structural details of the DB2277-DNA complex a more detailed analysis of DB2277 binding to different DNA sequences was undertaken. It is important to understand the origin of the different compound orientations and how they can be modulated by DNA sequences. It seemed possible that formation of a unique complex could be achieved with DNA sequences that are less symmetric than the AAAGTTT motif in G-hp1.

2.5.3 *G-hp2 (AAGTTT Binding Site)*

A 5' A•T bp in G-hp1 (G-hp2) was removed to give a less symmetric binding site. The question to be addressed is whether changes of this type will give a single DB2277-DNA species with good binding affinity. Imino proton spectra of G-hp2 show one less A•T bp resonance peak (**Figure 2.2**). At a 1:1 binding ratio of DNA:DB2277, the fraction of the minor species was reduced, but not eliminated relative to the G-hp1-DB2277 complex. The largest differences in chemical shifts between the major and minor species are observed for A4 and T6 imino protons which are adjacent to the central G. The chemical shift of T6 signal for the major species is downfield of A4 signal. It is very interesting that the minor species T6 signal is the most upfield while the minor species A4 is very downfield. In other words, the protons shift position is in an opposite way for these

major and minor species, in agreement with an opposite orientation of the DB2277 for two species evident in the flipping exchange model (**Figure 2.1**). In addition, the chemical shifts for the major binding species are similar to the major species in G-hp1 complex but with one less A•T bp signal. The less symmetric binding site reduces the amount of minor species. This result indicates that the less symmetric binding site in G-Hp2 relative to G-Hp1 is an advantage for obtaining a single species. While this may seem obvious for asymmetric sites, not all asymmetric sequences are advantageous for yielding a single binding species.

Assignment of exchangeable protons of the G-hp2-DB2277 complex was done using 1D-NOE and 2D NOESY experiments (**Figures 2.3, 2.4, and 2.5**). The ambiguity with the imino proton assignment due to exchange peaks was resolved by performing 2D EXSY experiments (**Figure 2.6, 2.7 and 2.8**). 2D EXSY experiments are similar to very short mixing time 2D NOESY. They are useful to define which 2D cross peaks are due to exchange. Intense exchange cross peaks in the EXSY experiment showing exchange between major and minor species and NOE-build up from adjacent bp in NOESY experiment are clearly discernible in **Figure 2.6**.

2.5.4 *G-hp3 (AAAGTT Binding Site)*

Since the removal of a 5' A•T bp in G-hp1 improved the imino proton spectra, in the next step a 3' terminal T•A bp was deleted from G-hp1 to remove sequence similarity in a different way. Despite similar imino proton spectra of free G-hp2 and G-hp3 DNA, their titration spectra with DB2277 were quite different (**Figures 2.2 and 2.9** respectively). Surprisingly, the change at the 3' end of the binding site in G-hp3 resulted in sharp peaks with a significant increase in signal intensity of the minor species as compared to the G-hp2 complex. Clearly, the 3' TTT side of the binding site is critical for the major species while the minor species is more sensitive to the 5' AAA sequence. These observations show that the amount of the minor species is very sequence-

dependent, even for two asymmetric sequences, AAGTTT and AAAGTT. Thus, the binding of DB2277 is quite sensitive to the A•T bp sequence flanking the central G•C bp.

2.5.5 *G-hp4 (AACTTT Binding Site)*

A different possible method to improve the results for G-hp2 is to invert the central 5' G•C bp, again with the goal of reducing the minor species. It is surprising that this change in the central G•C bp causes the minor species to significantly increase relative to the G-hp2 complex (**Figure 2.10**) These results indicate that the DB2277 complex is also very sensitive to the orientation of the central G•C bp. The NMR results with G-hp1, 2, 3, and 4 with DB2277 illustrate the extreme sensitivity of the complex to the DNA sequence.

2.5.6 *G-hp5 (AAGATA Binding Site)*

The 3' TTT sequence clearly has a major role in directional binding with DB2277 since the removal of the T in G-hp3 causes large changes in the spectra. In G-hp5, the 3' TTT was changed to ATA and this change yielded a single set of peaks for the complex at a 1:1 ratio. Examination of several 3' modifications that reduce the local symmetry in the central region in different ways (data not shown) helped to pave the way towards a 3' ATA change. At the binding ratio of 1:0.5 (DNA:DB2277), double imino resonance peaks for the five A•T bp imino protons reveal the occurrence of one binding species and the free DNA. The almost equal intensity of these binding species shows two equally populated states. Single bp resonance peaks in the imino spectral region in the G-hp5-DB2277 complex at the 1:1 binding ratio show one selective binding orientation (**Figure 2.11**). The imino proton assignment of the G-hp5-DB2277 complex was done using a 2D NOESY experiment (**Figure 2.12**). The G-hp5-DB2277 complex at the ratio of 1:1 provides selective binding as desired. The reduction of minor species in G-hp2 and single binding species in G-hp5 is clearly seen in the spectral comparisons in **Figure 2.13**. Decrease in intensity of the

peak at 14.5 ppm can be used as a probe to observe the reduction of the minor species from the G-hp1 complex to G-hp5 complex.

2.5.7 *G-hp6 (AACATA Binding Site)*

As observed with G-hp4, inversion of the central G•C in G-hp5 again results in the appearance of significant minor species signals in the spectra of DB2277-DNA complex (**Figure 2.14**). Therefore, the AAGATA binding site (G-hp5) sequence is the only one among over thirty oligomer hairpin duplexes investigated that provided very selective binding with DB2277. The success with G-hp5 relative to G-hp1 is due to two sequence changes. Removal of a 5' A base to give a less symmetric binding sequence clearly reduces the minor species. The reason for the decrease of the minor form when the 3' TTT is changed to ATA is less clear. Subtle changes in DNA microstructure, base substituent position and fit of DB2277 to the minor groove must be responsible for an enhancement of the major species. The DB2277 complex with the AAGATA sequence is now being used for high resolution NMR analysis of the structural details of the selective DB2277 complex.

2.5.8 *^{15}N-labeled central G in G-hp5*

To assign the peak at 11.5 ppm, a G-hp5 sequence was synthesized having the central G fully labeled with ^{15}N. The proton spectra of ^{15}N-labeled G-hp5 show five A•T bp imino proton peaks, as with the G-hp5 spectra (**Figure 2.15**). Splitting of the G5 imino proton resonance at 12.5 ppm is observed in the ^{15}N-labeled spectra due to ^{15}N-1H coupling (88 Hz) whereas a singlet is seen for G5 in the unlabeled DNA. Splitting of the G5 proton peak reverts to a singlet upon decoupling of ^{15}N (**Figure 2.16**).

On addition of DB2277, the additional peak emerges at 11.5 ppm as observed in all other G-hp-DNA complexes. However, this peak shows no effect of ^{15}N-labeling, which indicates that

this peak does not belong to G5. The aza-benzimidazole NH signal of DB2277 is generally downfield and exchanges with water. A strong H-bond of this proton to DNA should cause an additional downfield shift of the NH signal (41) and this signal should not be affected by ^{15}N coupling. Moreover, in the molecular model of the G-hp1-DB2277 complex (**Figure 2.17**), the aza-benzimidazole NH of DB2277 is observed to form a strong H-bond with the carbonyl oxygen of an adjacent T base. Therefore, in this case, the signal at 11.5 ppm, not affected by ^{15}N coupling, most likely represents the aza-benzimidazole NH of DB2277. Support for this assignment is also provided by NMR studies done by Leupin and coworkers (41) on Hoechst 33258 with A•T sequences. In this model, the benzimidazole NH proton faces into the minor groove and forms an H-bond with the carbonyl oxygen of a T base. The chemical shifts for the labile protons of benzimidazole NH are 11.3 and 12.3 ppm in the AATT complex (41), in general agreement with an observed signal between 11 and 11.5 ppm in **Figures 2.13 and 2.15**.

2.5.9 *UV-Vis Thermal Melting*

Thermal melting experiments were performed on over thirty G-hp DNA sequences by monitoring DNA unfolding at 260 nm to screen the complexes based on their relative binding affinity with DB2277. The addition of DB2277 to the hairpin oligomers in this study resulted in Tm increases (ΔTm) of 10-14 °C. The DB2277 strongly stabilizes the G-hp1-5 as clearly indicated by their ΔTm values (**Table 2.3**) whereas, reversing the central G•C bp in G-hp6 decreases the thermal stability of the complex (ΔTm value is 8.7 °C).

2.5.10 *Surface Plasmon Resonance*

Biosensor-SPR methods provide a quantitative estimation of the interaction of small molecules with DNA and other biomolecules (42). Very fast association and slow dissociation rates make DB2277 a very strong binder for single G•C-containing mixed sequences. SPR data (**Figure 2.18**)

reveals that DB2277 binds strongly with the single G•C bp sequence in both AAGTTT (KD = 1.3 nM) and AAGATA (KD = 4.4 nM). This represents a small difference in ΔG° for binding to these two oligomers (Δ(Δ(reveals that DB2277 binds strongly with the single number of factors. Certainly, the increased number of species for G-hp2 contributes to the more favorable ΔG° for binding. DB2277 forms a 5-fold weaker monomeric complex with AAGATA which has DB2277 bound in one orientation.

2.5.11 *Dynamics of Bound DB2277*

At a 1:1 binding ratio, only bound DNA is observed with separate peaks for the major and minor species. 2D EXSY provided clear evidence for the exchange between major and minor species of bound DB2277 and allowed evaluation of the exchange rate between both species. Strong cross peaks observed in EXSY spectra of the G-hp2-DB2277 complex shows the chemical exchange between species (**Figures 2.6, 2.7 and 2.8**). Exchange cross peaks observed in this short time experiment are not due to NOE from the adjacent bp, but due to the chemical exchange for nuclei in different environments. The exchange rate between the major and minor species is 6.8 s^{-1} at 303 K and 2.8 s^{-1} at 285 K. The exchange rate was calculated by considering the chemical exchange between major and minor species for A•T bp imino protons (A4, T6 and T7 at 303 K and A4 and T8 at 285 K) which show sufficiently resolved signals (**Table 2.4 and 2.5**) (43,44). The exchange rate calculated from the EXSY spectrum at 50 ms mixing time, is in close agreement with the exchange rate calculated at 150 ms mixing time.

2D EXSY also provided information about the exchange of the aza-benzimidazole NH proton of DB2277. The aza-benzimidazole NH interacts with G-hp DNA in two different chemical environments as reported previously (41). A clearly visible exchange cross peak has been observed for the labile NH proton of aza-benzimidazole of DB2277 in the EXSY spectrum (**Figure 2.7**). In

addition, 1D-NOE differences in the spectra clearly show the change in signal intensity at the central G and adjacent T base on irradiating the NH proton signal of the aza-benzimidazole moiety of DB2277 resonating at 11.25 ppm (**Figure 2.2**). The exchangeable NH proton of aza-benzimidazole in different environments characterizes the major and minor binding species and is at 11.3 ppm and at 12.7 ppm in complex spectra, respectively (**Figure 2.2**).

An interesting comparison between macroscopic dissociation rate constants from SPR data with the total exchange rate constant from 2D EXSY for DB2277 binding with G-hp1 is shown in **Table 2.2**. The apparent half-life for the exchange between different species is around 100 ms by NMR whereas the half-life for the bound DB2277 with G-hp2 and G-hp5 is in the range of 30-60 s by SPR. Thus, the exchange between the major and minor species is much faster than the macroscopic dissociation of DB2277 from the complex. The long dissociation half-life of the complex compared to half-life of the exchange between species shows that DB2277 can flip rapidly between major and minor species while remaining bound to minor groove of DNA. Such microscopic exchange between two orientations of bound DB2277 is clearly represented in the model shown below. In addition, the free energy change for exchange of major to minor species can be calculated using intensities of resonances originating from the different species ($\Delta G^\circ = -RT \ln [(I_M)/(I_m)]$ where $[(I_M)/(I_m)] = K_{eq}$) . The free energy difference between major and minor species is 0.95 ± 0.03 Kcal/mol at 303 K and 0.97 ± 0.03 Kcal/mol at 285 K (**Table 2.5**). The configurational entropy change associated with different orientations of the major and minor species is 0.4 kcal/mol·K using Boltzmann's entropy formula, $S = k_B \ln (W)$ where k_B is the Boltzmann's constant and W is the number of possible configurations. Both dynamic amidine moieties and the $–CH_2O$ linker (as shown in the model in **Figure 2.17**) provide flexibility in the minor groove of G-hp2. This flexibility contributes to the ability of DB2277 to adopt a favorable

conformation and make optimum contacts with the minor groove structure. Removal of the CH_2O linker, for example, results in a loss of both selectivity and affinity (32). From the clustering histogram of 200 Autodock runs (**Figure 2.19**), it has also been observed that the conformer in **Figure 2.17 (A)** is the most populated and most favorable with the lowest binding energy.

The rapid, microscopic dynamics of DB2277 and netropsin (40) suggest that such dynamics may be a common feature of minor groove binding compounds, even those with quite low K_D values. The flipping motion of DB2277 in different orientations is seen in the model in **Figure 2.17** and is surprising for a compound bound to a DNA sequence.

2.5.12 *Molecular Modeling of G-hp1-DB2277 Complex*

DB2277 fits quite well into the minor groove of the AAAGTTT binding site with an optimum curvature to follow the groove shape. Models of the two possible orientations (**Figure 2.1**) represented in **Figure 2.17**, are in general agreement with the experimental results and predict the observed strong binding of DB2277 into this sequence. The flexible -CH_2O- linker is essential for the aza-benzimidazole moiety to track along the minor groove curvature and predicts a strong H-bond with the exocyclic G-NH_2 that projects into the minor groove of the binding site. The H-bond distance of aza-N with G-NH_2 is 1.9 Å and 2.2 Å in conformers represented in **Figure 2.17 (A) and (B)**, respectively. The NH of the aza-benzimidazole moiety also nicely projects towards the minor groove of the binding site and can form H-bonds with the carbonyl group of 5' T7 on the same strand and 3' T5 on the complementary strand as shown in **Figure 2.17 (A) and (B)** respectively (T=O---H-N distances are 2.0 Å and 1.9 Å, respectively). By van der Waals interactions, the two phenyl groups adequately stack parallel to both sides of the minor groove walls and provide an extended binding interaction with the flanking bases. Both amidine groups fit into the minor groove and dynamically form strong H-bonds with both N3 of A and O2 of T at

the flanking side of mixed DNA sequences (**Figure 2.17**). The inherent dynamic nature also allows the amidines to form H-bonds to deoxyribose O and the negatively charged phosphate backbone. The -CH_2O- linker of DB2277 molecules in this model is serving primarily to add needed flexibility and appropriate distances to the groups of DB2277 to allow optimum DNA interactions.

2.6 Discussion

An imino proton NMR investigation on different single G•C bp-containing DNA sequences with an aza-benzimidazole compound, DB2277, revealed that rationally designed diamidine molecules can act as very sequence-specific probes at the minor groove of mixed sequences of DNA. The strong binding efficiency and highly dynamic binding nature of DB2277-G-hp complexes allow the ligand to form two reversed oriented microstructures in the minor groove of G-hp sequences. A key finding here is that the large dissociation off-rate of G-hp2 complex allows the DB2277 to microscopically rearrange in the complex without its complete dissociation from target DNA. The study also illustrates the different orientation preferences of DB2277 in G-hp DNA due to different microstructures and base substituent positions and how well the nature of the minor groove can mesh with the functional groups of the DB2277. Another major finding of a DNA sequence with a single species orientation for DB2277 with good binding affinity is essential for detailed NMR structural studies of the DB2277-DNA complex. The examination of binding spectra of DB2277 with ^{15}N-labeled G5 clearly shows the pivotal role of the exocyclic G-NH_2 in H-bonding with the aza-benzimidazole moiety of DB2277. This research has provided exciting information about sequence specificity and exchange behavior of DNA minor groove binders which will help expand the scope of rationally designed molecules for DNA minor groove recognition.

2.7 Experimental protocol

2.7.1 *Sample Preparation*

Sequences for selected hairpin DNAs are provided in **Table 2.1** (Integrated DNA Technologies, Coralville, IA, USA). The heterocyclic diamidine, DB2277, was synthesized as previously described (32). ^{15}N-labeled central G in G-hp5 was synthesized with ^{15}N-labeled phosphoramidites using a 391 DNA synthesizer and then desalted by FPLC using an AKTA purifier. The purity of DNA was checked using denaturing gel electrophoresis. The DB2277 was prepared as 2 mM stock solution in double distilled water (ddH_2O) and its concentration was calibrated against a known concentration of DSS (4, 4-dimethyl-4-silapentane-1-sulfonic acid), being used as an internal standard.

2.7.2 *NMR Experiments*

1D and 2D NOE data were collected. Water suppression in NMR spectra was obtained using 1-1 pulse sequence. In 1D-NOE difference experiment, selective imino proton resonances are pre-irradiated to observe the NOE build up at proton resonances in proximity. Phase-sensitive NOESY spectra were recorded with 2048 × 400 data points in the two dimensions at different mixing times (150 ms and 50 ms). The spectral width was set at 24 ppm and 1 s relaxation delay was used for experiments in H2O. Both dimensions were acquired with cosine bell functions (SSB = 2). From 2D Exchange Spectroscopy (2D EXSY), total chemical exchange rate is calculated for a two-site exchange using the following equations (**Table 2.2**):

$$k = (1 / tm) \ln ((r + 1) / (r - 1))$$

$$r = 4 X_M X_m ((I_M + I_m) / (I_{M-m} + I_{m-M})) - (X_M - X_m)^2$$

The term r corresponds to unequal populations of the major and minor species. I_M and I_m are the diagonal peak intensities of major and minor exchangeable resonances in the EXSY spectrum and

I_M-m and I_m-$_M$ are the intensities of the exchange cross peaks, tm is the mixing time. X_M and X_m are the mole fractions of the major and minor species (43, 50).

2.7.3 *Surface Plasmon Resonance*

As the Biosensor-SPR approach responds to mass, it is an excellent method for comparative studies of dications that have very large differences in properties and equilibrium binding constants (K). The SPR experiments were performed at 25 °C in degassed and filtered Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl. 1 mM EDTA, pH 7.4). Steady-state binding analysis was performed with multiple injections of compound concentration as previously described (33). Kinetic analysis was performed by globally fitting the binding results for the entire concentration series using a standard 1:1 kinetic model with integrated mass transport-limited binding parameters as described previously (42, 45). The signal at saturation of binding sites allows direct determination of stoichiometry.

2.7.4 *Molecular Modeling and Docking*

The center of the macromolecule is the grid center with a grid size of 72 Å×68 Å×108 Å and a grid spacing of 0.375 Å. Docking runs were performed using the Lamarckian genetic algorithm (LGA) with no modifications of docking parameters. LGA was used because of the existence of rotatable bonds in the ligands and to evaluate the correct conjugate DNA conformation, as it is known to reproduce various experimental ligand−DNA complex structures. Initially, we used a population of random individuals (population size of 150), a maximal number of 2500000 energy evaluations, a maximal number of evaluations of 2700, and a mutation rate of 0.02 fs. Two hundred independent flexible docking runs were conducted for each ligand, and then the lowest-energy dock conformation obtained from the Flexible docking was resubmitted for rigid docking to

remove the internal energy of the ligand (steric clashes) and retain the hydrogen bonding interaction with ds-DNA bases.

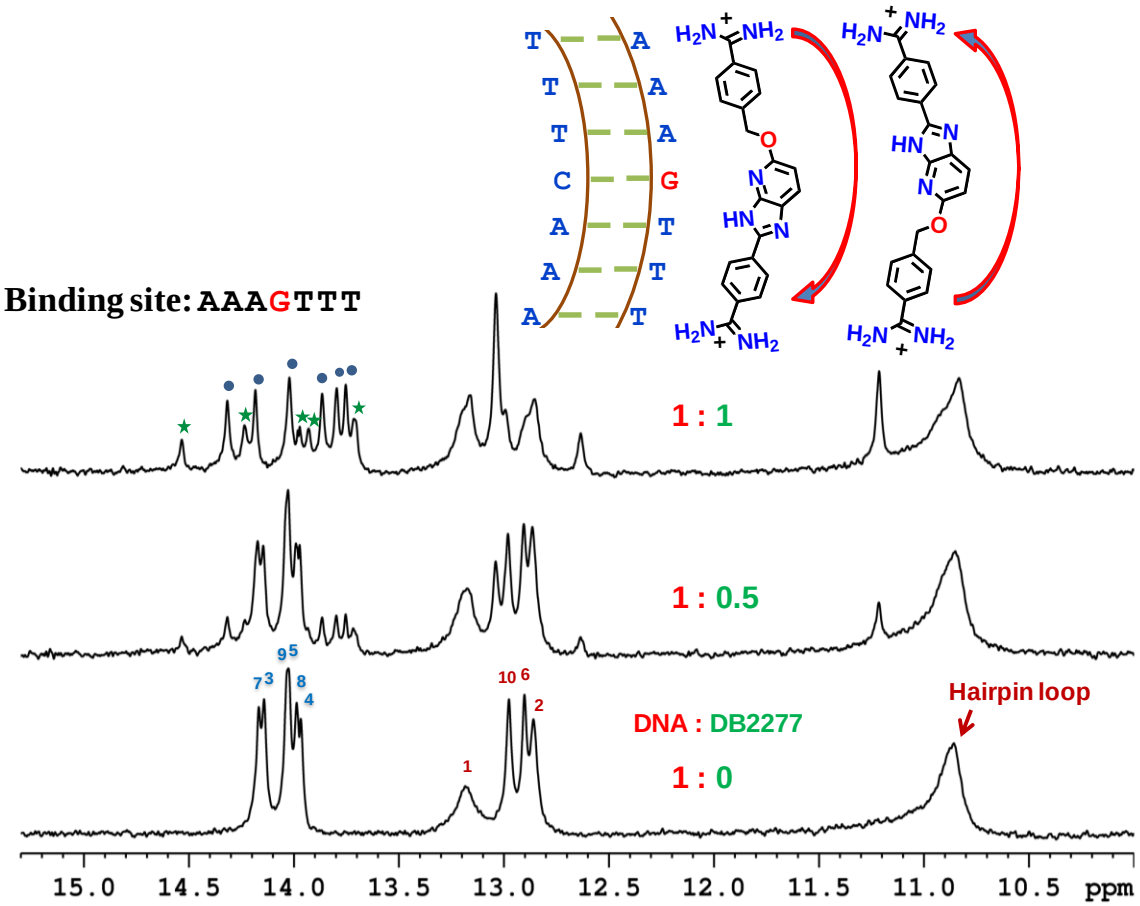


Figure 2.1 Shifts in the imino spectral peaks at different G-hp1:DB2277 ratios at 285 K.

The major and minor binding species of DNA-DB2277 complex are represented by "●" and "★" labels, respectively. A schematic model showing two possible orientations of the DB2277 complex with G-hp1 is at the top. The NMR sample consisted of 100 µM DNA at pH 6.9.

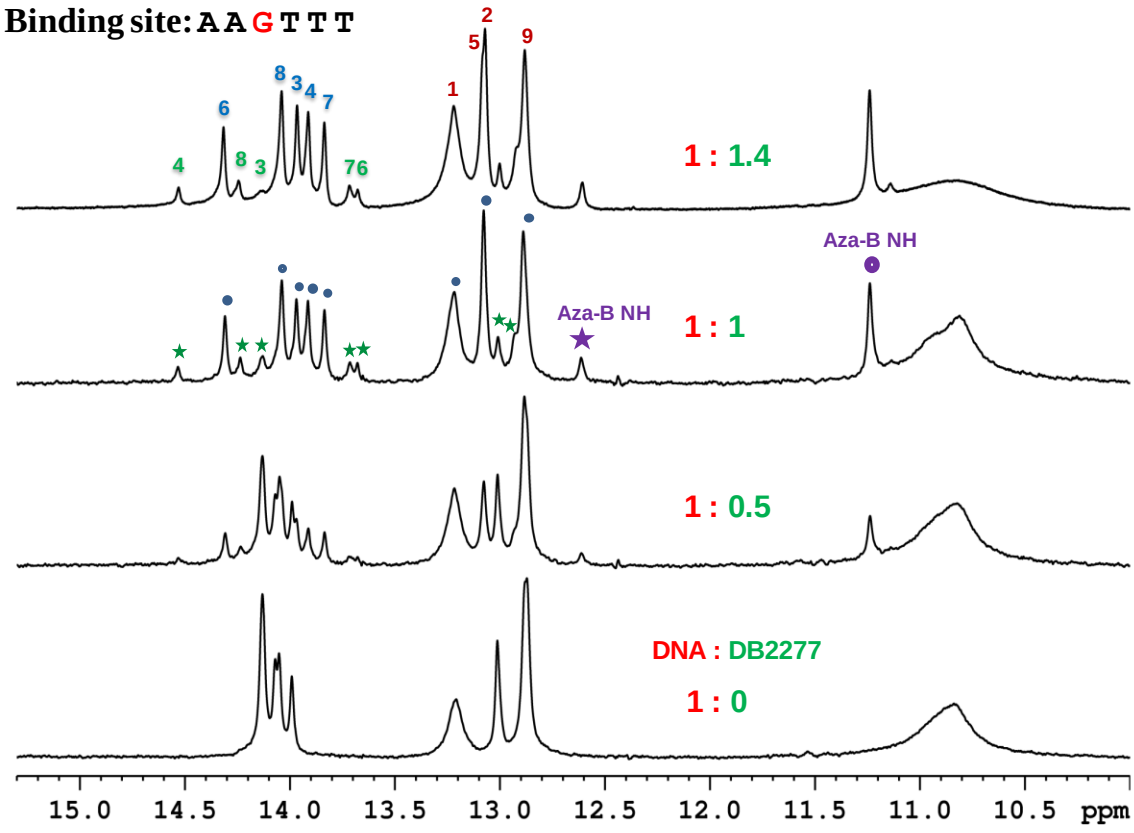


Figure 2.2 Shifts in the imino proton peaks at different G-hp2:DB2277 ratios at 285 K.

The major and minor binding species of the DNA-DB2277 complex are represented by "● " and "★" labeled peaks respectively along with the exchangeable protons of aza-benzimidazole NH that are assigned from 1D-NOE and 2D NOESY shown in Figure 2.3 and 2.7. The NMR sample consisted of 100 µM DNA at pH 6.7.

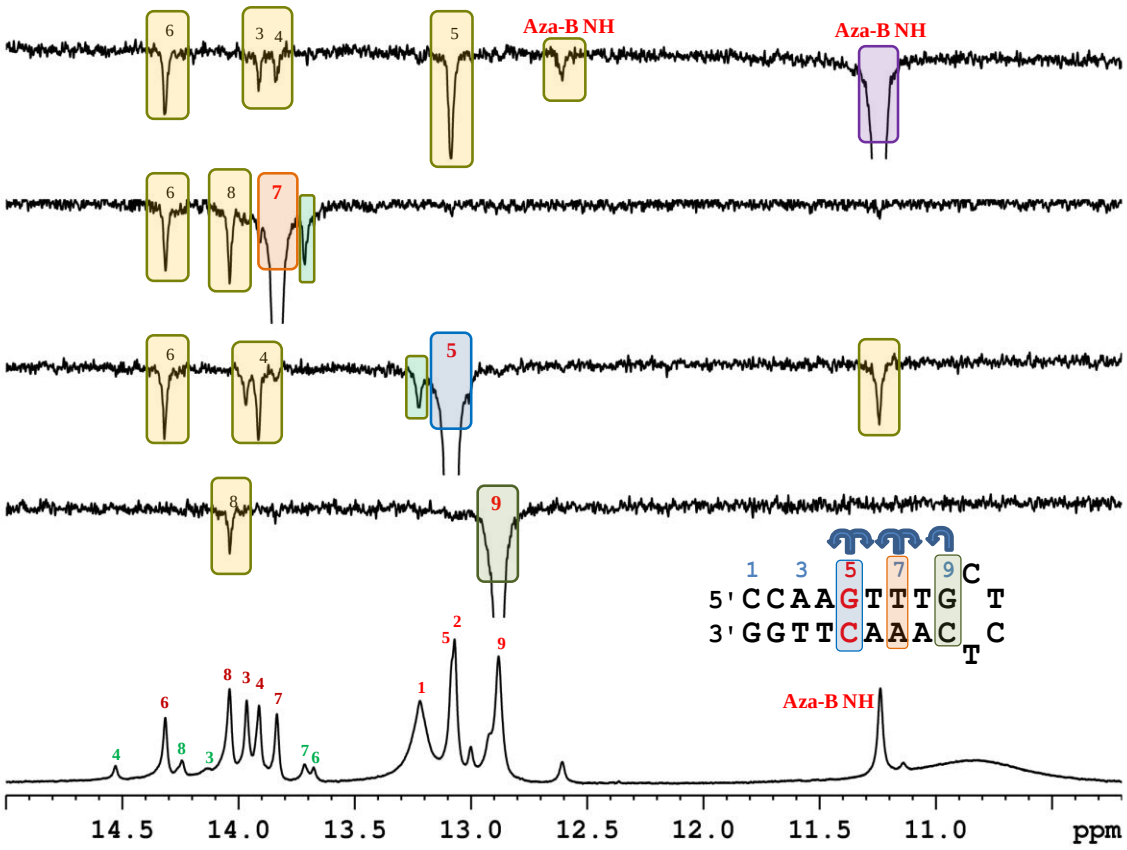


Figure 2.3 Sequential assignment of imino protons of G-hp2 complex with DB2277 using 1D-NOE difference spectra.

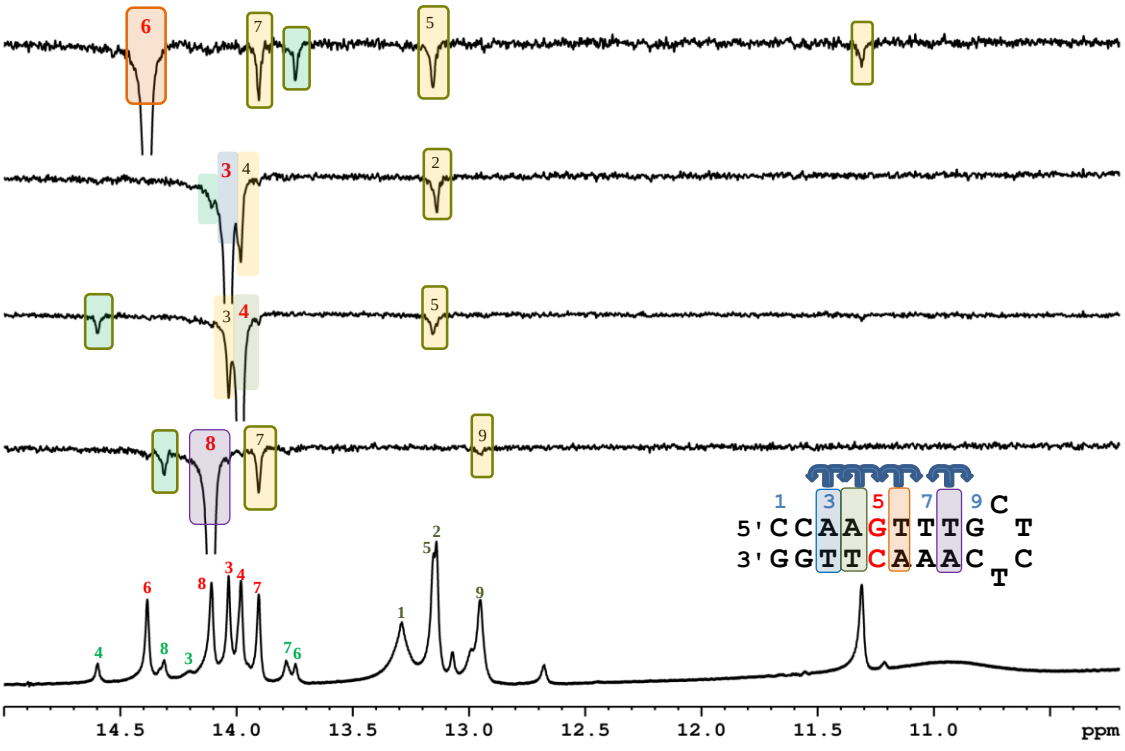


Figure 2.4 Sequential assignment of imino protons of G-hp2 complex with DB2277 using several 1D-NOE difference spectra in conjunction with 2D NOESY experiments.

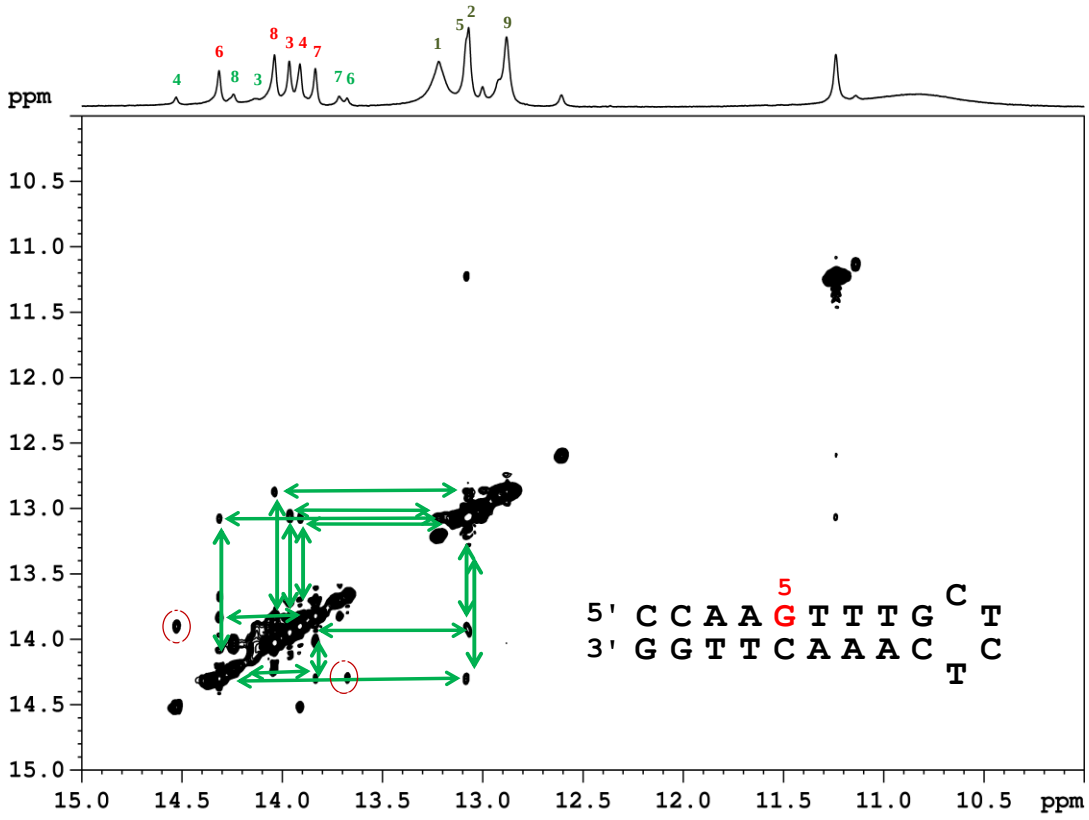


Figure 2.5 Imino proton region of 2D NOESY spectrum of G-hp2-DB2277 complex using a 150 ms mixing time at 285 K.
Both NOE and exchange peaks (in circle) are visible.

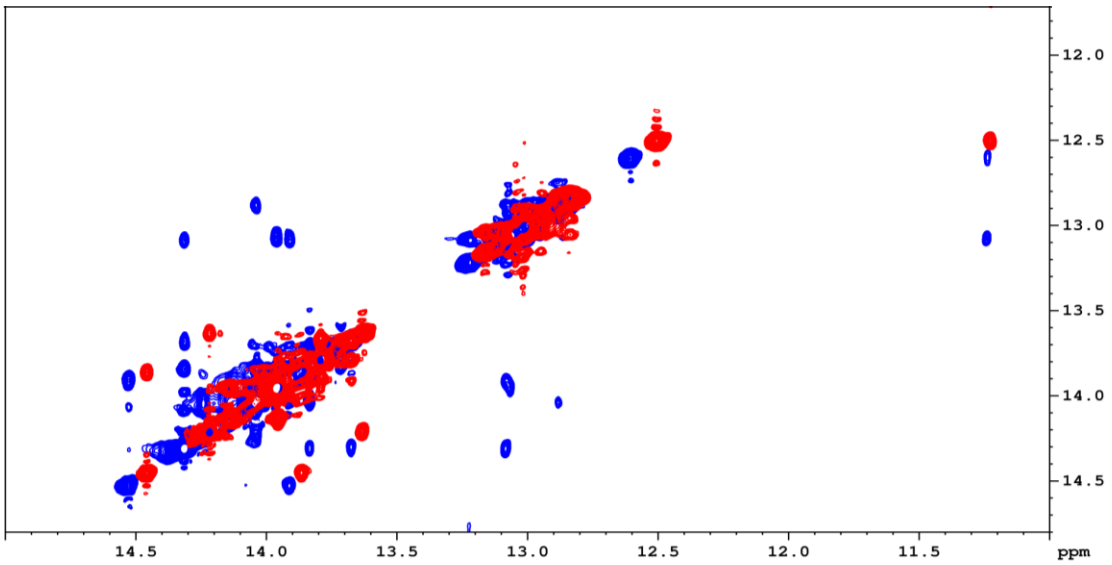


Figure 2.6 Overlay of imino proton region of the NOESY spectrum (in blue) at 285 K and the EXSY spectrum (in red) at 303 K of G-hp2-DB2277 complex.
EXSY experiment conducted at short mixing time 50 ms whereas NOESY experiment was conducted at 150 ms mixing time). Red cross peaks show exchange between major and minor species. Shifts in imino proton peaks of NOESY and EXSY are due to the temperature difference.

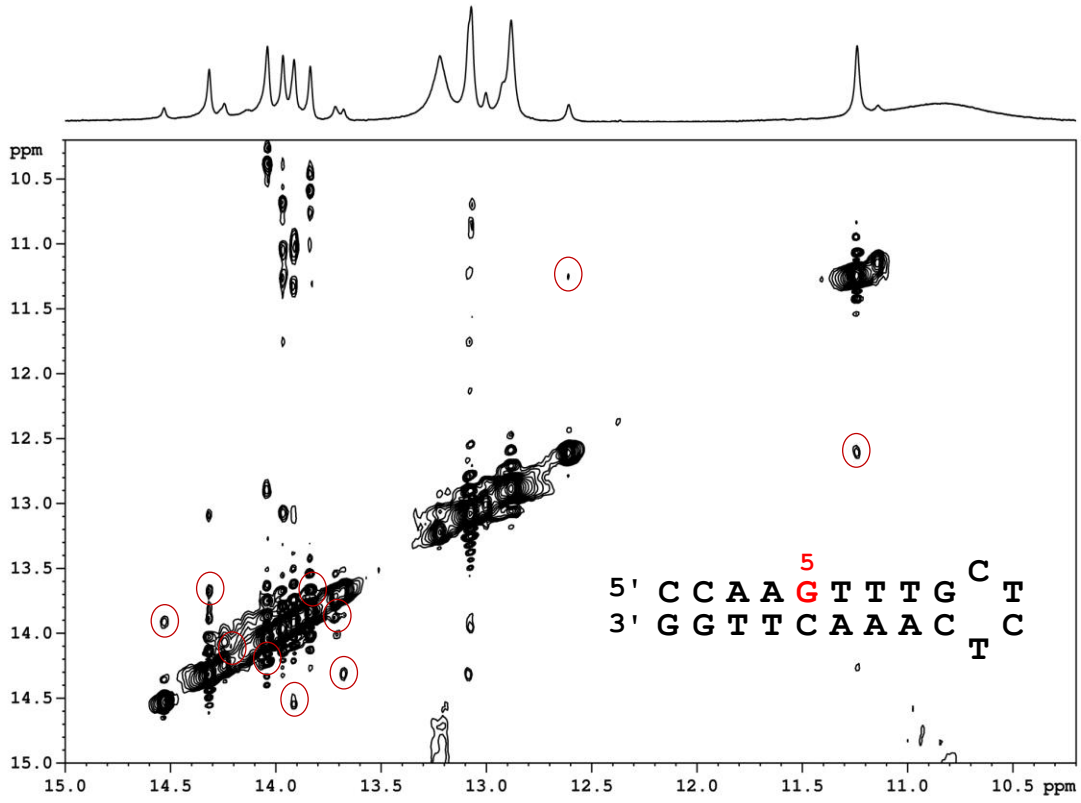


Figure 2.7 Imino proton region of EXSY spectrum of G-hp2-DB2277 complex at 285 K. *2D NOESY recorded at short mixing time of 50 ms. Exchange cross peaks are marked with circles.*

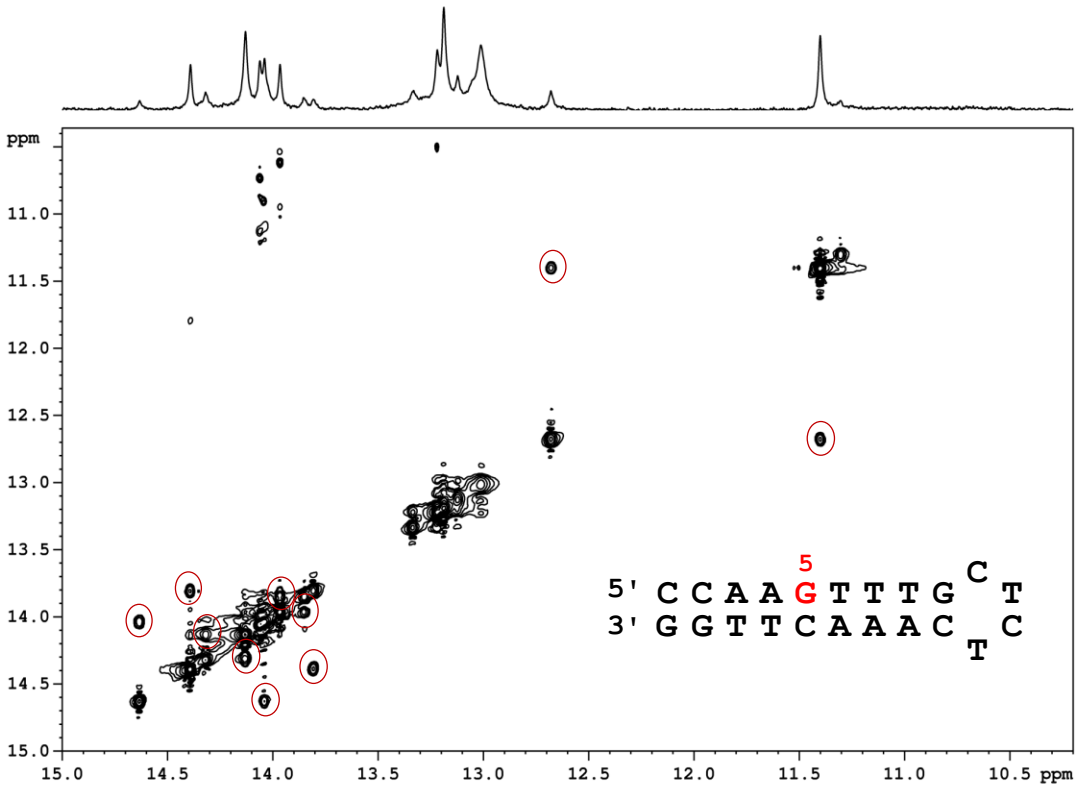


Figure 2.8 Imino proton region of EXSY spectrum of G-hp2-DB2277 complex at 303 K. *EXSY spectrum was recorded at a short mixing time 50 ms. Exchange cross peaks are marked in circle.*

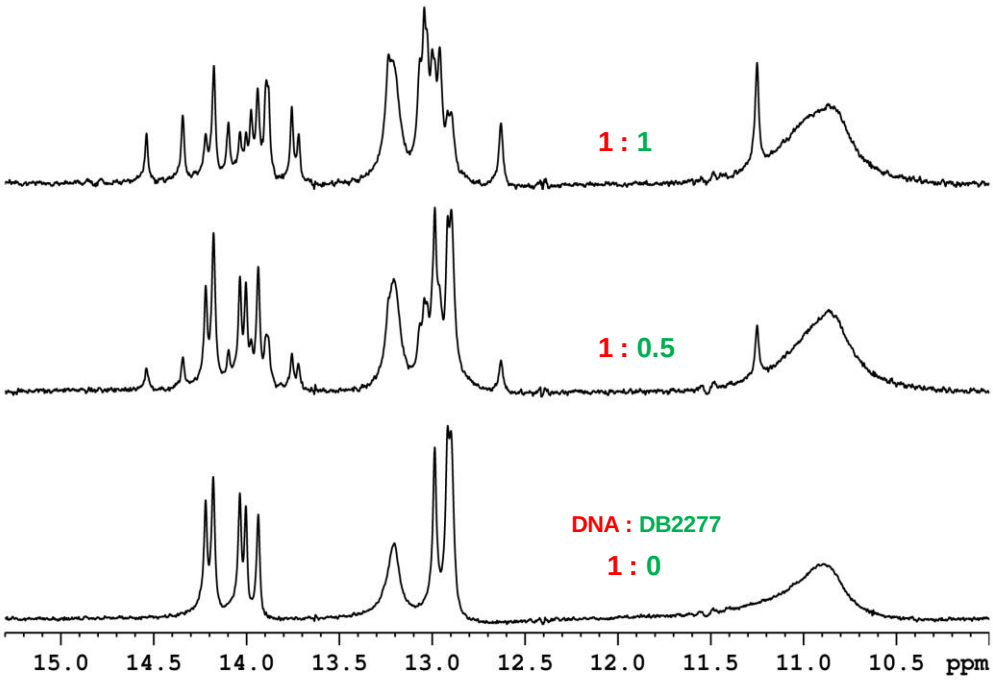


Figure 2.9 Shifts in the imino spectral region peaks at different G-hp3:DB2277 ratios at 285 K.

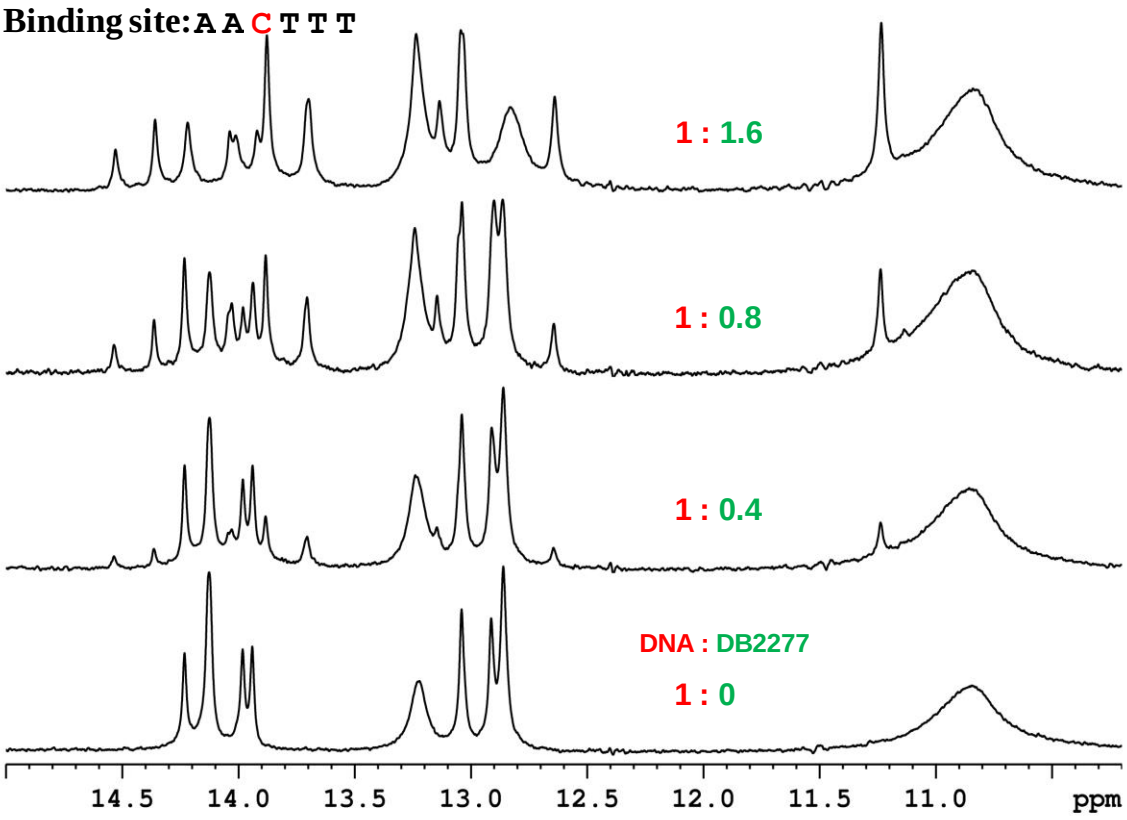


Figure 2.10 Shifts in the imino spectral region peaks at different G-hp4:DB2277 ratios at 285 K.

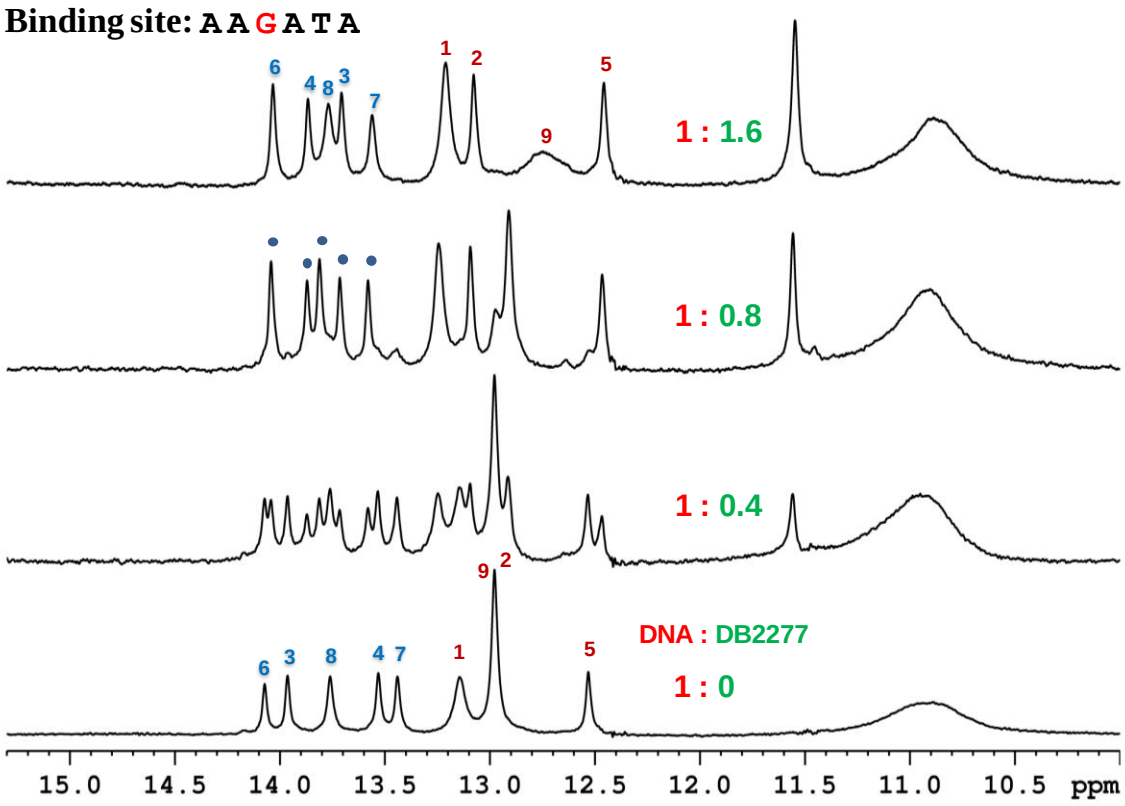


Figure 2.11 Shifts in the imino spectral region peaks at different G-hp5:DB2277 ratios at 285 K.

Unique and selective binding species of the DNA-DB2277 complex are represented by the labeled peaks. The NMR sample consisted of 100 µM DNA at pH 6.7.

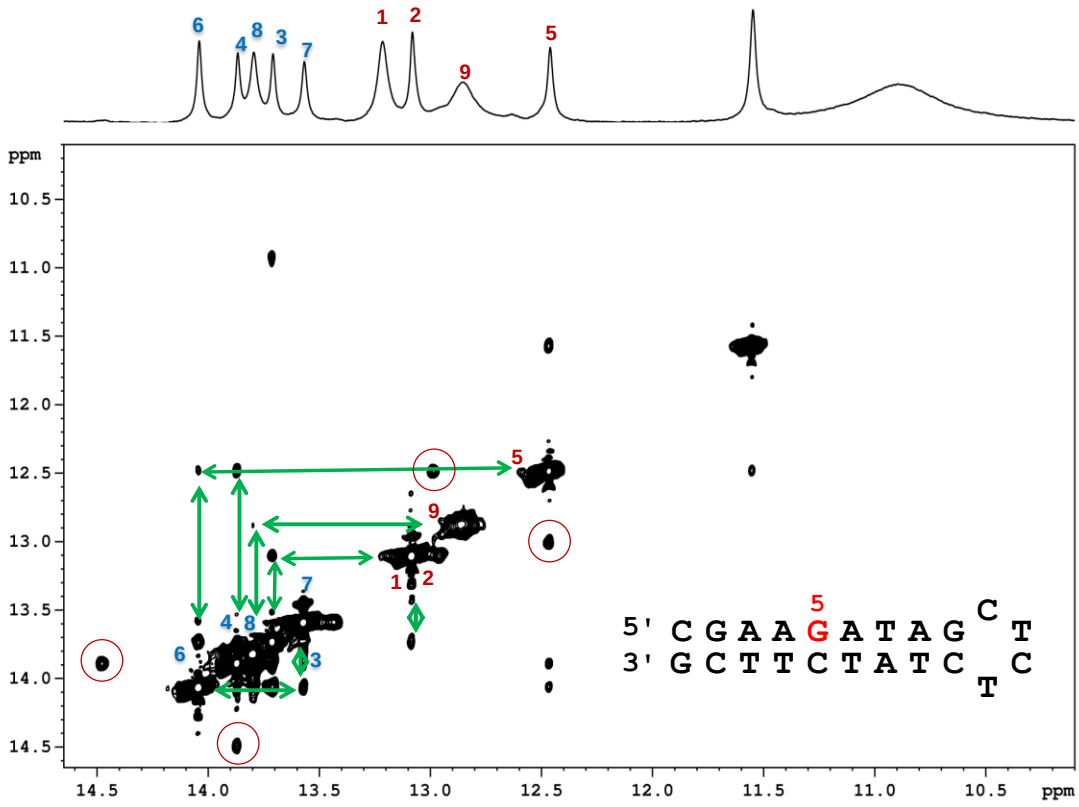


Figure 2.12 Imino proton assignment of G-hp5-DB2277 complex using 2D NOESY at 150 ms mixing time.
Exchange cross peaks are marked in circle.

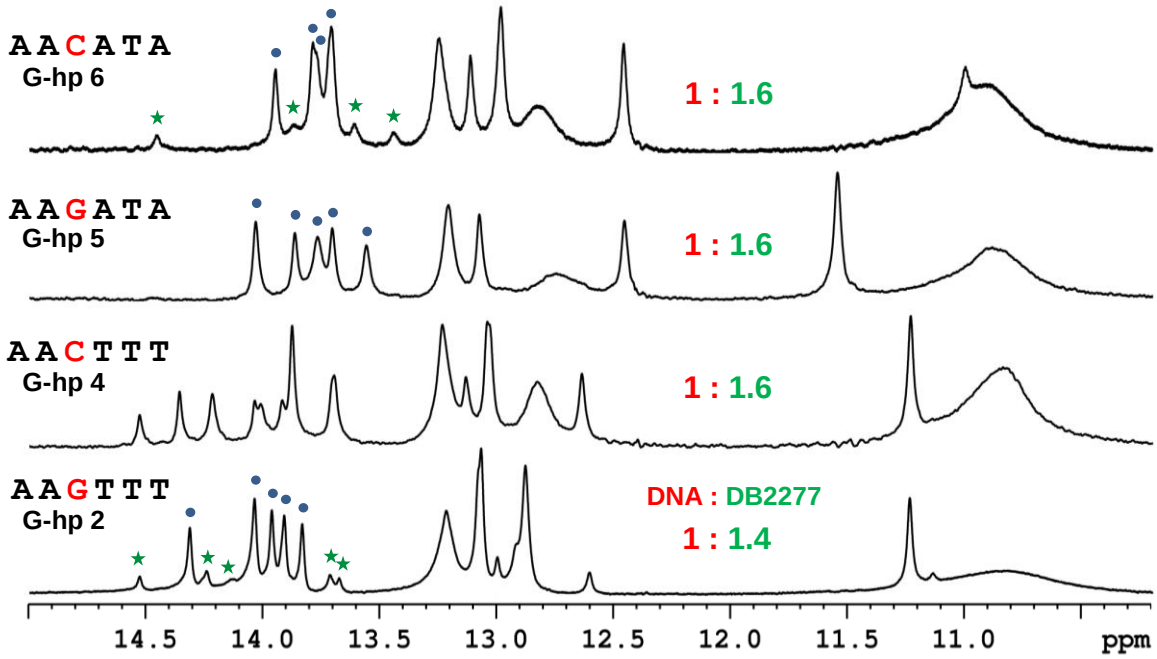


Figure 2.13 Comparison of imino proton peaks of G-hp DNA-DB2277 complexes at 285 K.

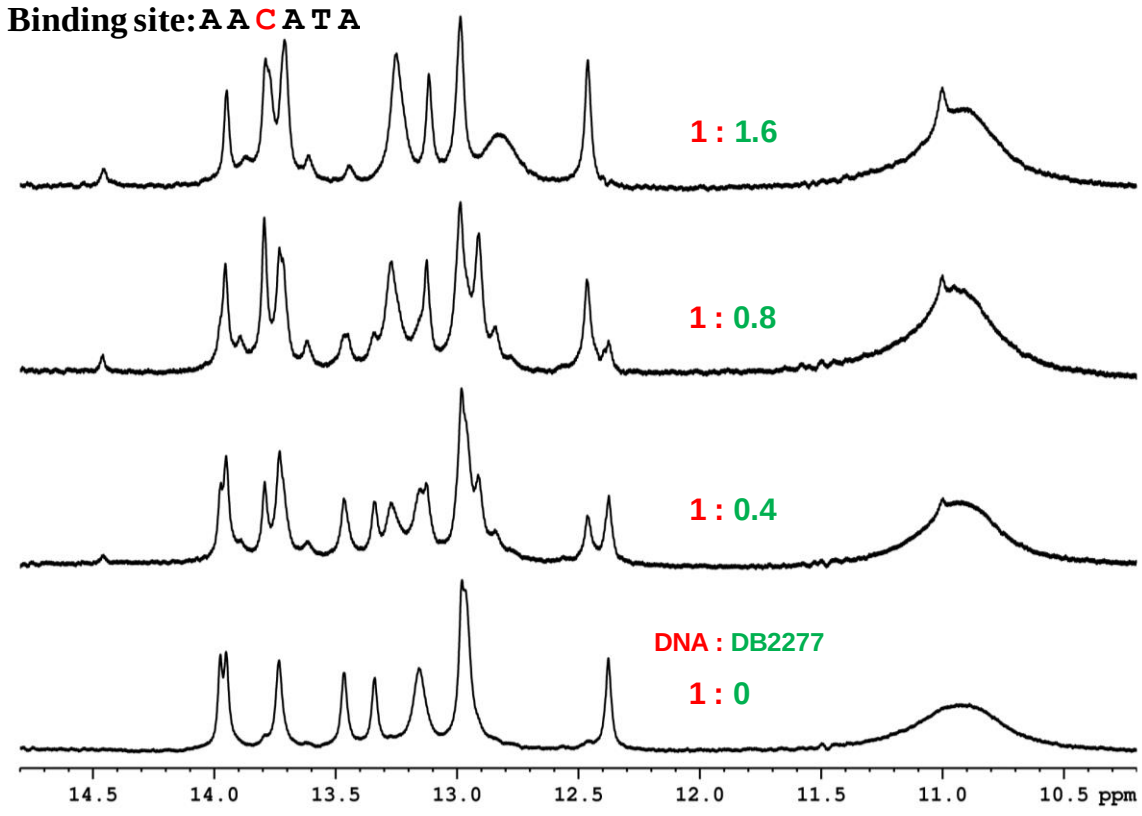


Figure 2.14 Shifts in the imino spectral region peaks at different G-hp6:DB2277 ratios at 285 K.

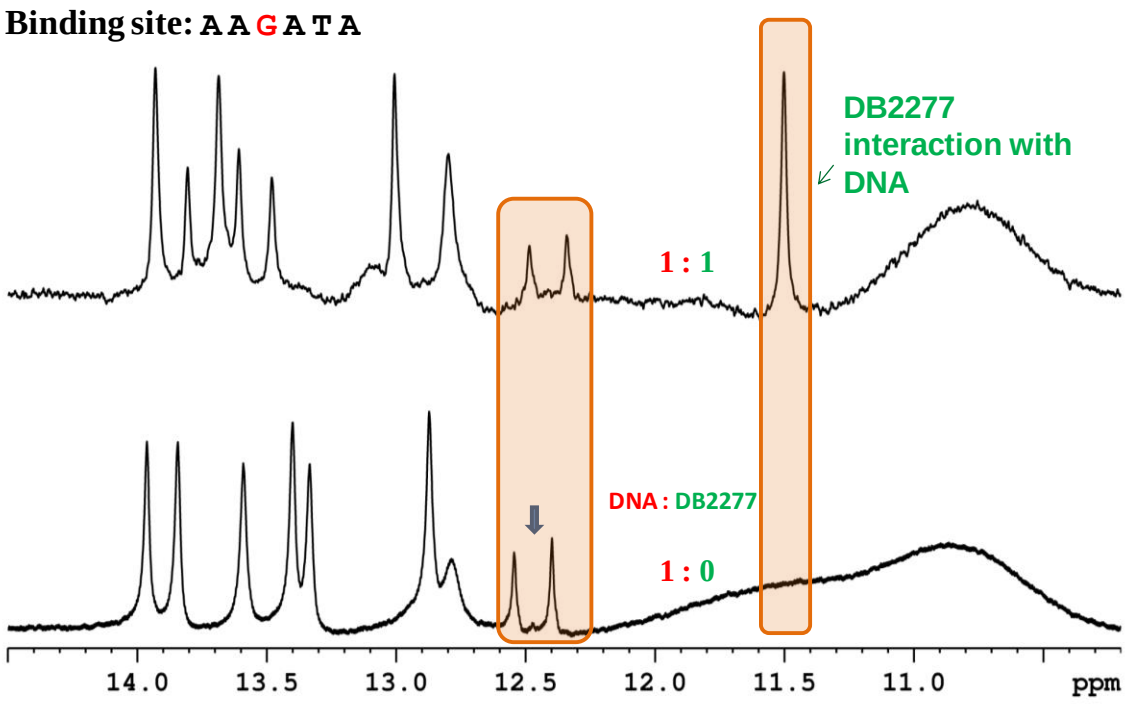


Figure 2.15 Shift of imino proton peaks at 308 K in ^{1}H NMR spectra of ^{15}N labeled G-hp5. *Splitting of imino proton of G-5 bp is highlighted along with the emergence of a new peak at 11.5 ppm. The NMR sample consisted of 100 µM DNA at pH 6.7.*

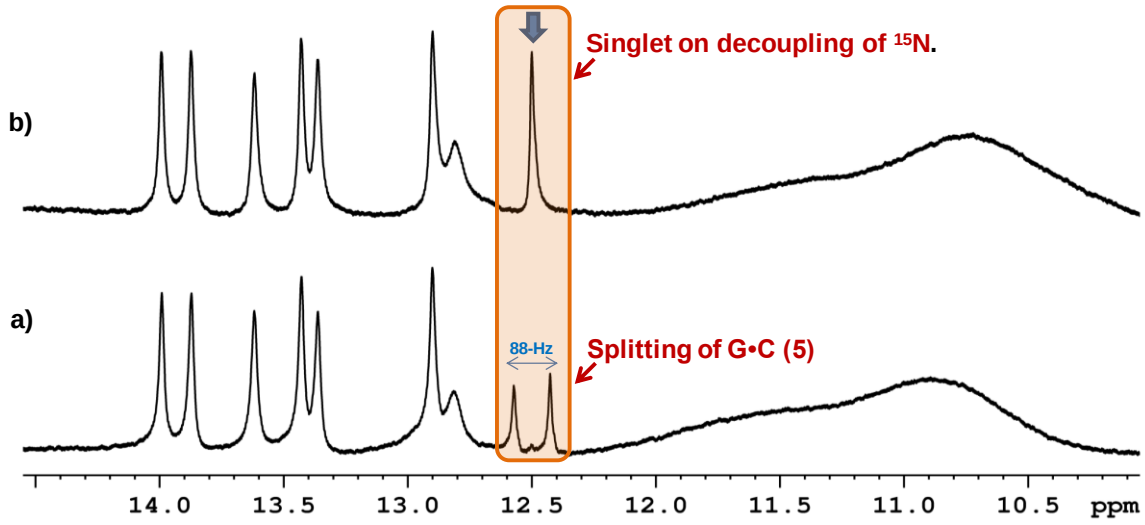


Figure 2.16 a) Splitting of G•C (5) imino proton peak due to ^{15}N-^{1}H coupling in the imino spectral region of ^{1}H NMR spectrum at 308 K. b) Splitting of G•C (5) imino proton peak reverts back to singlet upon decoupling of ^{15}N.

a) Splitting of G•C (5) imino proton peak due to ^{15}N-^{1}H coupling in the imino spectral region of ^{1}H NMR spectrum at 308 K. b) Splitting of G•C (5) imino proton peak reverts back to singlet upon decoupling of ^{15}N.

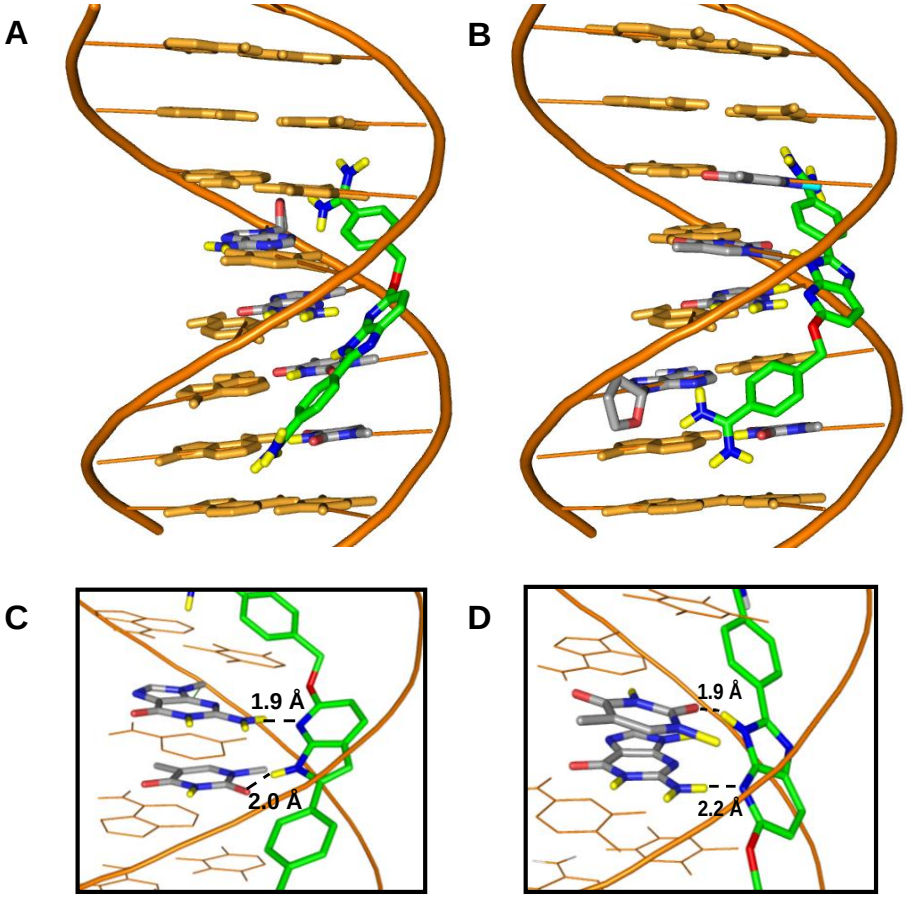


Figure 2.17 Minor groove views of the docked conformations of DB2277 with the AAAGTTT. *(A, B) The stick in green-blue-red-yellow represents DB2277 ligand. The DNA backbone is represented as a tube form in orange, DB2277 interacting bases are in stick with grey-blue-red-yellow and other bases are in stick with orange. (C, D) Important H-bond interactions between aza-N- with the exocyclic G6-NH_2 and aza-benzimidazole NH either with N3 of A or the carbonyl group of T at bases flanking the central G.*

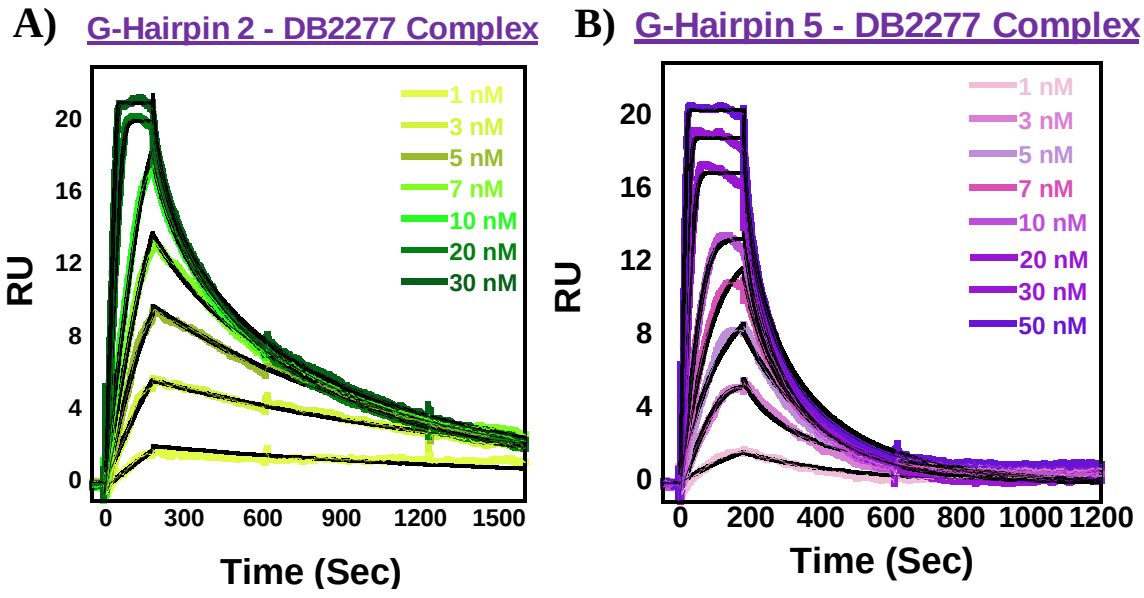


Figure 2.18 Representative SPR sensorgrams for DB2277 binding to G-hp2 and G-hp5. *Inset, the increase in the concentration of DB2277 as injected on immobilized DNA sequence. The solid black lines are best-fit values for the global kinetic fitting of the results with a single site function.*

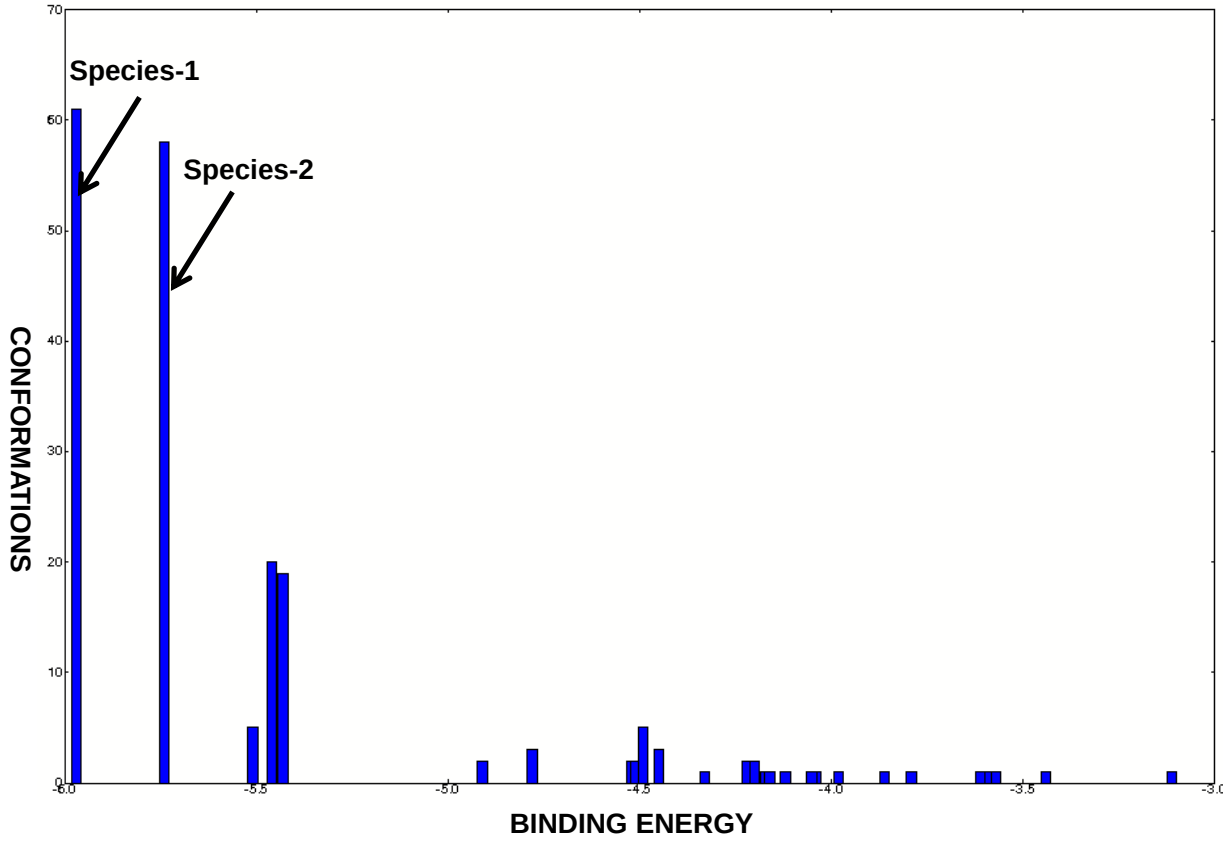


Figure 2.19 Clustering Histogram of multi-member conformations found out of 200 Lamarckian Genetic Algorithm docking runs.

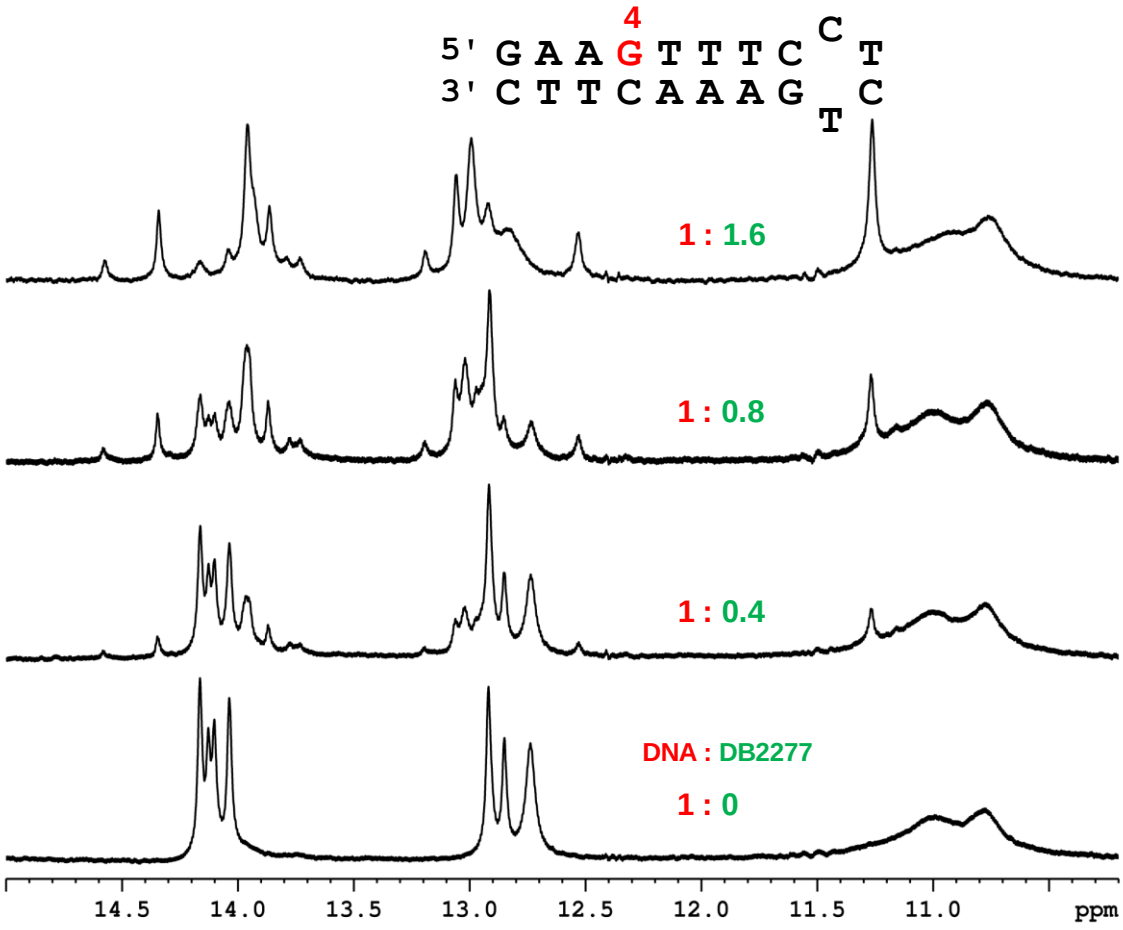


Figure 2.20 Shifts in peaks at different G-hp7-:DB2277 ratios in imino proton spectrum at 285K. Spectra show significant minor form.

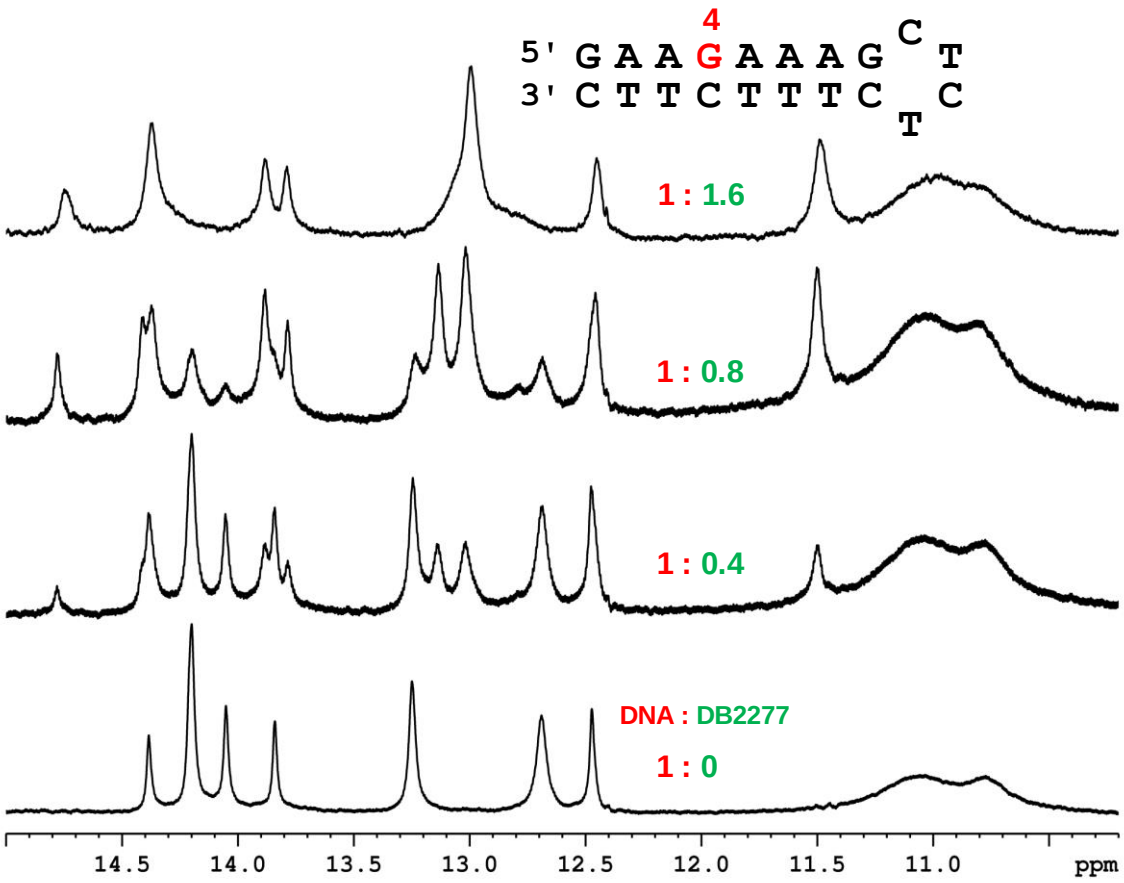


Figure 2.21 Shifts in peaks at different G-hp7-:DB2277 ratios in imino proton spectrum at 285K. Spectra show significant minor form.

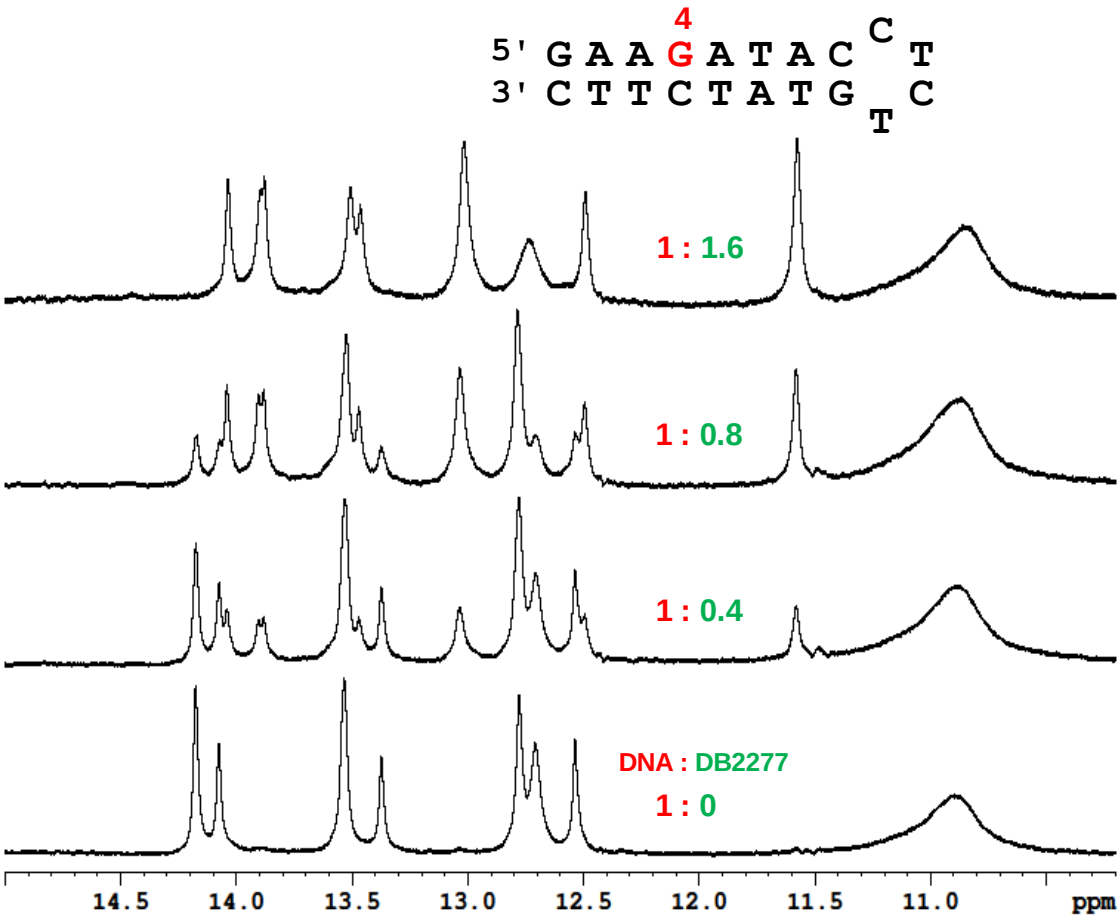


Figure 2.22 Shifts in peaks at different G-hp9:DB2277 ratios in imino proton spectrum at 285K. In this sequence context the minor form is suppressed.

Table 2.1 Selected mixed-sequence hairpin DNAs.

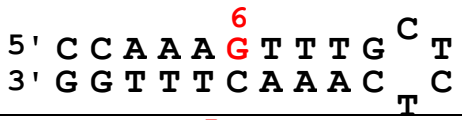
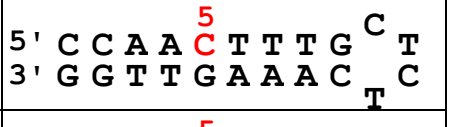
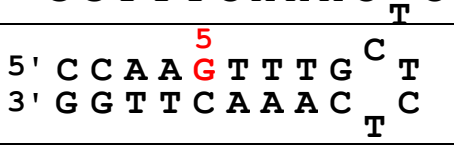
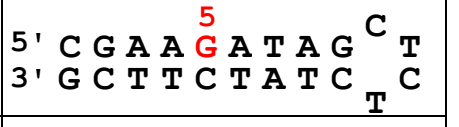
Abbrev.	Sequence	Abbrev.	Sequence
G-hp1 **AAAGTTT**		**G-hp4** **AACTTT**	
G-hp2 **AAGTTT**		**G-hp5** **AAGATA**	
G-hp3 **AAAGTT**		**G-hp6** **AACATA**	

Table 2.2 Kinetic and equilibrium parameters calculated using SPR and NMR.

Complex	k_a ($M^{-1}s^{-1}$) [a]	k_d (s^{-1}) [a]	K_D (nM) [a]	$1/k_d$ (s) [a]	k_{ex} (s^{-1}) [b]	$1/k_{ex}$ (s) [b]
G-hp2-DB2277	$(1.4 \pm 0.8) \times 10^7$	0.02 ± 0.01	1.3 ± 0.1	55.5	6.8 ± 1.3	0.15
G-hp5-DB2277	$(0.6 \pm 0.4) \times 10^7$	0.03 ± 0.02	4.4 ± 0.3	33.5	-	-

[a] determined from SPR. [b] determined from NMR at short mixing time (50 ms) at 303 K (40).

Table 2.3 Tm results of DNA hairpin oligomers and their complexes with DB2277 are reported within ± 0.5 °C.

Hairpin DNA	**Native DNA T_m (°C)**	**ΔT_m (DNA-DB2277 complex) (°C)**
G-hp1	68.5	13.5
G-hp2	68.1	13.8
G-hp3	68.5	12.0
G-hp4	68.1	11.9
G-hp5	65.2	10.3
G-hp6	66.9	8.7

Table 2.4 Intensity and molar fraction of major and minor well-resolved peaks at 303 K and 285 K.

Peak No. A•T bp (303 K)	Intensity Major peak I_M	Intensity Minor peak I_m	Intensity Major to minor peak I_{M-m}	Intensity Minor to major peak I_{m-M}	Molar fraction X_M	Molar fraction X_m
4	2.38E+09 2.42E+09 2.29E+09	3.13E+08 3.25E+08 3.22E+08	9.02E+07 9.36E+07 8.96E+07	9.83E+07 1.09E+08 9.83E+07	0.88 0.88 0.88	0.12 0.12 0.12
6	1.97E+09 1.93E+09 1.93E+09	4.68E+08 4.09E+08 4.59E+08	9.79E+07 9.95E+07 9.95E+07	8.47E+07 8.09E+07 8.51E+07	0.81 0.81 0.81	0.19 0.19 0.19
7	2.04E+09 2.05E+09 1.91E+09	5.65E+08 5.95E+08 5.46E+08	1.43E+08 1.40E+08 1.35E+08	1.65E+08 1.67E+08 1.67E+08	0.78 0.78 0.78	0.22 0.22 0.22

Peak No. A•T bp (285 K)	Intensity Major peak I_M	Intensity Minor peak I_m	Intensity Major to minor peak I_{M-m}	Intensity Minor to major peak I_{m-M}	Molar fraction X_M	Molar fraction X_m
4	8.72E+08 8.72E+08 8.25E+08	1.16E+08 1.17E+08 1.16E+08	1.43E+07 1.33E+07 1.43E+07	9.61E+06 9.61E+06 9.61E+06	0.89 0.89 0.88	0.11 0.11 0.12
8	8.63E+08 9.86E+08 1.03E+09	2.16E+08 2.14E+08 2.52E+08	2.50E+07 2.66E+07 3.64E+07	2.38E+07 2.51E+07 2.51E+07	0.80 0.82 0.80	0.20 0.18 0.20

Table 2.5 Exchange rate calculation of G-hp2-DB2277 complex from EXSY at 303 K and 285 K.

Peak No. AT bp	k_{ex} (at 303 K)	ln(K_{eq}) (at 303 K)	ΔG (Kcal/mol) (at 303 K)	k_{ex} (at 285 K)	ln (K_{eq}) (at 285K)	ΔG (Kcal/mol) (at 285 K)
4	7.7 ± 0.3	- 2.00	1.20 ± 0.02	2.5 ± 0.1	- 2.00	1.13 ± 0.02
6	5.3 ± 0.1	- 1.47	0.89 ± 0.04	-	-	-
7	7.4 ± 0.3	- 1.26	0.76 ± 0.02	-	-	-
8	-	-	-	3.0 ± 0.1	- 1.44	0.82 ± 0.04

[a]*The reported errors are calculated using standard deviation from three repeated measurements on the same sample.*

2.8 References

1. Raskatov, J. A., Nickols, N. G., Hargrove, A. E., Marinov, G. K., Wold, B., and Dervan, P. B. (2012) Gene expression changes in a tumor xenograft by a pyrrole-imidazole polyamide. Proc. *Natl. Acad. Sci. U.S.A.,* **109**, 16041-16045.

2. Wilson, W. D., Tanious, F. A., Mathis, A., Tevis, D., Hall, J. E., and Boykin, D. W. (2008) Antiparasitic compounds that target DNA. *Biochimie,* **90**, 999-1014.

3. Ding, P., McFarland, K. A., Jin, S., Tong, G., Duan, B., Yang, A., Hughes, T. R., Liu, J., Dove, S. L., Navarre, W. W., and Xia, B. (2015) A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. *PLoS Pathog.,* **11**, 1-26.

4. Rodriguez, J., Mosquera, J., Couceiro, J. R., Vazquez, M. E., and Mascarenas, J. L. (2015) The AT-Hook motif as a versatile minor groove anchor for promoting DNA binding of transcription factor fragments. *Chem. Sci.,* **6**, 4767-4771.

5. Hashimoto, Y. (1994) Chemical control of gene expression. *Yakugaku Zasshi*, **114**, 357-73.

6. Nguyen, B., Hamelberg, D., Bailly, C., Colson, P., Stanek, J., Brun, R., Neidle, S., and Wilson, W. D. (2004) Characterization of a Novel DNA Minor-Groove Complex. *Biophys. J.,* **86**, 1028-1041.

7. Chavda, S., Liu, Y., Babu, B., Davis, R., Sielaff, A., Ruprich, J., Westrate, L., Tronrud, C., Ferguson, A., Franks, A., Tzou, S., Adkins, C., Rice, T., Mackay, H., Kluza, J., Tahir, S. A., Lin, S., Kiakos, K., Bruce, C. D., Wilson, W. D., Hartley, J. A., and Lee, M. (2011) Hx, a Novel Fluorescent, Minor Groove and Sequence Specific Recognition Element: Design,

Synthesis, and DNA Binding Properties of p-Anisylbenzimidazole-imidazole/pyrrole-Containing Polyamides. *Biochemistry,* **50**, 3127-3136.

8. Kang, J. S., Meier, J. L., and Dervan, P. B. (2014) Design of Sequence-Specific DNA Binding Molecules for DNA Methyltransferase Inhibition. *J. Am. Chem. Soc.,* **136**, 3687-3694.

9. Liu, Y., Kumar, A., Boykin, D. W., and Wilson, W. D. (2007) Sequence and length dependent thermodynamic differences in heterocyclic diamidine interactions at AT base pairs in the DNA minor groove. *Biophys. Chem.,* **131**, 1-14.

10. Yamamoto, M., Bando, T., Morinaga, H., Kawamoto, Y., Hashiya, K., and Sugiyama, H. (2014) Sequence-Specific DNA Recognition by Cyclic Pyrrole-Imidazole Cysteine-Derived Polyamide Dimers. *Chem. Eur. J.,* **20**, 752-759.

11. Alniss , H. Y., Salvia, M.-V., Sadikov, M., Golovchenko, I., Anthony, N. G., Khalaf, A. I., MacKay, S. P., Suckling, C. J. and Parkinson, J. A. (2014) Recognition of the DNA Minor Groove by Thiazotropsin Analogues. *ChemBioChem.*, **15**, 1978–1990.

12. Barrett, M.P., Gemmell, C.G. and Suckling, C.J. (2013) Minor groove binders as anti-infective agents. *Pharmacol. Ther.*, **139**, 12-23.

13. Baraldi, P. G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M. A., Pavani, M. G., and Romagnoli, R. (2004) DNA minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.,* **24**, 475-528.

14. Xinbo, Z., Zhang, S. C., Dejun, S., Jiang, H., Wali, A., Pass, H., Fernandez-Madrid, F., Harbut, M. R., and Naimei, T. (2011) New Insight into the Molecular Mechanisms of the Biological Effects of DNA Minor Groove Binders. *PLoS ONE.,* **6**, 1-12.

15. Matovu, E., Stewart, M. L., Geiser, F., Brun, R., Mäser, P., Wallace, L. J. M., Burchmore, R. J., Enyaru, J. C. K., Barrett, M. P., Kaminsky, R., Seebeck, T., and de Koning, H. P. (2003) Mechanisms of Arsenical and Diamidine Uptake and Resistance in Trypanosoma brucei. *Eukaryotic Cell,* **2**, 1003-1008.

16. Paine, M. F., Wang, M. Z., Generaux, C. N., Boykin, D. W., Wilson, W. D., De Koning, H. P., Olson, C. A., Pohlig, G., Burri, C., Brun, R., Murilla, G. A., Thuita, J. K., Barrett, M. P., and Tidwell, R. R. (2010) Diamidines for human African trypanosomiasis, *Curr. Opin. Investig. Drugs,* **11**, 876-83.

17. Sturk, L. M., Brock, J. L., Bagnell, C. R., Hall, J. E., and Tidwell, R. R. (2004) Distribution and quantitation of the anti-trypanosomal diamidine 2,5-bis(4-amidinophenyl)furan (DB75) and its N-methoxy prodrug DB289 in murine brain tissue. *Acta Trop.,* **91**, 131-143.

18. Munde, M., Kumar, A., Nhili, R., Depauw, S., David-Cordonnier, M.-H., Ismail, M. A., Stephens, C. E., Farahat, A. A., Batista-Parra, A., Boykin, D. W., and Wilson, W. D. (2010) DNA Minor Groove Induced Dimerization of Heterocyclic Cations: Compound Structure, Binding Affinity, and Specificity for a TTAA Site. *J. Mol. Biol.,* **402**, 847-864.

19. Nanjunda, R., and Wilson, W. D. (2012) Binding to the DNA Minor Groove by Heterocyclic Dications: From AT-Specific Monomers to GC Recognition with Dimers. In *Current Protocols in Nucleic Acid Chemistry*, **Chapter 8**, Unit 8.8, Wiley, New York.

20. Munde, M., Kumar, A., Peixoto, P., Depauw, S., Ismail, M. A., Farahat, A. A., Paul, A., Say, M. V., David-Cordonnier, M.-H., Boykin, D. W., and Wilson, W. D. (2014) The Unusual Monomer Recognition of Guanine-Containing Mixed Sequence DNA by a Dithiophene Heterocyclic Diamidine. *Biochemistry,* **53**, 1218-1227.

21. Liu, Y., Chai, Y., Kumar, A., Tidwell, R. R., Boykin, D. W., and Wilson, W. D. (2012) Designed compounds for recognition of 10 base pairs of DNA with two at binding sites. *J. Am. Chem. Soc.,* **134**, 5290-9.

22. Rahimian, M., Kumar, A., Say, M., Bakunov, S. A., Boykin, D. W., Tidwell, R. R., and Wilson, W. D. (2009) Minor Groove Binding Compounds That Jump a GC Base Pair and Bind to Adjacent AT Base Pair Sites. *Biochemistry,* **48**, 1573-1583.

23. Munde, M., Ismail, M. A., Arafa, R., Peixoto, P., Collar, C. J., Liu, Y., Hu, L., David-Cordonnier, M.-H., Lansiaux, A., Bailly, C., Boykin, D. W., and Wilson, W. D. (2007) Design of DNA Minor Groove Binding Diamidines That Recognize GC Base Pair Sequences: A Dimeric-Hinge Interaction Motif. *J. Am. Chem. Soc.,* **129**, 13732-13743.

24. Hunt, R. A., Munde, M., Kumar, A., Ismail, M. A., Farahat, A. A., Arafa, R. K., Say, M., Batista-Parra, A., Tevis, D., Boykin, D. W., and Wilson, W. D. (2011) Induced topological changes in DNA complexes: influence of DNA sequences and small molecule structures. *Nuc. Acids Res.,* **39**, 4265-4274.

25. Nguyen, B., Neidle, S., and Wilson, W. D. (2009) A Role for Water Molecules in DNA−Ligand Minor Groove Recognition. *Acc. Chem. Res.,* **42**, 11-21.

26. Miao, Y., Lee, M. P. H., Parkinson, G. N., Batista-Parra, A., Ismail, M. A., Neidle, S., Boykin, D. W., and Wilson, W. D. (2005) Out-of-Shape DNA Minor Groove Binders: Induced Fit Interactions of Heterocyclic Dications with the DNA Minor Groove. *Biochemistry,* **44**, 14701-14708.

27. Munde, M., Lee, M., Neidle, S., Arafa, R., Boykin, D. W., Liu, Y., Bailly, C., and Wilson, W. D. (2007) Induced Fit Conformational Changes of a "Reversed Amidine" Heterocycle: Optimized Interactions in a DNA Minor Groove Complex. *J. Am. Chem. Soc.,* **129**, 5688-5698.

28. He, G., Vasilieva, E., Davis, G., Jr., Koeller, K. J., Bashkin, J. K., and Dupureur, C. M. (2014) Binding studies of a large antiviral polyamide to a natural HPV sequence. *Biochimie,* **102**, 83−89.

29. Paul, A., Nanjunda, R., Kumar, A., Laughlin, S., Nhili, R., Depauw, S., Deuser, S.S., Chai, Y., Chaudhary, A.S., David-Cordonnier, M. H., Boykin, D. W., and Wilson, W. D. (2015) Mixed up minor groove binders: Convincing A·T specific compounds to recognize a G·C base pair. *Bioorg. Med. Chem. Lett.*, **25**, 4927-4932.

30. Satam, V., Babu, B., Patil, P., Brien, K. A., Olson, K., Savagian, M., Lee, M., Mepham, A., Jobe, L. B., Bingham, J. P., Pett, L., Wang, S., Ferrara, M., Bruce, C. D., Wilson, W. D., Lee, M., Hartley, J. A., and Kiakos, K. (2015) AzaHx, a novel fluorescent, DNA minor groove and G•C recognition element: Synthesis and DNA binding properties of a p-anisyl-4-aza-benzimidazole-pyrrole-imidazole (azaHx-PI) polyamide. *Bioorg. Med. Chem. Lett.,* **25**, 3681-3685.

31. Wei, D., Wilson, W. D., and Neidle, S. (2013) Small-molecule Binding to the DNA Minor Groove Is Mediated by a Conserved Water Cluster. *J. Am. Chem. Soc.,* **135**, 1369-1377.

32. Chai, Y., Paul, A., Rettig, M., Wilson, W. D., and Boykin, D. W. (2014) Design and synthesis of heterocyclic cations for specific DNA recognition: From AT-rich to mixed-base-pair DNA sequences. *J. Org. Chem.,* **79**, 852-66.

33. Paul, A., Chai, Y., Boykin, D. W., and Wilson, W. D. (2015) Understanding mixed sequence DNA recognition by novel designed compounds: the kinetic and thermodynamic behavior of azabenzimidazole diamidines. *Biochemistry,* **54**, 577-87.

34. Rettig, M., Germann, M. W., Wang, S., and Wilson, W. D. (2013) Molecular basis for sequence-dependent induced DNA bending. *Chembiochem*., **14**, 323-31.

35. Fede, A., Billeter, M., Leupin, W. and Wuthrich, K. (1993) Determination of the NMR solution structure of the Hoechst 33258-d(GTGGAATTCCAC)2 complex and comparison with the X-ray crystal structure. *Structure*, **1**, 177-186.

36. Trotta, E., and Paci, M. (1998) Solution structure of DAPI selectively bound in the minor groove of a DNA T.T mismatch-containing site: NMR and molecular dynamics studies. *Nuc. Acids Res.,* **26**, 4706-4713.

37. Conte, M. R., Jenkins, T. C., and Lane, A. N. (1995) Interaction of Minor-Groove-Binding Diamidine Ligands with an Asymmetric DNA Duplex. *Eur. J. Biochem.,* **229**, 433-444.

38. Rydzewski, J. M., Leupin, W., and Chazin, W. (1996) The Width of the Minor Groove Affects the Binding of the Bisquaternary Heterocycle SN-6999 to Duplex DNA. *Nuc. Acids Res.,* **24**, 1287-1293.

39. Lewis, E. A., Munde, M., Wang, S., Rettig, M., Le, V., Machha, V., and Wilson, W. D. (2011) Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nuc. Acids Res.,* **39**, 9649-9658.

40. Rettig, M., Germann, M. W., Ismail, M. A., Batista-Parra, A., Munde, M., Boykin, D. W., and Wilson, W. D. (2012) Microscopic rearrangement of bound minor groove binders detected by NMR. *J. Phys. Chem. B,* **116**, 5620-7.

41. Fede, A., Labhardt, A., Bannwarth, W., and Leupin, W. (1991) Dynamics and binding mode of Hoechst 33258 to d(GTGGAATTCCAC)2 in the 1:1 solution complex as determined by two-dimensional proton NMR. *Biochemistry,* **30**, 11377-11388.

42. Liu, Y. and Wilson, W.D. (2010) Quantitative analysis of small molecule-nucleic acid interactions with a biosensor surface and surface plasmon resonance detection. *Methods Mol. Biol. (Totowa, NJ, U. S.)*, **613**, 1-23.

43. Zhang, J., and Germann, M. W. (2011) Characterization of secondary amide peptide bond isomerization: Thermodynamics and kinetics from 2D NMR spectroscopy. *Biopolymers,* **95**, 755-762.

44. Bain, A. D. (2003) Chemical exchange in NMR. *Prog. Nucl. Magn. Reson. Spectrosc,* **43**, 63-103.

45. Nguyen, B., Tanious, F. A., and Wilson, W. D. (2007) Biosensor-surface plasmon resonance: Quantitative analysis of small molecule–nucleic acid interactions. *Methods*, **42**, 150-161.

46. Nanjunda, R., Munde, M., Liu, Y., and Wilson, W. D. (2011) In Methods for Studying DNA/Drug Interactions, (Wanunu, M., and Tor, Y., Eds.), **Chapter 4**, CRC Press-Taylor & Francis Group: Boca Raton, FL.

47. Spartan'10 Tutorial and User's Guide (2010) Wavefunction, Inc., Irvine, CA.

48. Sanner, M. F. (1999) Python: A programming language for software integration and development. *J. Mol. Graphics Modell.*, **17**, 57−61.

49. Oleg Trott and Arthur J. Olson (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, 455-461.

50. Perrin, C.L. and Dwyer, T.J. (1990) Application of two-dimensional NMR to kinetics of chemical exchange. *Chem. Rev.*, **90**, 935-967.

3 FIRST STRUCTURE OF A DESIGNED MINOR GROOVE BINDING HETEROCYCLIC CATION THAT SPECIFICALLY RECOGNIZES MIXED DNA BASE PAIR SEQUENCES

Narinder K. Harika, Markus W. Germann* and W. David Wilson*

Harika, N. K., et al. *Chemistry*, 2017, 23: 17612-17620.

My contribution to this chapter was in sample preparation and NMR spectral analyses, restrained molecular dynamics studies, and writing.

3.1 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.

3.2 Key words

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

3.3 Abbreviations

Deoxyribonucleic acid (DNA); Nuclear Magnetic Resonance (NMR); Minor groove binder (MGB); Molecular Dynamics (MD) Simulations; Solution Structure

3.4 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 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 3.1** (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 3.1**). 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 3.1B**). This sequence has been used in the high-resolution NMR investigations of the DB2277–AAGATA complex structure reported here. Both 1D and 2D ^{1}H and ^{31}P spectra were collected in H_2O and D_2O (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 H_2O-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.

3.5 Results

3.5.1 *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 ^{1}H 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. ^{31}P 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 3.1** with their numbering schemes.

3.5.2 *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 3.2**). 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% H_2O/10% D_2O. 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 3.2**) (14). 1D proton NMR experiments of the DNA hairpin with ^{15}N-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).

3.5.3 *2D NMR*

NOESY spectra in D_2O 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% D_2O were used to observe the specific binding of the exchangeable aza 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 3.3 and 3.4**). Protons of free DNA, free DB2277 and the DB2277-DNA complex in solution were assigned from the NOESY spectrum in D_2O solution. These proton assignments were further confirmed and supported by employing 2D TOCSY, ^{31}P-correlated 2D spectra (HPCOR) and ^{13}C-^{1}H HSQC experiments (see details in the Methods section). A proton-decoupled ^{31}P spectrum with phosphorus assignments for free DNA is shown in **Figure 3.5** and chemical shifts of a ^{31}P (δP in ppm) signal for both free DNA and DB2277-DNA complex are plotted in **Figure 3.6**.

B1 and B1' proton signals of DB2277 (**Figure 3.1**) 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 3.13** 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 3.1**. Assignments of free DNA and the DB2277-DNA

complex protons in the "NOE walk" of H6/H8 base-H1' sugar region is shown in the supplementary **Figure 3.4** and **Figure 3.3** respectively.

3.5.4 *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 3.7**. Chemical shift differences of H2'', H3' sugar protons, H6/H8 base protons and H5 proton of cytosine are shown in **Figure 3.8**, and imino proton shift differences are shown in **Figure 3.9**. 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 3.7**).

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 3.7**. The H4' proton of G5 is situated on top of the phenyl ring attached to the OCH_2 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 3.10**. 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 3.7**) 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 3.7**). Small chemical shift differences were observed for protons in the flanking region of the DNA which are not in the vicinity of DB2277 (**Figure 3.8 and Figure 3.9**).

A loss of connectivity through the H1' sugar protons of T7 and T20 in the sequential walk from H1' to base protons (**Figure 3.3**) 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 3.13** 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 3.11**.

3.5.5 *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 3.2**. 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 3.2**. The interactions of the DB2277 protons with DNA protons are also represented in the schematic model shown in **Figure 3.12**.

The aza 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 3.2**). 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 3.13**. Strong intermolecular NOEs are also observed between H1' of G5 and B2 phenyl proton of DB2277 (**Figure 3.3**). 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 3.3**. 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 3.13**). 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 3.2**). 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 3.2**). 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 3.13 and Table 3.2**).

3.5.6 *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 3.13**). For example, H2 protons of A3 and A4 are located in proximity to the phenyl ring attached to the OCH_2 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 3.14**. The heavy atom rmsd value for free DNA in the binding site is 1.7 Å (**Figure 3.14**). 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 3.15**. 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 3.15**). 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 3.16** for free DNA and in **Figure 3.17** 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 3.3**. 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 3.18**. 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.

3.5.7 *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 3.19**. 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 3.20** 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 (ε-ζ > 0°) from the canonical BI conformations (ε-ζ < 0°) 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 3.4**) (36). Calculated ε values were used for rMD simulations for both free DNA and the DB2277-DNA complex. ε-ζ 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 3.21**. Other bases also show little or no variation in their ε-ζ values from the standard BI state. Values for all dinucleotide steps fluctuate around -90° except the C2pA3 step that shows more variation in ε-ζ values (**Figure 3.21**) but remains in the BI region.

3.6 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 3.22A and 3.18**).

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 3.10, 3.11 and 3.22** and the schematic in **Figure 3.12** 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 aza BI –NH and the keto group of C18 in the minor groove based on the model shown in **Figure 3.10**. The downfield shift of the aza 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 3.22C**, 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 3.22B** 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 3.22B**). 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-CH_2– 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 3.22C**) (21). These DB2277-DNA interactions coupled to the dynamic stabilizing hydration network provide a very favorable complex for specific DNA recognition.

3.7 Conclusion

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 aza 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 indicate 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.

3.8 Experimental procedures

3.8.1 *Sample Preparation*

HPLC-purified DNA hairpin with the binding site 5'-AAGATA-3' was purchased from Integrated DNA Technologies, Coralville, IA, USA. The DNA sample was dissolved in diluted NaCl solution, and its concentration was measured using the extinction coefficient at 260 nm with a Cary UV/Vis spectrophotometer at room temperature (39). The DNA sample was dialyzed three times against double deionized water by using a Spectra/Por (Spectrum Laboratories, Inc., Rancho Dominguez) dialysis membrane with a cutoff of 1000Da and lyophilized. Denaturing gel electrophoresis and NMR were used to confirm the purity of DNA samples. The heterocyclic diamidine, DB2277, was synthesized as previously reported (13). The stock solution of DB2277 was prepared in ddH_2O or 99.96% D_2O.

3.8.2 *NMR Experiments*

Spectra were obtained at Bruker Avance 600 MHz NMR equipped with a 5mm QXI probe, a Bruker Avance III HD 600 MHz NMR equipped with triple resonance probe and an Avance III HD 850 MHz NMR with a TCI cryoprobe. DNA concentrations were used around 100 μM for 1D NMR experiments and in the range of 0.5 mM - 0.8 mM for 2D NMR experiments. The NMR experiments involving exchangeable protons were prepared with cacodylate buffer (10 mM cacodylic acid/NaOH at pH 7.0), NaCl (10 mM) ethylene diamine tetra-acetic acid (EDTA), and 4, 4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal standard in 90% H_2O-10% D_2O at pH 6.7 (14). For 2D NMR experiments involving non-exchangeable protons, Tris-d11 buffer (10 mM in 99.996% D_2O) with pH* around 6.3 was used in 99.996% D_2O (40). Water suppression in 1D NMR spectra was obtained using 1-1 pulse sequence. 1D ^{31}P spectrum was obtained on Bruker Avance 500 MHz NMR for both free DNA and DB2277-DNA complex. 85% Phosphoric acid in a capillary was used as an external reference.

3.8.3 *2D NMR*

Phase-sensitive NOESY spectra were recorded for several mixing times (50 ms, 100 ms, 150 ms, 300 ms), collected with 2048×512 data points in t2 and t1 dimension and a relaxation delay of 3.0. Spectra were recorded at 285 K and 298 K with the spectral width of 11 ppm. NOESY spectra were also recorded in H_2O using 1-1 pulse sequence for water suppression, 150 ms mixing time and spectral width of 24 ppm. ^{1}H, ^{31}P-correlated 2D spectra (HPCOR) were used for the assignment of phosphorus resonances (41). Phase-sensitive TOCSY experiments were collected with 80 ms mixing time, 2048×1024 data points in t2 and t1 and 80 scans per t1 increment along with the spectral width of 11 ppm for the D_2O sample. Phase-sensitive ROESY spectra were also recorded with a 100 ms spin-lock, a 2.5 s relaxation delay and 64 scans. ^{13}C, ^{1}H HSQC experiments

for the DNA-DB2277 complex were recorded for complete resonance assignment. CTNOESY (constant time NOESY) NMR was employed on [(5'-CCAAGCTTGG-3')(5'-CCAAGCTTGG-3')). DNA to calculate epsilon torsional angles for this DNA sequence context. Data were processed with TOPSPIN (3.4 version, Bruker Biospin Inc., Billerica, MA). Spectra were zero-filled during the processing to create a 2K × 2K matrix and apodized with shifted sine bell function (SSB = 2).

3.8.4 *Distance Restraints*

Spectra of free DNA and DNA-DB2277 complex were assigned and integrated using SPARKY version 3.114 (42). Integration was performed using the Gaussian function with data used above the lowest displayed contour level for well-resolved peaks whereas data points within a selected rectangular box were used for partially-resolved peaks. NOESY data with 100 ms mixing time was used for integration of cross peaks volumes. Cytosine H5-H6 cross peak (2.45 Å), as a reference point, was used to calculate distance restraints from integrated cross peaks. NOEs were converted to interproton distances with the reference peak of known distance using the following equation:

$$r_{ij} = r_{ref} (I_{ij} / I_{ref})^{-1/6}$$

where r_{ij} is inter-proton distance, I_{ij} is NOE cross peak intensity between two protons i and j, r_{ref} is known distance of reference peak (Cytosine H5-H6) and I_{ref} is the intensity of reference peak.

Cross-peaks were classified as strong (< 3.5 Å), medium (< 4.0 Å) and weak (< 5.0 Å). Upper and lower error bounds of ± 25 %, ± 30 %, and ± 35 % respectively were used to derive distance restraints depending on their peak quality. For example, the percentage for upper error bound was increased for partially-overlapped peaks whereas 5.0 Å as an upper bound was used for well-assigned completely overlapped peaks and for long-range distances observed only in 300 ms mixing time NOESY data.

3.8.5 *DB2277 force field parametrization*

The initial geometry of DB2277 was optimized with an ab initio method (DFT/B3LYP theory with the 6-31+G* basis set) using Gaussian 09 (Gaussian, Inc., 2009, Wallingford, CT) with Gauss-view 5.09 (43,44). Electrostatic potential calculations were performed for the DB2277 molecule using the DFT/B3LYP theory with the 6-31+G* basis set (45). Partial charges were derived using the RESP fitting method (Restrained Electrostatic potential) (46,47). Charge fitting to the electrostatic potential was calculated according to the Merz-Singh-Kollman scheme (48).

All the parameters of the DB2277 molecule are required to run molecular dynamic simulations in the AMBER package. ANTECHAMBER, an AMBER tool, was used to generate force field parameter files using GAFF and parm99 force fields (49-51). The specific atom types adapted from the ff99 force field were used to assign atoms of the DB2277 molecule. Force field parameter files for the DB2528 molecule contains most of the parameters derived from the existing set of bonds, angles and dihedrals for the similar atom types in parm99 and GAFF force fields. The specified atom types used for the DB2277 molecule to generate the force field parameters are shown in **Figure 3.23**. The equilibrium bond length for CK-CA and the equilibrium bond angle for CA-CA-N2 were taken from the optimized geometry of DB2277 from Gaussian 09. Improper dihedrals were taken from the parm99 force field. Six torsional angles of DB2277 which are crucial in defining its conformation and its interaction with DNA, are as follows;

1) N2-CA-CA-CA
2) CA-CA-CT-H1 Need new parameters
   CA-CA-CT-OS Need new parameters
3) CA-CT-OS-CA Need new parameters
4) CT-OS-CA-CA Need new parameters

CT-OS-CA-NC Need new parameters

5) N*-CK-CA-CA Need new parameters

6) NB-CK-CA-CA

The values of torsional parameters for N2-CA-CA-CA and NB-CK-CA-CA dihedrals are already defined in previously reported parametrized data (52,53). The remaining dihedral angles (CA-CA-CT-OS, CA-CT-OS-CA, CT-OS-CA-NC and N*-CK-CA-CA) (not defined in GAFF/Parm99 force field) were re-parametrized. These dihedral angles specifically define the rotation around bonds connecting the phenyl group and aza-benzimidazole moiety. Ab-initio calculations were performed using DFT/6-31+G (d) level of theory in Spartan 16 for defining the required parameters (54). Potential energy profiles for dihedral angles were obtained using constrained equilibrium calculations with dihedral angles varying from 0 to 180 degrees (52,53). The curves were fitted with the least square fitting in kaleidaGraph software using the following cosine function equation:

$$\sum Vn/2\ [1 + \cos(n\Phi - \gamma)].$$

where Vn is the torsional barrier, n is periodicity (0,1,2,3,4 or 6), and γ is phase angle. The frcmod file (force field modification) modified with these newly parametrized dihedral angles of DB2277 is shown in **Table 3.5**.

3.8.6 *Molecular Dynamics (MD) Simulations*

For molecular dynamic (MD) calculations the hairpin loop was not used and calculations were done on a double strand DNA. Hairpin loop doesn't interfere in the complex formation (14). MD simulations were performed with Assisted Model Building with Energy Refinement (AMBER 14) (55). Canonical B-form dsDNA was built using a Nucleic Acid Builder (NAB) tool in the AMBER program (56).

The initial structure of the DNA-DB2277 complex was obtained using Autodock vina software (57). To perform simulations in explicit solvation conditions, free DNA and the DNA-DB2277 complex were solvated in a truncated octahedron periodic box filled with TIP3P water. The box wall was extended to 10.0 Å distance from all the DNA or DNA-DB2277 complex atoms in each dimension. Counterions (Na^+) were added to neutralize the system.

The system was equilibrated with an AMBER14 package using "ff14SB" and ff14SB/chi/ez (i.e., parmbsc0_chiOL4_ezOL1) force field modifications for DNA (39,58). During the equilibration, the SHAKE algorithm was used to constrain all bonds involving hydrogen atoms. The Particle Mesh Ewald method with a 10.0 Å cutoff was used for electrostatic interactions. The ratio of peak heights of attenuated and reference spectrum obtained from CTNOESY spectra was used to obtain epsilon torsional values for ds[(5'-CCAAGCTTGG-3')(5'-CCAAGCTTGG-3')]. DNA (36).

NOE restraints were applied as a potential with a flat bottom and parabolic sides with varying force constants as shown in the energy penalty plot in the **Figure 3.24** (55). The restraint file also specifies the rk2 and rk3 values, the force constants for the lower and upper bounds, respectively. Standard backbone torsion angles for B-DNA (β, γ and χ) and sugar pucker restraints with a force constant of 50 kcal mol^{-1} $Å^{-2}$ were used (59). Watson-Crick hydrogen bonding restraints were included using standard B-type DNA values and distance restraints from NMR were used with force constant of 20 kcal mol^{-1} $Å^{-2}$ (59). ε torsion angles for both the DNA and DB2277-DNA complex were derived from experimentally determined values for ds[(5'-CCAAGCTTGG-3')(5'-CCAAGCTTGG-3')). DNA. The restraint file containing all the above-mentioned restraints was used for further molecular dynamic simulations. In the first step of equilibration, the system was relaxed with 1000 steps of energy minimization using restraint file

for the DNA and DB2277-DNA complex. In the next step, the system was heated from 0 K to 300 K over 10 ps using restraint file under constant-volume conditions. A production run on the system was subsequently performed for 4 ns for DNA and 2 ns for the DB2277-DNA complex in the NPT conditions (isothermal, isobaric ensemble) using restraints. The ensemble of final structures for DNA and the DB2277-DNA complex was obtained from every 10 ps for the last 100 ps of rMD calculations and subsequently minimized.

The free MD simulation run in the explicit solvent was then performed without restraints for 200 ns for DNA and the DB2277-DNA complex on the PMEMD CUDA module of AMBER14. Analysis of helicoidal parameters of DNA was done using a CURVES+ program and cpptraj module (60,61). VMD 1.9 and UCSF Chimera molecular visualization software were used for graphics (62-64).

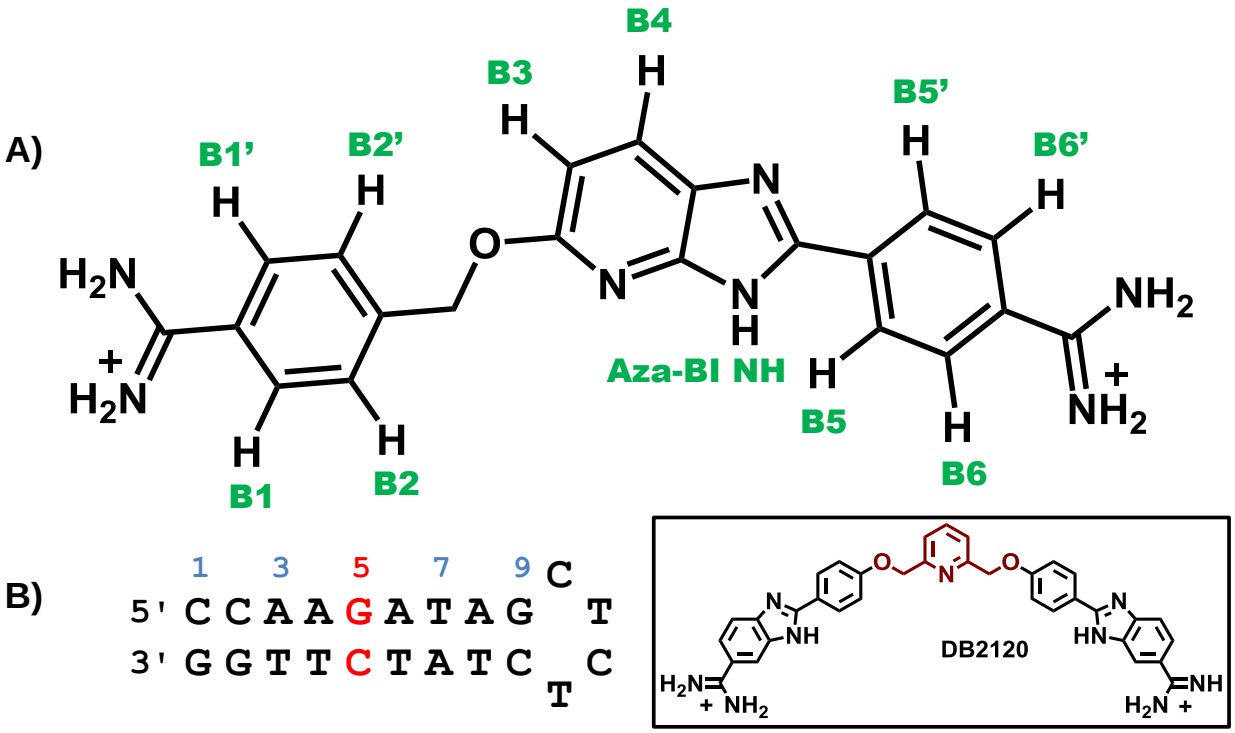


Figure 3.1 A) The molecular structure of DB2277, with numbering scheme for aromatic protons. B) DNA hairpin sequence used in the study.

The structure of DB2120 is shown in the inset.

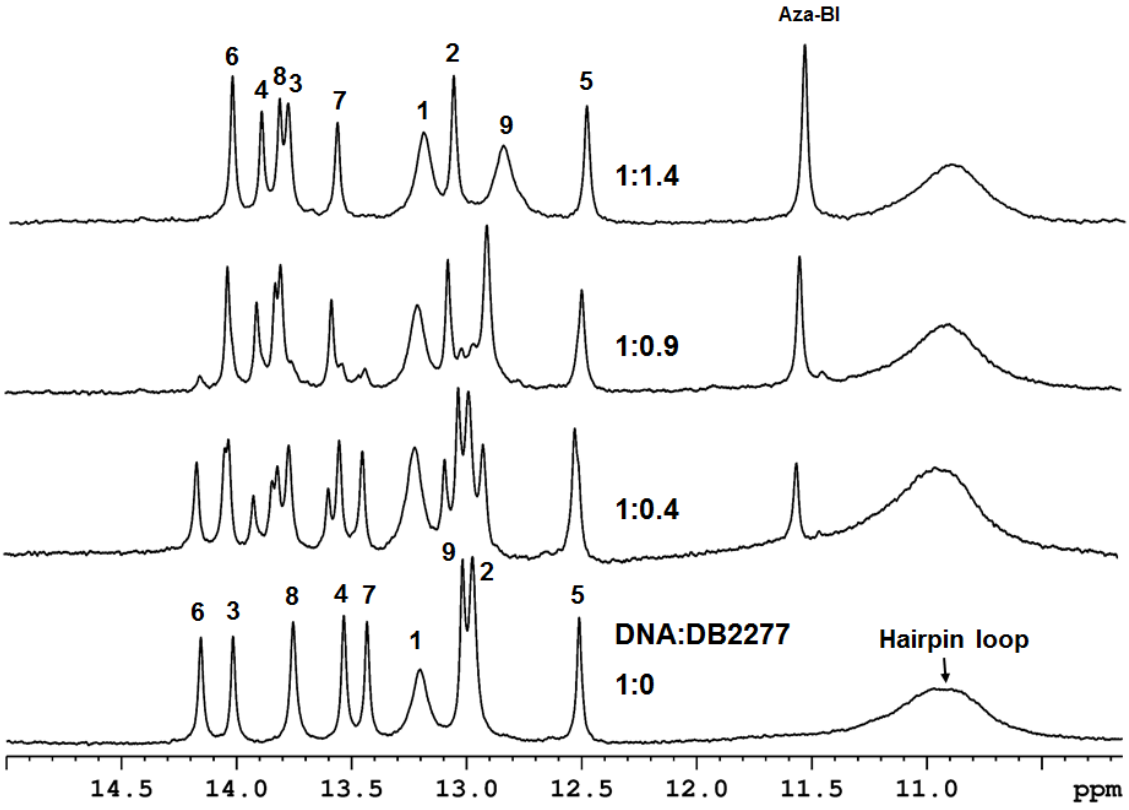


Figure 3.2 Imino proton spectra of DB2277 binding with DNA hairpin at several DNA:DB2277 binding ratios at 285 K.
Assignment of the imino proton resonances was done using 2D NOESY.

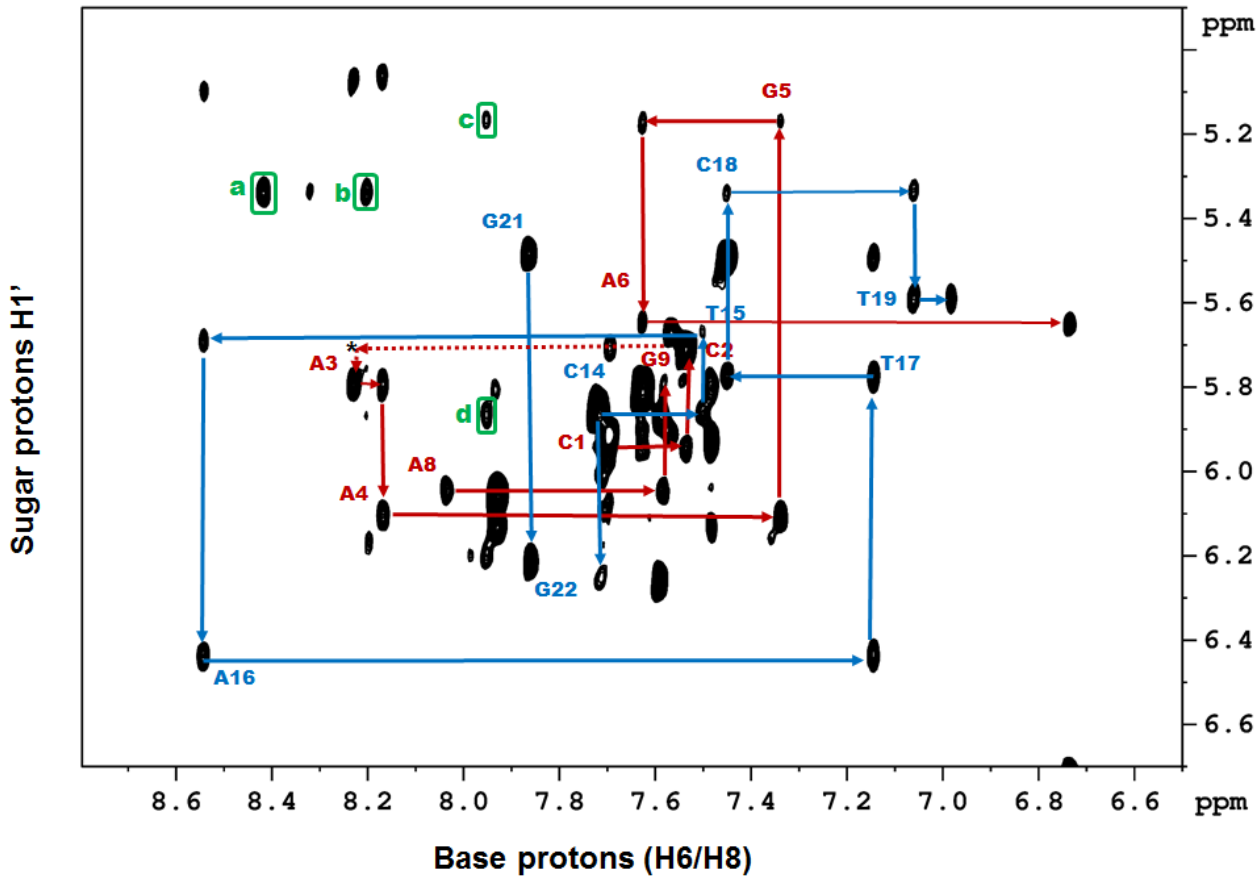


Figure 3.3 NOESY spectrum of DB2277:DNA complex at 1:1 binding ratio at 285 K.

Spectrum shows 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, CH_2O - B2. Missing peak (*) is connected through broken line. Additional peaks are assigned further in ***Figure 3.24****.*

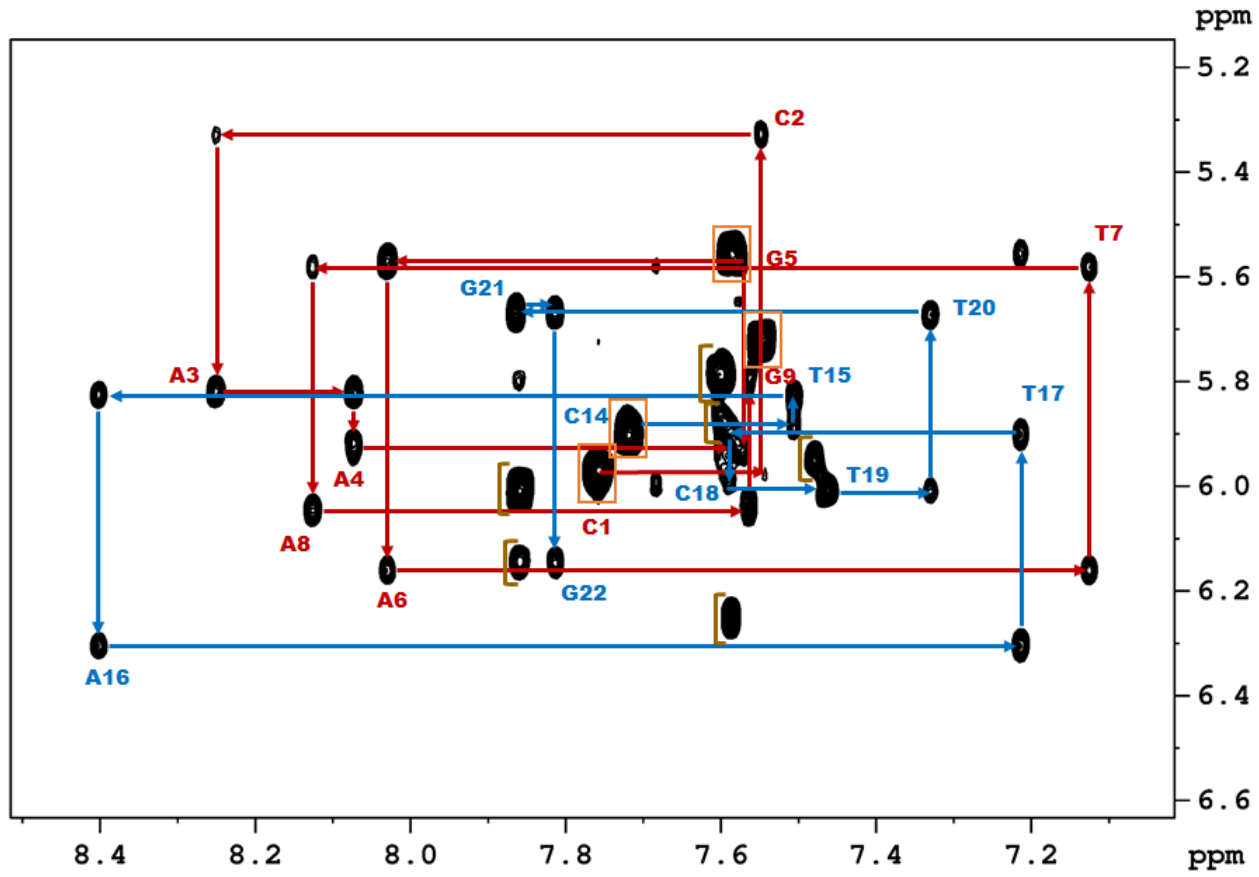


Figure 3.4 NOESY spectrum of a free DNA hairpin at 285 K.
Spectrum shows sequential walk through base protons to H1' sugar protons. Additional peaks marked are assigned as follows: Cytosine H5-H6 (orange box), loop region (highlighted in brown).

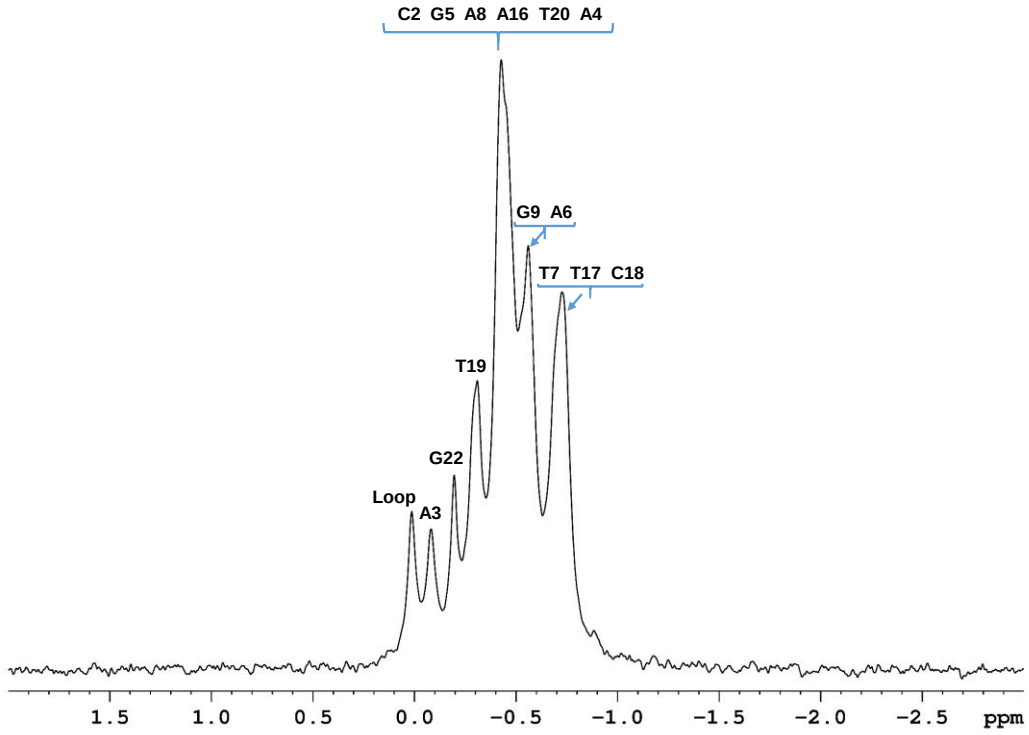


Figure 3.5 1D proton-decoupled ^{31}P spectrum with phosphorus assignments for free DNA at 285 K.

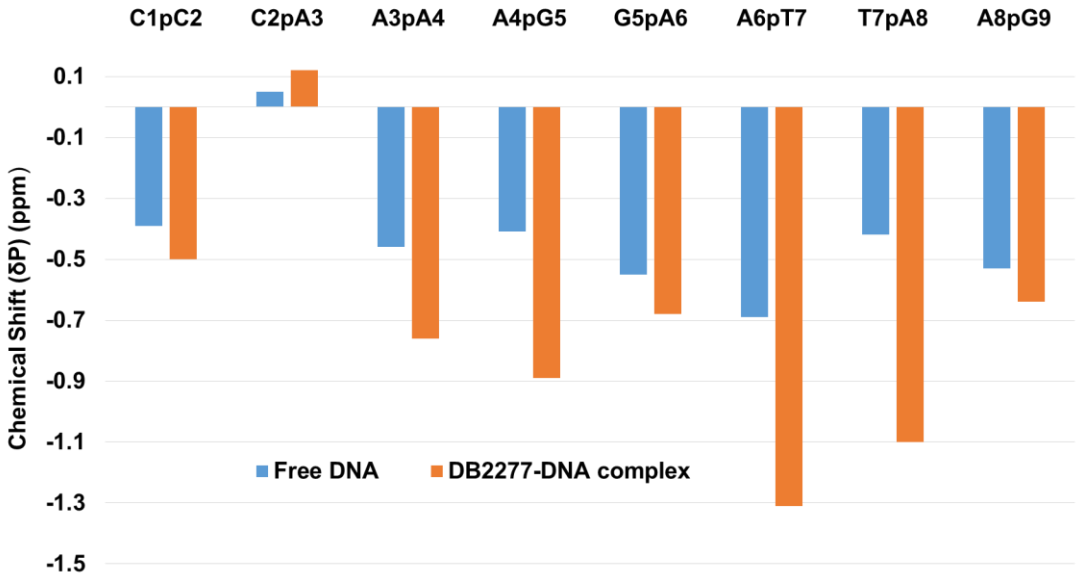


Figure 3.6 The plot of chemical shifts of ^{31}P (δP in ppm) for both free DNA and the DB2277-DNA complex.

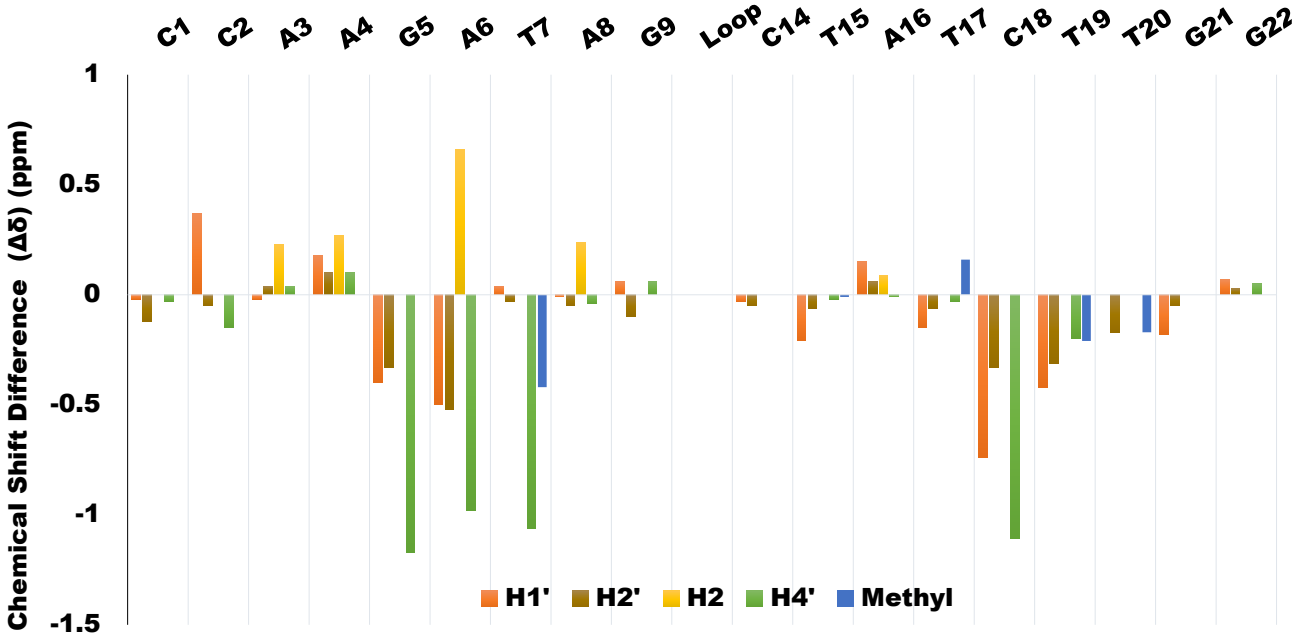


Figure 3.7 Chemical shift differences (Δδ in ppm) between free DNA and the DB2277-DNA complex for H1', H2' and H4' sugar protons, H2 protons of adenine and methyl protons of thymine.

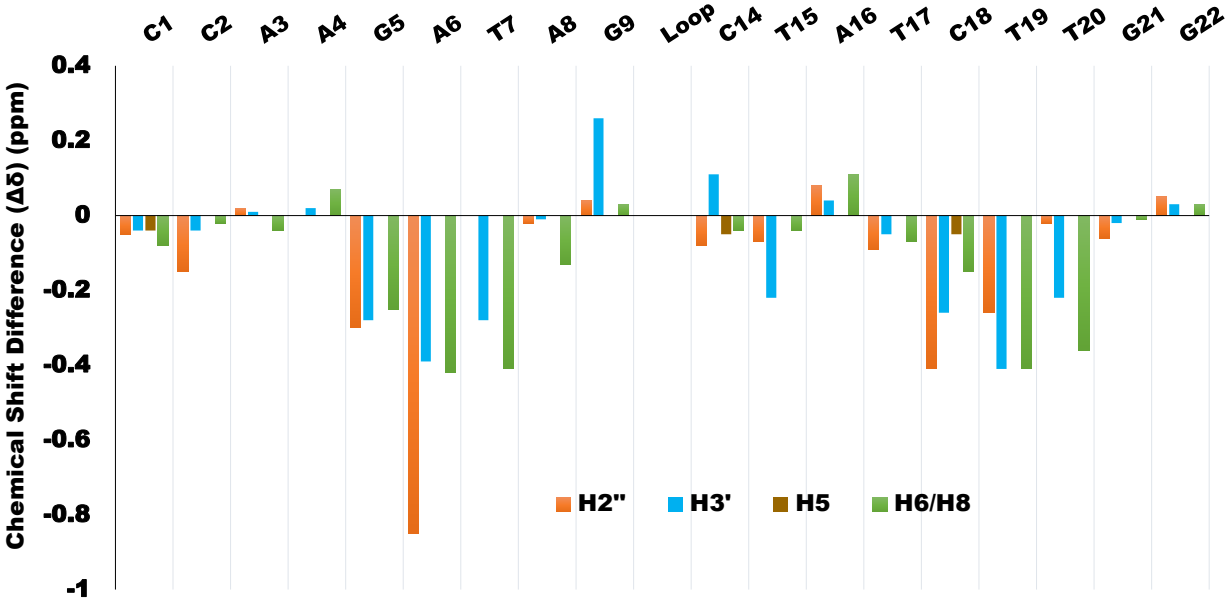


Figure 3.8 Chemical shift differences (Δδ) between free DNA and the DB2277-DNA complex is shown in the chart for H2'' and H3' sugar protons, H5 of cytosine, H6/H8 base protons of the DNA.

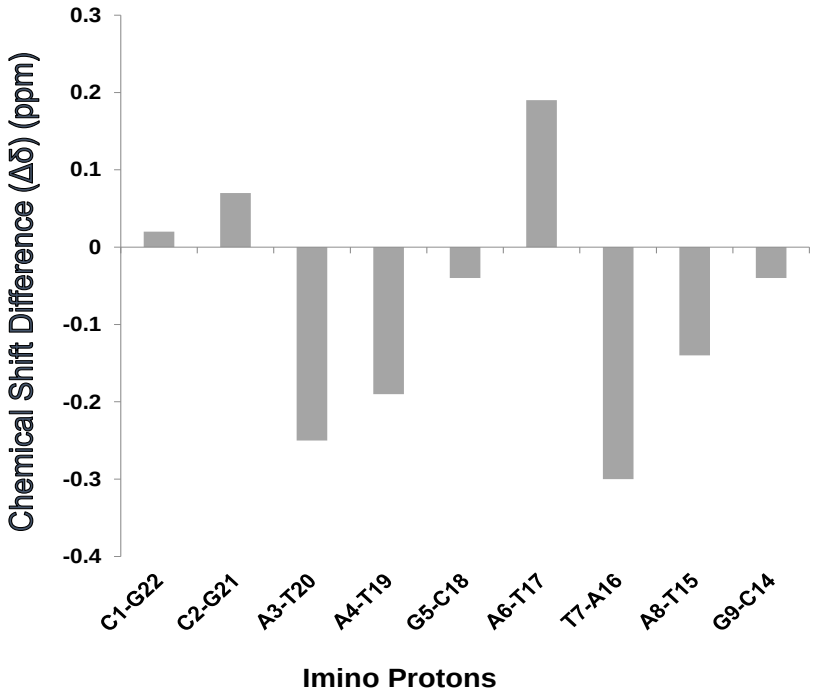


Figure 3.9 Δδ between free DNA and the DB2277-DNA complex for imino protons of the DNA base pairs.

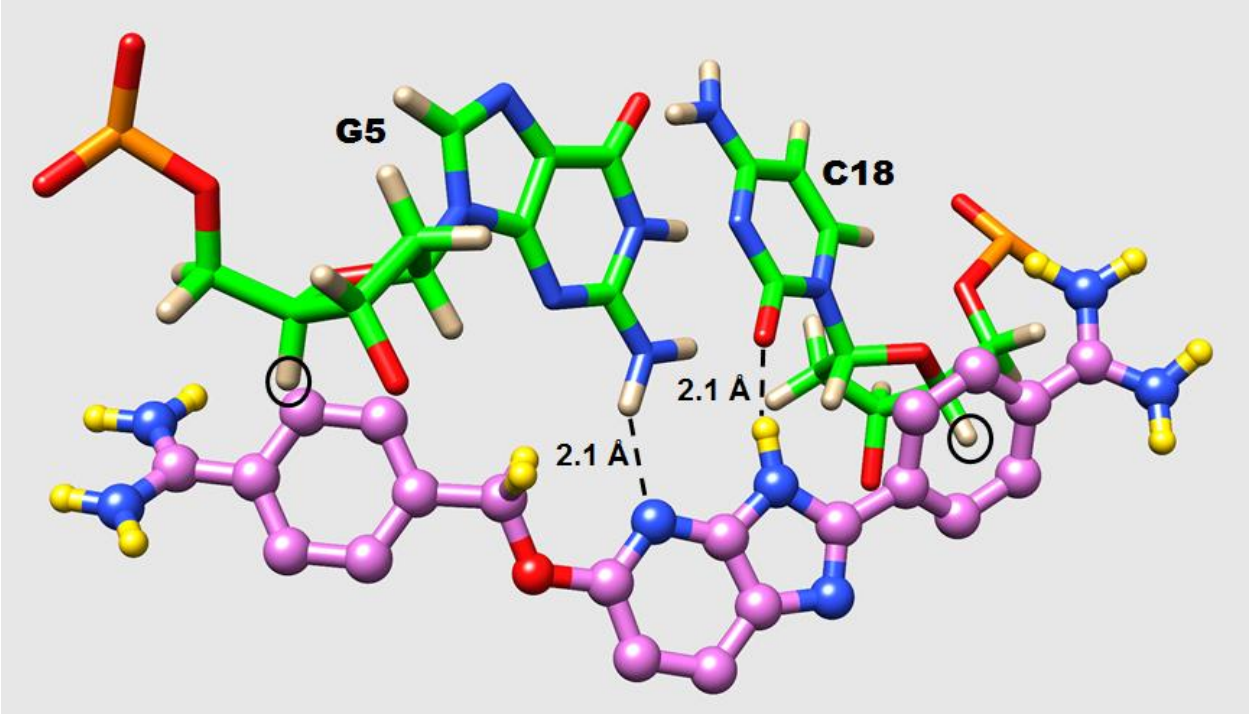


Figure 3.10 Specific interactions of G5 -NH_2 with aza BI -N and C18-O2 with aza BI -NH moiety of DB2277.
Ring current effects on H4' protons of G5 bp in relation to phenyl rings of DB2277 (black circle).

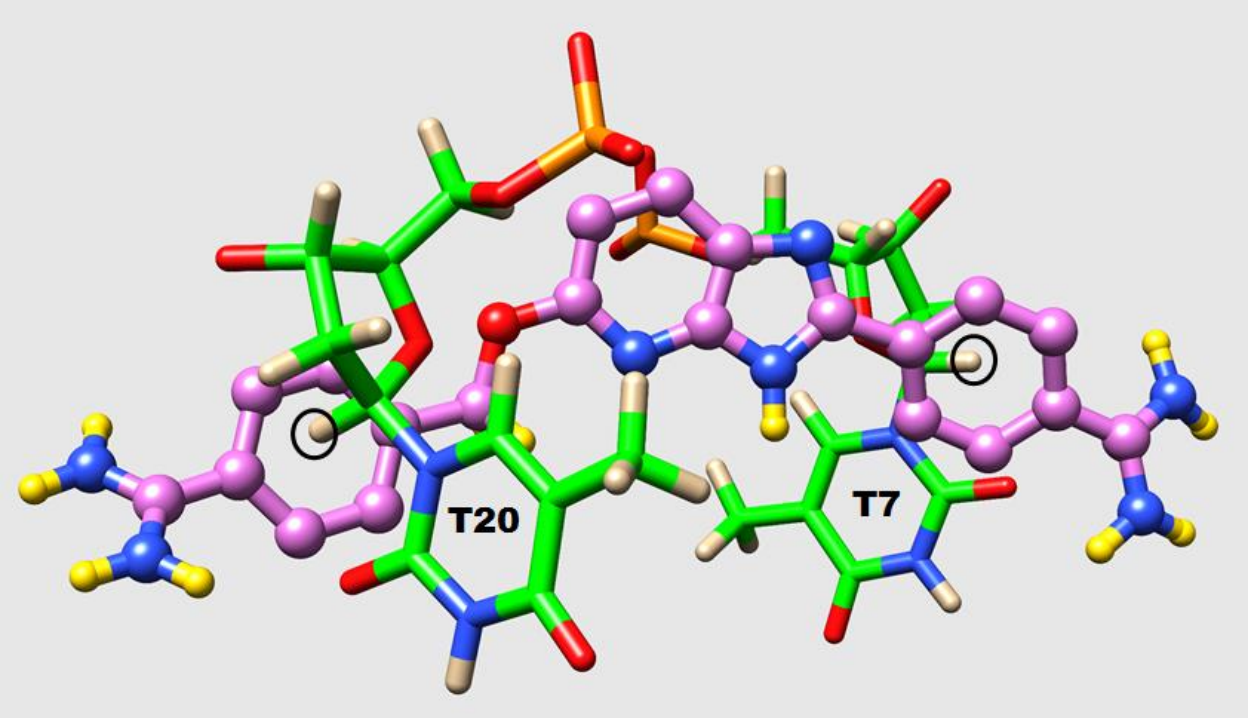


Figure 3.11 Ring current effects due to the mutual interaction of H1' protons of T7 and T20 with phenyl rings of DB2277 (black circle).
DB2277 is in ball and stick (C in magenta) and DNA is shown in stick representation (C in green).

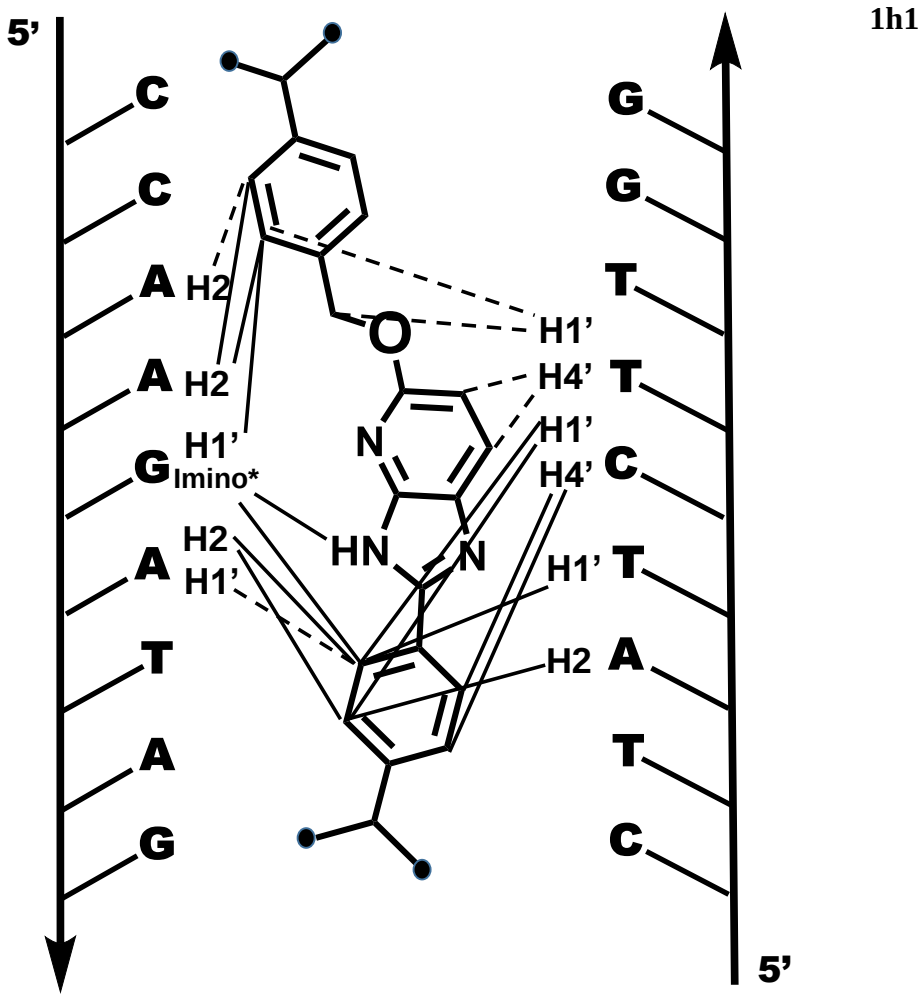


Figure 3.12 Schematic model of interactions of DB2277 protons with DNA protons at the binding site.

*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% H_2O/10% D_2O NOESY for exchangeable protons.*

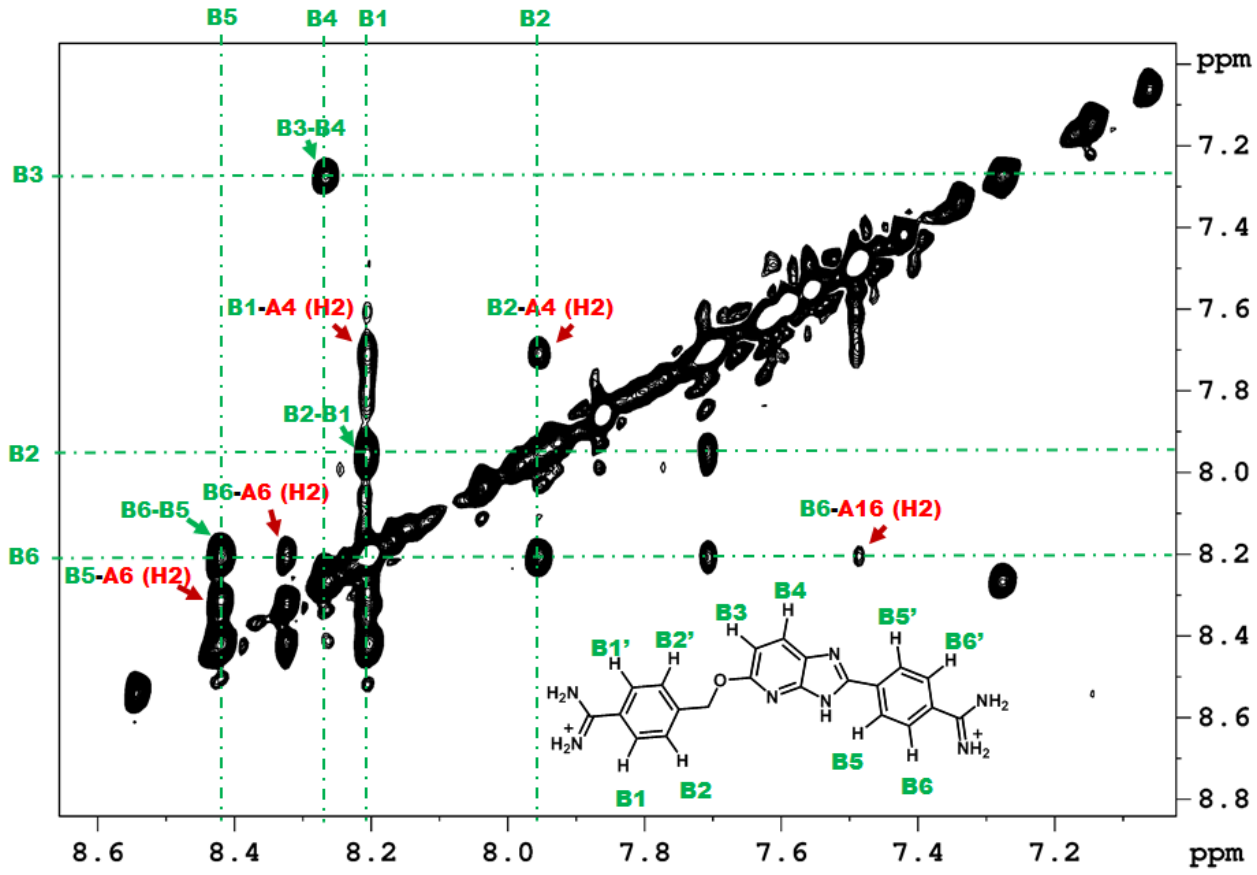


Figure 3.13 NOESY spectrum of the DB2277:DNA complex at 1:1 binding ratio at 285 K. *Assignments of intramolecular DB2277 contacts are marked with green arrows whereas intermolecular NOEs of DB2277 with adenine of DNA are assigned and marked with red arrows. DB2277 protons are shown in green and H2 protons of adenine are shown in red.*

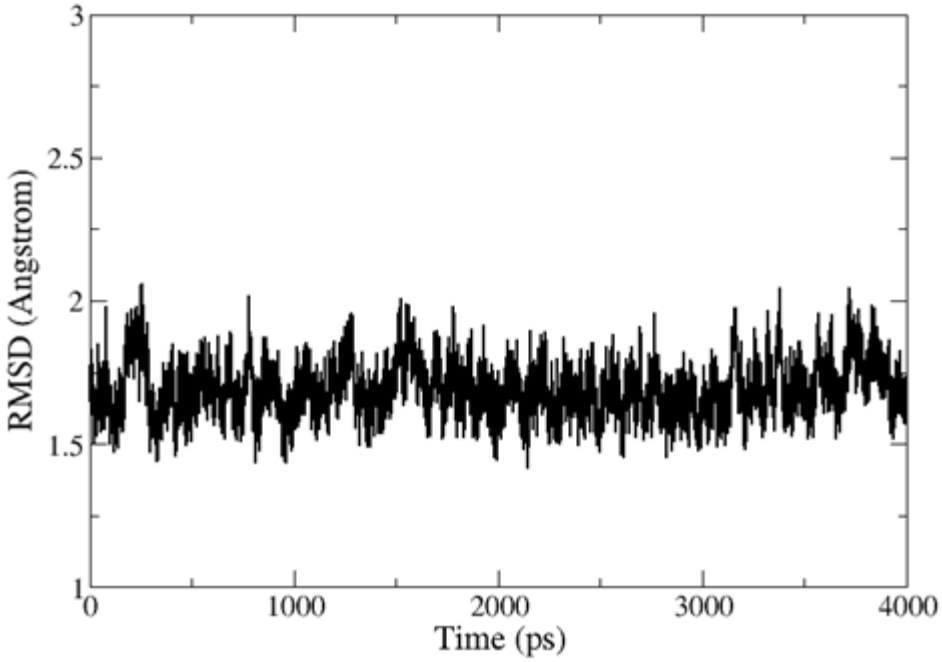


Figure 3.14 An rmsd plot for heavy atoms of free DNA in binding site –AAGATA–
Fitting applied to the average structure for 4 ns rMD.

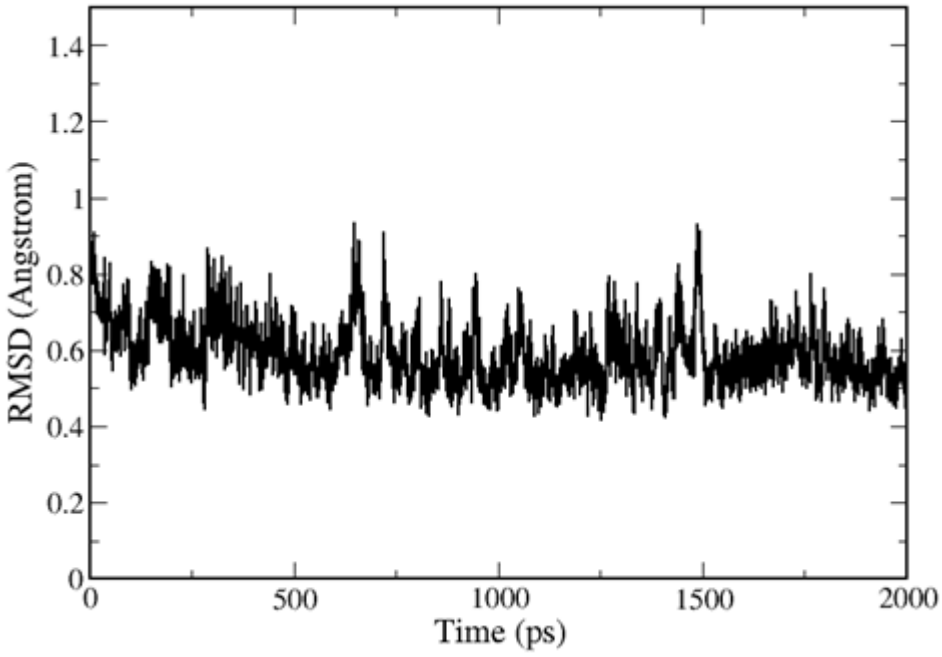


Figure 3.15 An rmsd plot for heavy atoms of the DB2277-DNA complex for the binding site –AAGATA–
Fitting applied to the average structure for 2 ns rMD.

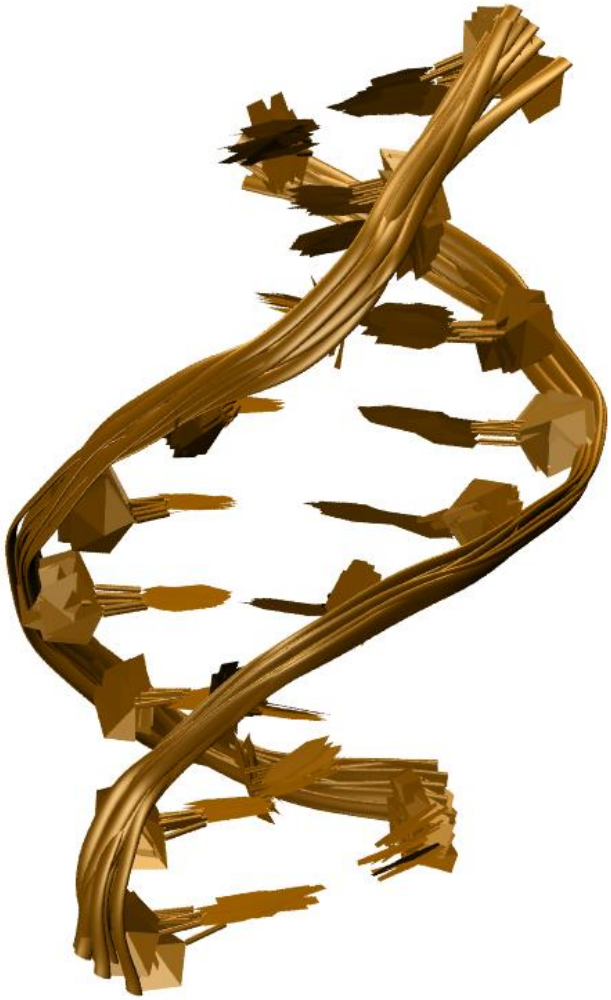


Figure 3.16 NMR structure ensemble of free DNA.
Snapshots were taken every 10 ps for the last 100 ps of rMD simulation and subsequently minimized. DNA is represented in new ribbon style (Ochre color). Hydrogen atoms and water molecules are omitted for clarity.

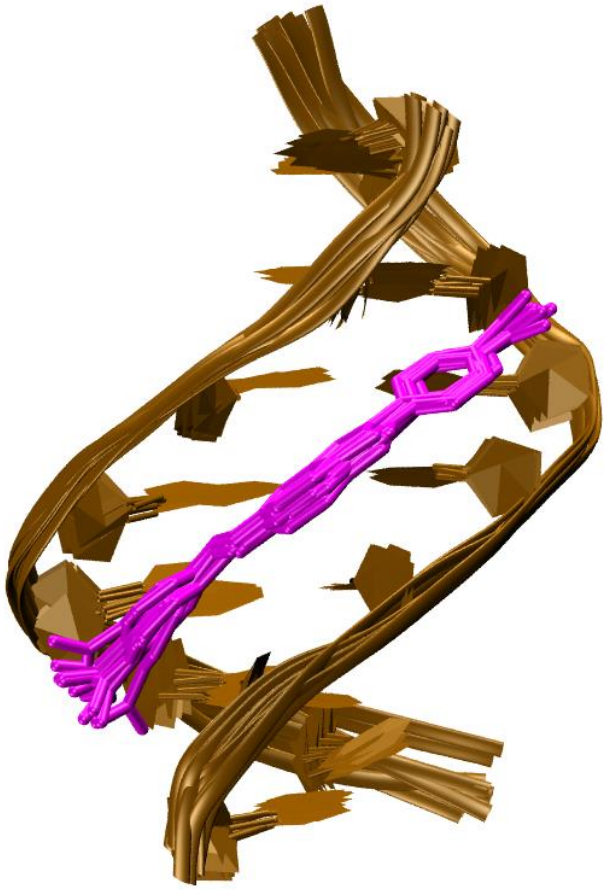


Figure 3.17 NMR structure ensemble of the DB2277-DNA complex.
Snapshots were taken every 10 ps for the last 100 ps of rMD simulation and subsequently minimized. DNA is represented in new ribbon style (ochre) and DB2277 is shown in licorice (magenta). Hydrogen atoms, terminal bp and water molecules are omitted for the sake of clarity.

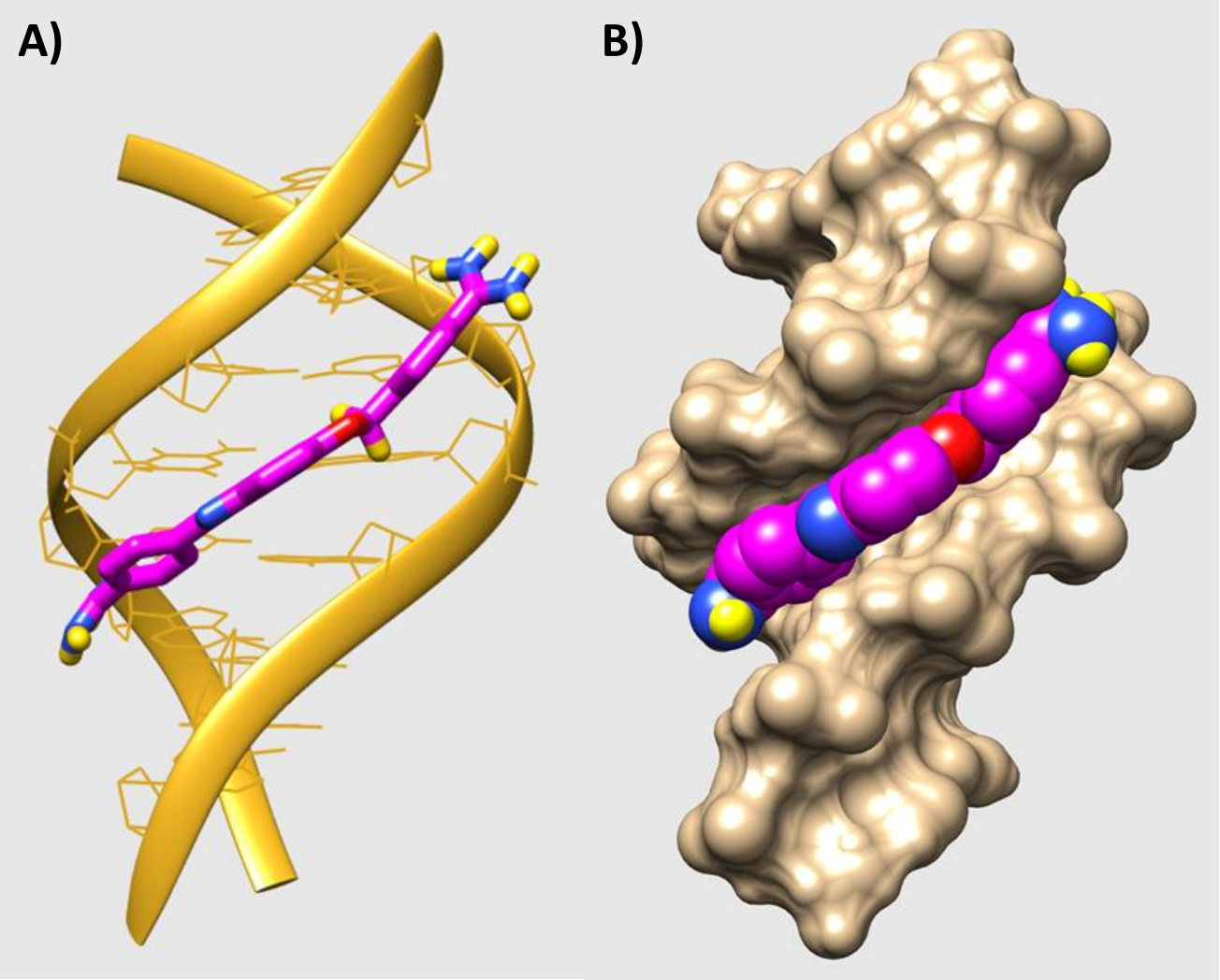


Figure 3.18 A) Minor groove view of final structure of the DB2277-DNA complex. B) Space-fill representation of the DB2277-DNA complex.
The DNA backbone is shown in ribbon style whereas the sugar and base region is shown in wire representation (gold color). DB2277 is shown in licorice (yellow for H, magenta for C, red for O and blue for N). Hydrogen atoms of DNA and water molecules are omitted for clarity.

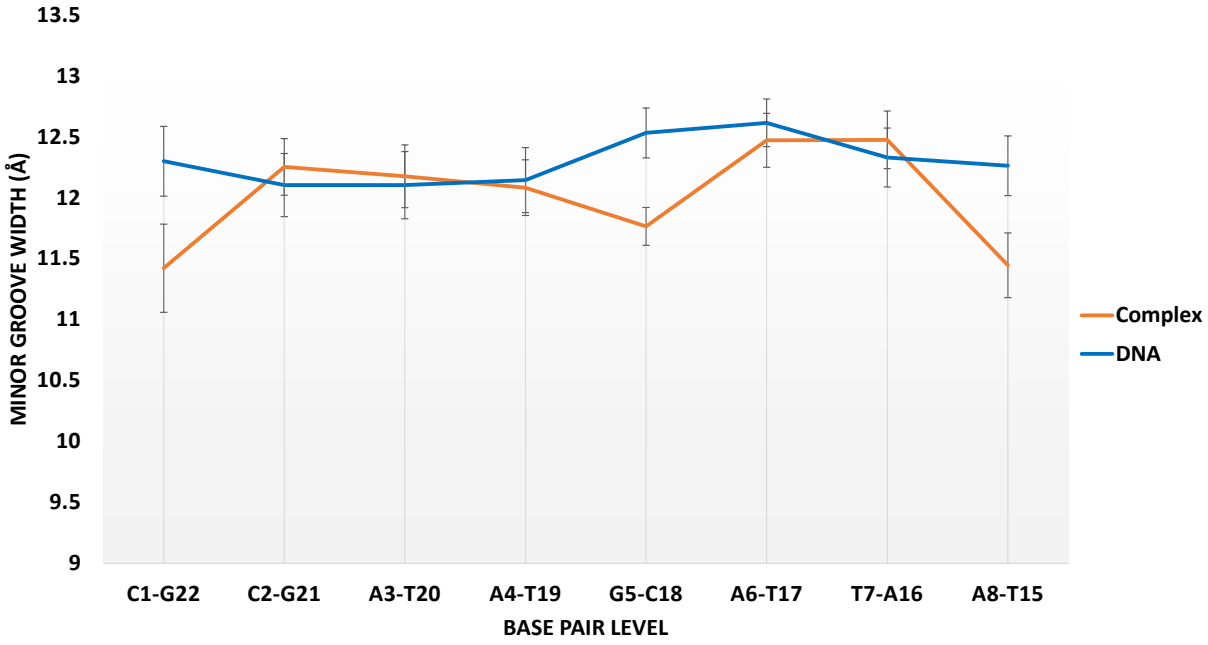


Figure 3.19 Minor groove widths of free DNA and the DB2277-DNA complex.
Free DNA (blue) and the DB2277-DNA complex (orange). Standard deviation calculated from last 100 ps trajectory of rMD.

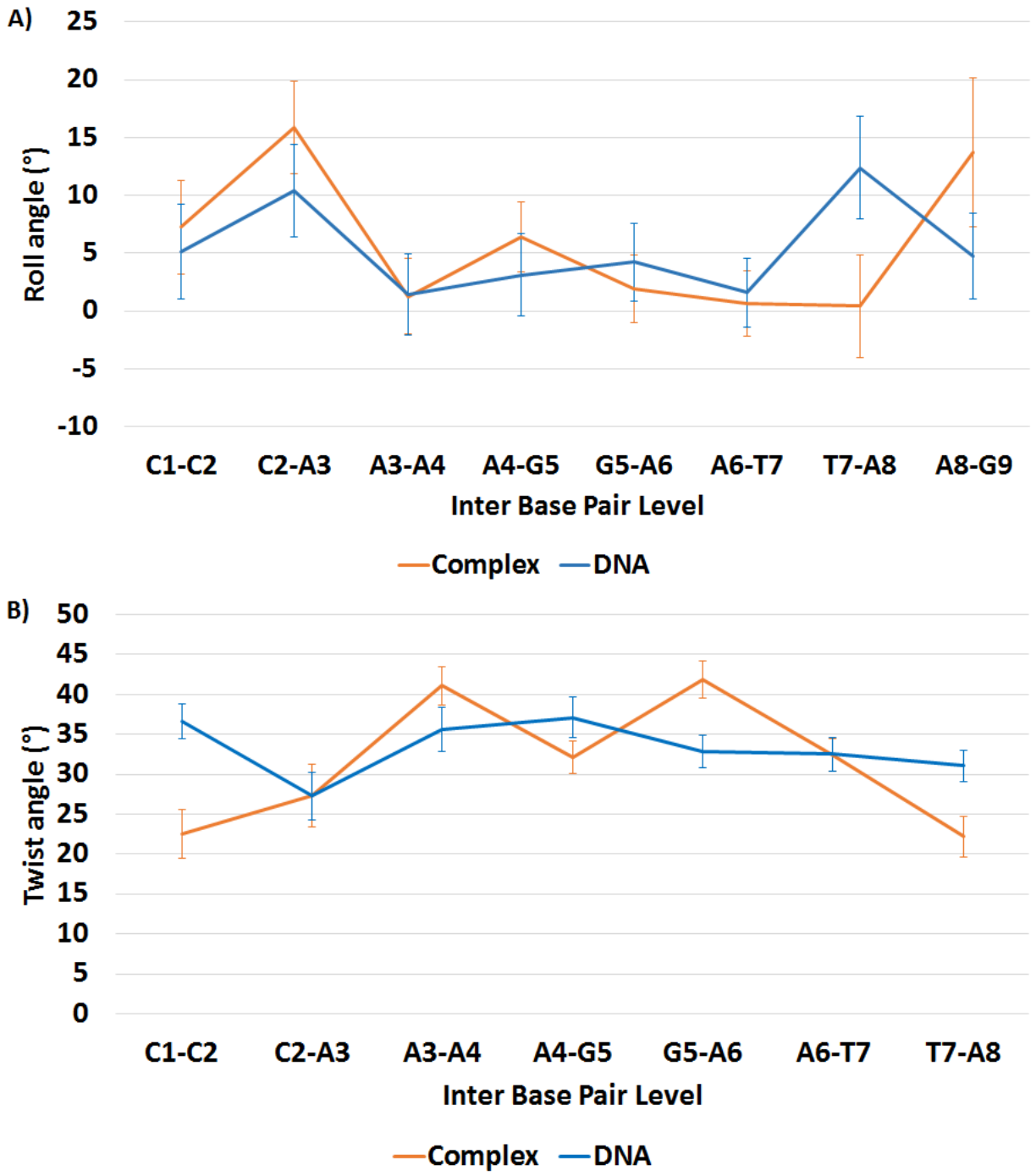


Figure 3.20 A) Roll angles for free DNA and the DB2277-DNA complex. B) Twist angles for free DNA and the DB2277-DNA complex.

Free DNA (blue) and the DB2277-DNA complex (orange). Standard deviations are calculated over all frames (each frame of 0.5 ps) for the last 100 ps trajectory of rMD.

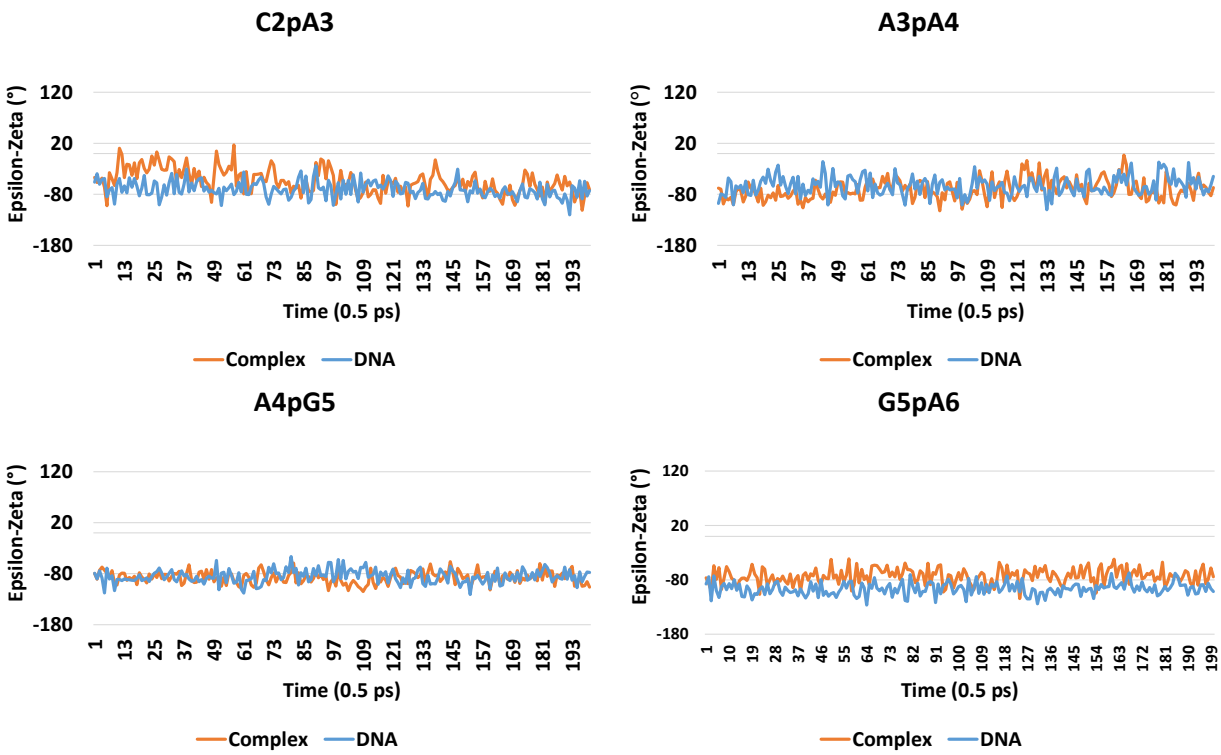


Figure 3.21 Plots for ε-ζ dihedral angles vs time frames for the last 100 ps of the rMD simulation for the free DNA and DB2277-DNA complex.
Free DNA (blue) and the DB2277-DNA complex (orange).

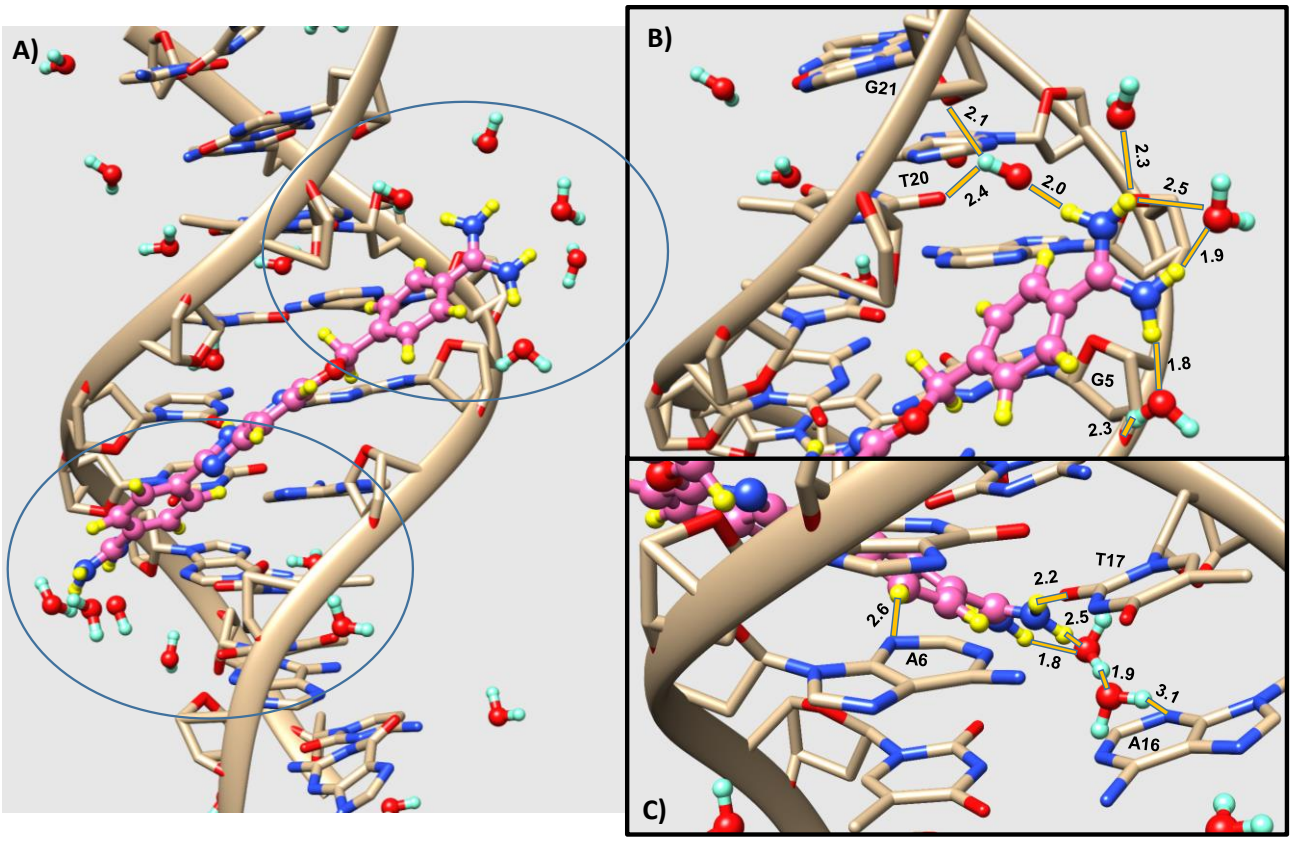


Figure 3.22 The rMD model with water molecules.
A) Minor groove view of the DB2277-DNA complex with proximal water molecules. Inset: B) Dynamic water network with amidine-NH and an interfacial water molecule that forms an H-bond bridge between amidine-NH and T20-O2 at the floor of the minor groove (Upper 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.

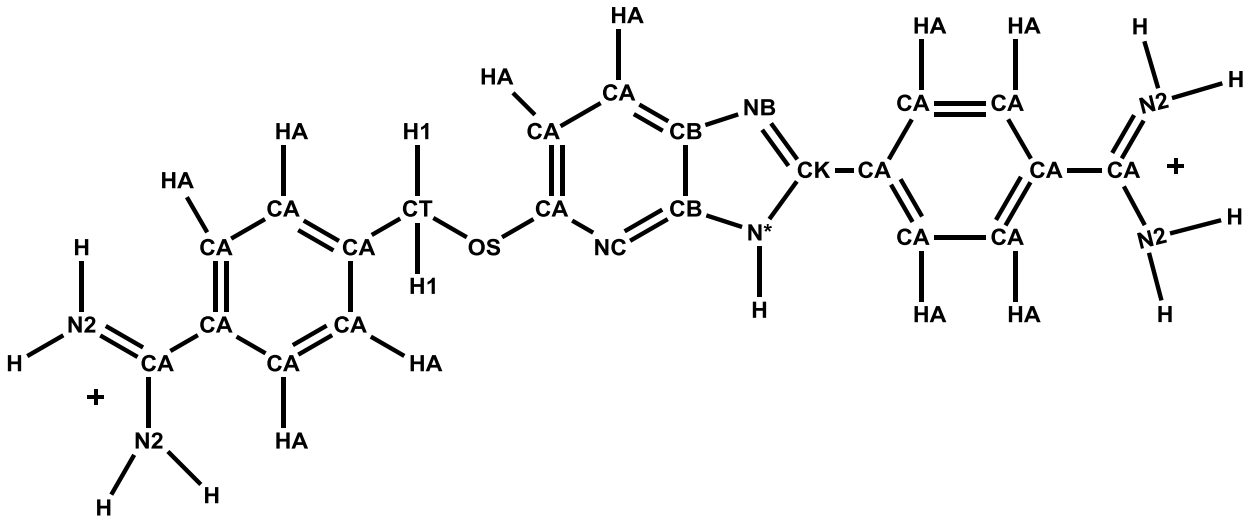


Figure 3.23 Molecular structure with specific atom types used for the DB2277 molecule.

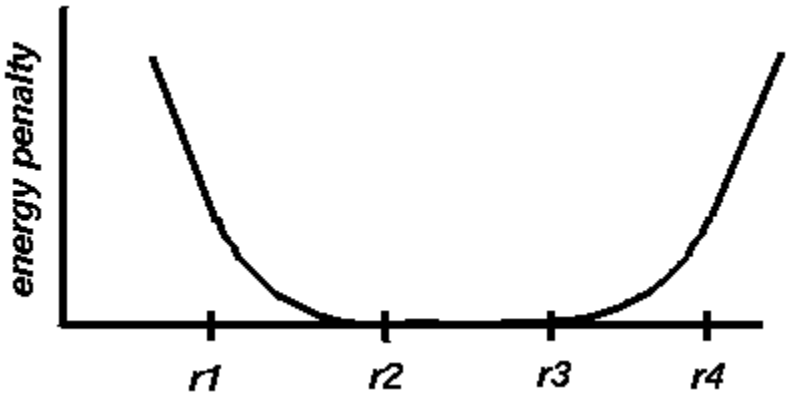


Figure 3.24 Energy penalty plot.
The violation energy is a well with a square bottom between r2 *and* r3*, with parabolic sides out to a defined distance (*r1 *and* r4 *for lower and upper bounds, respectively), and linear sides beyond that distance (55).*

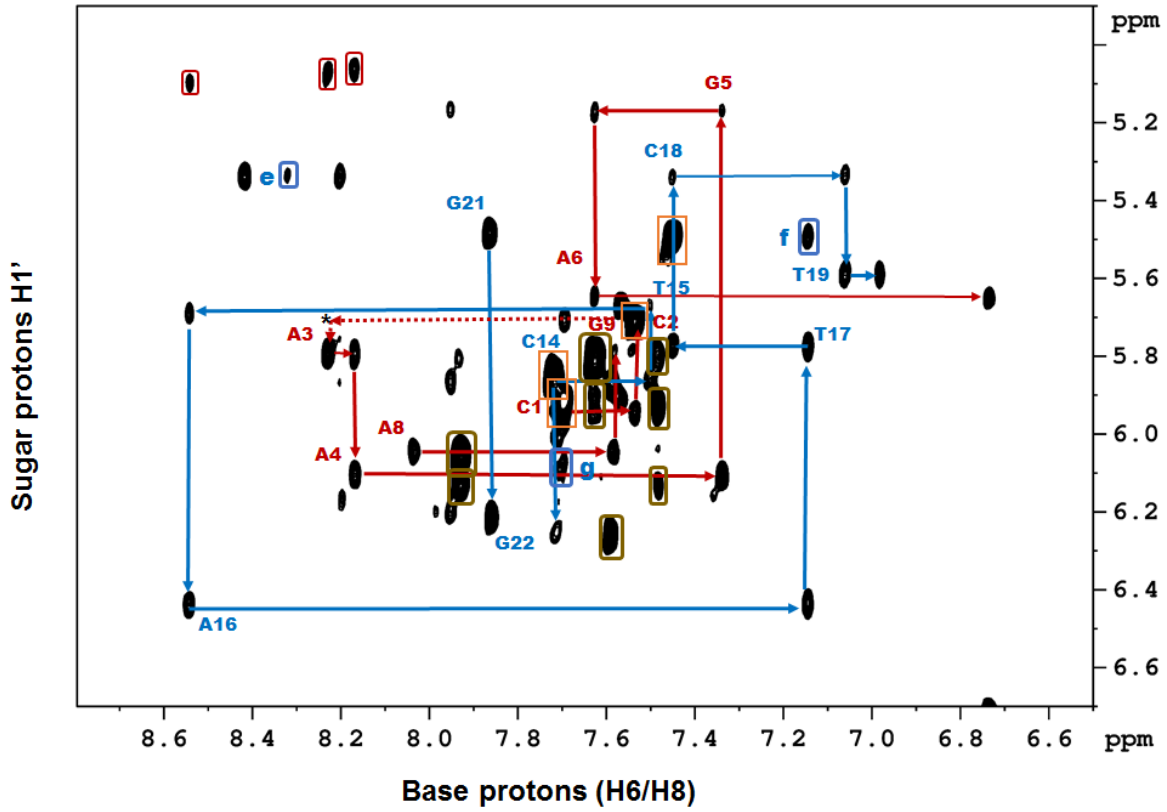


Figure 3.25 NOESY spectrum of the DB2277:DNA complex at 1:1 binding ratio at 285 K.
The spectrum shows sequential walk through base protons to H1' sugar protons (solid lines). Additional peaks marked are assigned as follows: Cytosine H5-H6 in orange box, loop region marked in brown. Remaining peaks marked in blue are assigned as e) C18 (H1') - A6 (H2), f) C18 (H5) - T17 (H6), g) A4 (H1') - A4 (H2). Peaks from sugar H3' protons marked in red boxes. Missing peak (*) is connected through broken lines.

Table 3.1 Chemical shifts (in ppm) of free DB2277 protons and DB2277 protons bound to DNA.

DB2277 protons	Free (ppm)	Bound (ppm)
B1/B1'	**7.77**	**8.21**
B2/B2'	**7.68**	**7.96**
B3	**6.86**	**7.28**
B4	**7.95**	**8.27**
B5/B5'	**8.11**	**8.41**
B6/B6'	**7.87**	**8.20**
OCH_2	**5.50**	**5.86**

Table 3.2 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% H_2O/10% D_2O.* [b] *Weak NOEs or NOEs visible only at long mixing times.*

Table 3.3 NMR refinement statistics for free DNA and the DB2277-DNA complex.

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*

Structure statistics calculated from final converged 10 snapshots taken every 10 ps for last 100 ps of rMD simulation. Experimental details are provided in supplementary materials and methods.

Table 3.4 $J_{H3'-P}$ coupling constant values determined from CTNOESY NMR for duplex DNA ds[(5'-CCAAGCTTGG-3')(5'-CCAAGCTTGG-3')].

Duplex **CCAAGCTTGG**	**$J_{H3'-P}$** **Coupling (Hz)**
C1	**5.7**
C2	**5.4**
A3	**3.9**
A4	**3.7**
G5*	**--**
C6	**3.7**
T7	**3.0**
T8	**4.8**
G9	**4.5**

** Value could not be measured due to extensive overlap with water signal.*

Table 3.5 The frcmod file showing complete parameters for the DB2277 molecule.

```
MASS
N2 14.01         0.530              parm99
H  1.008         0.161              parm99
CA 12.01         0.360              parm99
HA 1.008         0.167              parm99
OS 16.00         0.465              parm99
CB 12.01         0.360              parm99
N* 14.01         0.530              parm99
NB 14.01         0.530              parm99
NC 14.01         0.530              parm99
CK 12.01         0.360              parm99
H1 1.008         0.135              parm99
CT 12.01         0.878              parm99

BOND
CA-CA  469.0    1.400       parm99
CA-CB  469.0    1.404       parm99
CA-CT  317.0    1.510       parm99
CA-NC  483.0    1.339       parm99
CA-HA  367.0    1.080       parm99
CA-N2  481.0    1.340       parm99
CB-CB  520.0    1.370       parm99
CB-NC  461.0    1.354       parm99
CB-N*  436.0    1.374       parm99
CB-NB  414.0    1.391       parm99
CK-N*  440.0    1.371       parm99
CK-NB  529.0    1.304       parm99
CK-CA  469.0    1.460       force const:parm99,length;CA-CB:Gaussian
H -N2  434.0    1.010       parm99
CT-H1  340.0    1.090       parm99
CT-OS  320.0    1.410       parm99
CA-OS  372.4    1.3730      gaff, ca-os

ANGLE
N2-CA-N2    70.0      120.00    parm99
H -N2-H     35.0      120.00    parm99
CA-CA-N2    70.0      119.99    force const: parm99,CM-CA-N2:Gaussian
CA-N2-H     50.0      120.00    parm99
CA-CA-CA    63.0      120.00    parm99
CA-CA-HA    50.0      120.00    parm99
CA-CA-CT    70.0      120.00    parm99
CA-CT-H1    50.0      109.50    parm99, CA-CT-HC
H1-CT-H1    35.0      109.50    parm99
H1-CT-OS    50.0      109.50    parm99
```

```
CA-CT-OS    60.0      109.50    parm99, C -CT-OS
CT-OS-CA    62.3      119.76    force const: gaff,c3-os-ca:Gaussian
CA-CA-OS    70.5      115.46    force const: gaff,ce-c2-os:gaussian
NC-CA-OS    73.8      119.56    force const gaff, nd-cc-os:gaussian
CA-CA-NC    69.2      122.63    gaff, ca-ca-nb
CA-CA-CK    63.0      120.00    parm99, CA-CA-CN
CA-NC-CB    68.6      115.86    gaff, ca-nb-ca
CB-CB-NC    70.0      127.70    parm99
CA-CB-CB    63.0      117.30    parm99
CA-CB-NB    70.0      132.40    parm99
CB-CB-NB    70.0      110.40    parm99
CB-CB-N*    70.0      106.20    parm99
N*-CB-NC    70.0      126.20    parm99
CB-N*-CK    70.0      105.40    parm99
CB-N*-H     50.0      125.80    parm99
CK-N*-H     50.0      128.80    parm99
CB-NB-CK    70.0      103.80    parm99
N*-CK-NB    70.0      113.90    parm99
CA-CK-NB    70.0      120.00    parm99, CC-CV-NB
CA-CK-N*    70.0      120.00    parm99, CC-CW-NA

DIHE
N2-CA-N2-H     4    9.60        180.0          2.0    parm 99, X -CA-N2-X
CA-CA-N2-H     4    9.60        180.0          2.0    parm 99, X -CA-N2-X
N2-CA-CA-CA    4    0.789       327.000       -4.0    Cheatham
N2-CA-CA-CA    4   -3.118         0.000       -2.0    Cheatham
N2-CA-CA-CA    4    0.609        90.000        1.0    Cheatham
CA-CA-CA-CA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CA-CA-CA-CT    4   14.50        180.0          2.0    parm99
CA-CA-CT-H1    6    0.72        180.0          2.0    New parameter
CA-CA-CT-OS    6    0.72        180.0          2.0    New parameter
CA-CT-OS-CA    3    6.88          0.0          1.0    New parameter
CT-OS-CA-CA    2    3.29        180.000        2.000  New parameter
CT-OS-CA-NC    2    3.29        180.000        2.000  New parameter
OS-CA-CA-CA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
OS-CA-NC-CB    2    9.60        180.0          2.0    Parm99, X -CA-NC-X
CA-CA-CA-CB    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CA-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
HA-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
HA-CA-CA-CT    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CB-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
NC-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
OS-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CK-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CA-CA-CA-HA    4   14.50        180.0          2.0    parm99, X -CA-CA-X
CA-CA-CB-CB    4   14.00        180.0          2.0    parm99, X -CA-CB-X
HA-CA-CB-NB    4   14.00        180.0          2.0    parm99, X -CA-CB-X
```

```
HA-CA-CB-CB    4   14.00        180.0          2.0    parm99, X -CA-CB-X
CA-CA-CB-NB    4   14.00        180.0          2.0    parm99, X -CA-CB-X
CB-CB-NC-CA    2    8.30        180.0          2.0    parm99, X -CB-NC-X
N*-CB-NC-CA    2    8.30        180.0          2.0    parm99, X -CB-NC-X
NC-CB-N*-CK    4    6.60        180.0          2.0    parm99, X -CB-N*-X
CB-CB-N*-CK    4    6.60        180.0          2.0    parm99, X -CB-N*-X
CB-CB-N*-H     4    6.60        180.0          2.0    parm99, X -CB-N*-X
NC-CB-N*-H     4    6.60        180.0          2.0    parm99, X -CB-N*-X
CA-CB-CB-NC    4   21.80        180.0          2.0    parm99, X -CB-CB-X
CA-CB-CB-N*    4   21.80        180.0          2.0    parm99, X -CB-CB-X
CA-CB-NB-CK    2    5.10        180.0          2.0    parm99, X -CB-NB-X
N*-CK-NB-CB    2   20.00        180.0          2.0    parm99, X -CK-NB-X
CA-CK-NB-CB    2   20.00        180.0          2.0    parm99, X -CK-NB-X
NB-CK-N*-CB    4    6.80        180.0          2.0    parm99, X -CK-N*-X
CA-CK-N*-CB    4    6.80        180.0          2.0    parm99, X -CK-N*-X
CA-CK-N*-H     4    6.80        180.0          2.0    parm99, X -CK-N*-X
N*-CK-CA-CA    4    3.42        180.0          2.0    New parameter
NB-CK-CA-CA    4   -0.6         180.0         -4.0    DB921
NB-CK-CA-CA    4    3.1         180.0         -2.0    DB921
NB-CK-CA-CA    4   -0.7         360.0          1.0    DB921

IMPROPER
CA-CA-CA-HA          1.1          180.0          2.0          parm99
CA-CA-CA-CA          1.1          180.0          2.0          parm99
CA-N2-CA-N2          1.1          180.0          2.0          parm99
CA-NC-CA-OS          1.1          180.0          2.0          parm99
CB-N*-CB-NC          1.1          180.0          2.0          parm99
CB-CK-N*-H           1.1          180.0          2.0          parm99
CA-N*-CK-NB          1.1          180.0          2.0          parm99
CK-CA-CA-CA          1.1          180.0          2.0          parm99
CA-CA-CA-HA          1.1          180.0          2.0          parm99
CA-CA-CA-HA          1.1          180.0          2.0          parm99
CA-CA-CA-CA          1.1          180.0          2.0          parm99
CA-N2-CA-N2          1.1          180.0          2.0          parm99
CB-CA-CB-NB          1.1          180.0          2.0          parm99
```

3.9 References

1. J. O. Trent, G. R. Clark, A. Kumar, W. D. Wilson, D. W. Boykin, J. E. Hall, R. R. Tidwell, B. L. Blagburn and S. Neidle, *Journal of Medicinal Chemistry* **1996**, *39*, 4554-4562.

2. A. Abu-Daya, P. M. Brown and K. R. Fox, *Nucleic Acids Research* **1995**, *23*, 3385-3392.

3. D. E. Wemmer, *Annual Review of Biophysics & Biomolecular Structure* **2000**, *29*, 439.

4. A. I. Khalaf, A. A. H. Al-Kadhimi and J. H. Ali, *Acta Chimica Slovenica* **2016**, *63*, 689-704.

5. M. R. Conte, T. C. Jenkins and A. N. Lane, *European Journal of Biochemistry* **1995**, *229*, 433-444.

6. F. J. Scott, R. J. O. Nichol, A. I. Khalaf, F. Giordani, K. Gillingwater, S. Ramu, A. Elliott, J. Zuegg, P. Duffy, M.-J. Rosslee, L. Hlaka, S. Kumar, M. Ozturk, F. Brombacher, M. Barrett, R. Guler and C. J. Suckling, *European Journal of Medicinal Chemistry* **2017**, *136*, 561-572.

7. D. W. Boykin, *Journal of the Brazilian Chemical Society* **2002**, *13*, 763-771.

8. A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M.-H. David-Cordonnier, D. W. Boykin and W. D. Wilson, *Bioorganic & Medicinal Chemistry Letters* **2015**, *25*, 4927-4932.

9. R. Nanjunda and W. D. Wilson, *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2012**, *Chapter*, Unit 8.8.

10. Y. Miao, M. P. H. Lee, G. N. Parkinson, A. Batista-Parra, M. A. Ismail, S. Neidle, D. W. Boykin and W. D. Wilson, *Biochemistry* **2005**, *44*, 14701-14708.

11. A. Paul, A. Kumar, R. Nanjunda, A. A. Farahat, D. W. Boykin and W. D. Wilson, *Organic & Biomolecular Chemistry* **2017**, *15*, 827-835.

12. P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin and W. D. Wilson, *Chemistry – A European Journal* **2016**, *22*, 15404-15412.

13. Y. Chai, A. Paul, M. Rettig, W. D. Wilson and D. W. Boykin, *The Journal of Organic Chemistry* **2014**, *79*, 852-866.

14. N. K. Harika, A. Paul, E. Stroeva, Y. Chai, D. W. Boykin, M. W. Germann and W. D. Wilson, *Nucleic Acids Research* **2016**, *44*, 4519-4527.

15. A. Paul, Y. Chai, D. W. Boykin and W. D. Wilson, *Biochemistry* **2015**, *54*, 577-587.

16. C. E. Bostock-Smith, C. A. Laughton and M. S. Searle, *Nucleic Acids Research* **1998**, *26*, 1660-1667.

17. A. Fede, A. Labhardt, W. Bannwarth and W. Leupin, *Biochemistry* **1991**, *30*, 11377-11388.

18. C. E. Bostock-Smith, K. J. Embrey and M. S. Searle, *Chemical Communications* **1997**, 121.

19. C. E. Bostock-Smith, S. A. Harris, C. A. Laughton and M. S. Searle, *Nucleic Acids Research* **2001**, *29*, 693-702.

20. S. M. Chen, W. Leupin, M. Rance and W. J. Chazin, *Biochemistry* **1992**, *31*, 4406-4413.

21. T. C. Jenkins, A. N. Lane, S. Neidle and D. G. Brown, *European Journal of Biochemistry* **1993**, *213*, 1175-1184.

22. M. Yoshida, D. L. Banville and R. H. Shafer, *Biochemistry* **1990**, *29*, 6585-6592.

23. I. Antony-Debre, A. Paul, J. Leite, K. Mitchell, H. M. Kim, L. A. Carvajal, T. I. Todorova, K. Huang, A. Kumar, A. A. Farahat, B. Bartholdy, S.-R. Narayanagari, J. Chen, A. Ambesi-Impiombato, A. A. Ferrando, I. Mantzaris, E. Gavathiotis, A. Verma, B. Will, D. W. Boykin, W. D. Wilson, G. M. K. Poon and U. Steidl, *Journal of Clinical Investigation* **2017**, In Press, Manuscript No.: 92504-JCI-RG-1.

24. W. D. Wilson, F. A. Tanious, A. Mathis, D. Tevis, J. E. Hall and D. W. Boykin, *Biochimie* **2008**, *90*, 999-1014.

25. T. C. Jenkins and A. N. Lane, *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1997**, *1350*, 189-204.

26. A. Fede, M. Billeter, W. Leupin and K. Wüthrich, *Structure* **1993**, *1*, 177-186.

27. G. R. Clark, D. W. Boykin, A. Czarny and S. Neidle, *Nucleic Acids Research* **1997**, *25*, 1510-1515.

28. F. J. Acosta-Reyes, C. Dardonville, H. P. de Koning, M. Natto, J. A. Subirana and J. L. Campos, *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70*, 1614-1621.

29. K. Wüthrich, *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. M. Rettig, M. W. Germann, S. Wang and W. D. Wilson, *ChemBioChem* **2013**, *14*, 323-331.

32. T. E. Cheatham and M. A. Young, *Biopolymers* **2000**, *56*, 232-256.

33. A. V. Fratini, M. L. Kopka, H. R. Drew and R. E. Dickerson, *Journal of Biological Chemistry* **1982**, *257*, 14686-14707.

34. D. G. Gorenstein, *Phosphorus-31 NMR: Principles and Applications*, Academic Press, New York, **1984**.

35. A. Lefebvre, O. Mauffret, E. Lescot, B. Hartmann and S. Fermandjian, *Biochemistry* **1996**, *35*, 12560-12569.

36. Z. Wu, N. Tjandra and A. Bax, *Journal of Biomolecular NMR*, **2001**, 19, 367-370.

37. S. Xhani, S. Esaki, K. Huang, N. Erlitzki and G. M. K. Poon, *The Journal of Physical Chemistry B* **2017**, *121*, 2748-2758.

38. Y. Liu, A. Kumar, S. Depauw, R. Nhili, M.-H. David-Cordonnier, M. P. Lee, M. A. Ismail, A. A. Farahat, M. Say, S. Chackal-Catoen, A. Batista-Parra, S. Neidle, D. W. Boykin and W. D. Wilson, *Journal of the American Chemical Society* **2011**, *133*, 10171-10183.

39. F. G. Loontiens, P. Regenfuss, A. Zechel, L. Dumortier and R. M. Clegg, *Biochemistry* **1990**, *29*, 9029-9039.

40. A. Krężel and W. Bal, *Journal of Inorganic Biochemistry* **2004**, *98*, 161-166.

41. D. G. Gorenstein, *Methods in Enzymology* **1992**, *211*, 254-286.

42. SPARKY, T. D. Goddard, D. G. Kneller, University of California, San Francisco (USA).

43. Gaussian 09, Revision B.01 Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT (2009) by M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al.

44. J. B. Foresman and A. E. Frisch, *Exploring Chemistry with Electronic Structure Methods*, 2/e, Gaussian, Inc., Pittsburgh, **1993**.

45. M. Orio, D. A. Pantazis and F. Neese, *Photosynthesis Research* **2009**, *102*, 443-453.

46. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *The Journal of Physical Chemistry* **1993**, *97*, 10269-10280.

47. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *Journal of the American Chemical Society* **1993**, *115*, 9620-9631.

48. U. C. Singh and P. A. Kollman, *Journal of Computational Chemistry* **1984**, *5*, 129-145.

49. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry* **2004**, *25*, 1157-1174.

50. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247-260.

51. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophysical Journal* **2007**, *92*, 3817-3829.

52. P. Athri and W. D. Wilson, *Journal of the American Chemical Society* **2009**, *131*, 7618-7625.

53. N. a. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza and J. Šponer, *Journal of the American Chemical Society* **2003**, *125*, 1759-1769.

54. D. Young, *Computational Chemistry*, Wiley-Interscience, **2001**. Appendix A. A.1.6 pg 330, Spartan.

55. Amber 14 (2014) by D. A. Case, V. Babin, J. T. Berryman, et al.

56. T. J. Macke and D. A. Case, *In Molecular Modeling of Nucleic Acids*, American Chemical Society, **1997**, Vol. 682; pp 379−393.

57. O. Trott and A. J. Olson, *Journal of computational chemistry* **2010**, *31*, 455-461.

58. E. J. Sorin, Y. M. Rhee and V. S. Pande, *Biophysical Journal* **2005**, *88*, 2516-2524.

59. W. Saenger, *In Principles of Nucleic Acid Structure*, Springer, New York, **1984**, pp 51−104.

60. R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute and K. Zakrzewska, *Nucleic Acids Research* **2009**, *37*, 5917-5929.

61. D. R. Roe and T. E. Cheatham, *Journal of Chemical Theory and Computation* **2013**, *9*, 3084-3095.

62. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33-38.

63. G. S. Couch, D. K. Hendrix and T. E. Ferrin, *Nucleic Acids Research* **2006**, *34*, e29-e29.

64. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of Computational Chemistry* **2004**, *25*, 1605-1612.

4 BOUND COMPOUND, INTERFACIAL WATER, AND PHENYL RING ROTATION DYNAMICS OF A COMPOUND IN THE DNA MINOR GROOVE

Narinder K. Harika and W. David Wilson.

Harika, N. K. and Wilson, W. D. *Biochemistry*. **2018** Jul 26. doi: 10.1021/acs.biochem.8b00647.

My contribution to this chapter was molecular dynamics studies, and writing.

4.1 Abstract

DB2277, a heterocyclic diamidine, is a successful design for mixed base pair (bp) DNA sequence recognition. The compound has a central aza-benzimidazole group that forms two H-bonds with a GC bp that has flanking AT bps. The nuclear magnetic resonance structure of the DB2277−DNA complex with an AAGATA recognition site sequence was determined, and here we report extended molecular dynamics (MD) simulations of the structure. DB2277 has two terminal phenyl-amidine groups, one of which is directly linked to the DB2277 heterocyclic core and the other through a flexible -OCH_2- group. The flexibly linked phenyl is too far from the minor groove floor to make direct H-bonds but is linked to an AT bp through water-mediated H-bonds. The flexibly linked phenyl-amidine with water-mediated H-bonds to the bases at the floor of the minor groove suggested that it might rotate in time spans accessible in MD. To test this idea, we conducted multi-microsecond MD simulations to determine if these phenyl rotations could be observed for a bound compound. In a 3 μs simulation, highly dynamic torsional motions were observed for the -OCH_2-

linked phenyl but not for the other phenyl. The dynamics periodically reached a level to allow 180° rotation of the phenyl while it was still bound in the minor groove. This is the first observation of rotation of a phenyl bound to DNA, and the results provide mechanistic details about how a rotation can occur as well as how mixed bp recognition can occur for monomer compounds bound to the minor groove.

4.2 Introduction

We are systematically designing entirely new heterocyclic amidine derivatives with the goalof recognizing a range of biologically important DNA sequences that is much broader than what has been previously possible with these and related agents (1−8). Additional goals, which are addressed in the research described in this paper, are to better understand the dynamics of bound minor groove complexes and the roles that waters of interaction play in the structure, binding selectivity, and energetics of complex formation. Heterocyclic amidine cations, such as pentamidine, furamidine, and berenil, whose DNA complexes have been extensively investigated, are uniformly specific for binding to AT DNA sequences. To also recognize G·C base pairs (bps), it is necessary to incorporate H-bond acceptor units, as noted many years ago by Dickerson and Lown (9−12). To broaden the applications of heterocyclic amidines, particularly in therapeutic applications, it is essential to expand their sequence recognition abilities (1−8). It should be noted that pentamidine and berenil are already used in human and animal therapy, and furamidine has recently reached Phase III clinical trials against human parasitic diseases. These therapeutic successes suggest that the heterocyclic amidines are excellent prospects for expanded uses, and this has recently been confirmed in leukemia (13).

The compounds listed above have H-bond donors for recognition of A·T base pair AN3 and T2 C O groups but no available acceptors. DB2277 (**Figure 4.1**) is a member of a new class

of designed heterocyclic diamidines that specifically recognize a single G·C bp with a broad range of flanking AT sequences (1,2,5,7). There are quite a large number of nuclear magnetic resonance (NMR) and crystal structures of heterocyclic amidines binding to pure AT DNA sequences but no previous structures with a centrally bound GC base pair (14−20). To begin to understand GC recognition by the new agents, a number of mixed sequence DNAs were used in NMR studies to understand DNA sequence-dependent effects of GC recognition by DB2277 in a minor groove complex (5). The formation of a unique 1:1 complex for NMR studies was achieved with a DNA sequence that has a central -AAGATA-binding site (7).

The structure of the AAGATA−DB2277 complex was determined by two-dimensional NMR and found to be quite unusual. The aza-benzimidazole-phenyl-amidine end of the complex fits tightly into the minor groove with numerous van der Waals contacts, H-bonds, and solvent interactions. The experimental structure models (**Figure 4.2**) show that the G5-NH in the minor groove of AAGATA forms an excellent H-bond with the aza-N of the aza-BI group, as expected from the compound design (7). The DB2277−G·C interaction is further locked down by the unexpected formation of an H-bond between the aza-BI-NH group and the C14-O2 group (cytidine at position 14) in the minor groove (7). The flanking AT sequences also strongly interact with DB2277. The amidine on the phenyl-amidine attached to the aza-BI group forms a direct H-bond in the minor groove to an AT base pair (T O···H−N-amidine). The amidine−DNA interaction is additionally stabilized by an ensemble of dynamic water H-bonds in the minor groove (7). The complex is also stabilized by phenyl C−H protons that are near the floor of the minor groove. These aromatic protons carry a small positive charge and can form stabilizing interactions with the A-N3 and T O groups. The DB2277−DNA interactions coupled to the dynamic stabilizing

hydration network and van der Waals interactions provide a very favorable complex and specific DNA recognition for the aza-BI-phenyl-amidine module.

The other phenyl-amidine of the complex is connected by a flexible -OCH_2- to the aza-benzimidazole (**Figure 4.1**). As a result of the overall complex structure, this terminal phenyl-amidine rises away from the floor of the minor groove such that it cannot form direct interactions with the A·T bps at the floor of the groove. Instead, there is an interfacial water molecule that links the inner facing amidine-NH to T16-O2 (**Figure 4.1**) at the floor of the groove (−NH···O−H···O T). This water molecule, unlike the other waters that are well-known to externally stabilize minor groove complexes (21, 22) is an integral part of a trimeric DB2277−DNA−water complex. Both amidines are also stabilized by a dynamic, external, extended water network in the groove (**Figure 4.2**). Similar waters of stabilization are seen in essentially all minor groove complexes that have been determined at sufficient resolution and are clearly a common feature of DNA minor groove interactions (23, 24). The unique feature of the DB2277−water interactions is the interfacial water interaction that is a very rare feature of small molecule−DNA complexes. It should be emphasized that the complexes that we are designing for GC bp recognition bind to the DNA minor groove as monomers and are quite different from the extensively studied polyamides that form stacked dimer or hairpin binding units (25−28).

To better understand the binding of DB2277 to a mixed base pair DNA sequence, we have conducted long time scale MD investigations for several reasons: (i) to investigate the local dynamics of both AT and GC bp interactions of DB2277 to better characterize the critical design features that are essential for strong, specific complex formation with mixed base pair DNA sequences; (ii) to determine whether the water-linked phenyl could rotate while DB2277 was bound tightly in the minor groove and, if so, determine what changes in the bound molecule,

interfacial water, and DNA are required to allow the flip; (iii) to determine whether in long MD simulations rotation of the other, directly H-bonded phenyl in DB2277 could be observed; and (iv) to investigate the influence on complex structure and stability of water molecules that are external to bound DB2277. By conducting multi-microsecond MD calculations starting with the NMR-restrained DB2277 structure, we were able to observe the rotation of the flexible-linked phenyl as well as how this rotation affects the interfacial water component of the complex. To the best of our knowledge, this is the first observation of rotation of an aromatic group bound to the DNA minor groove, and the mechanistic characteristics of the dynamic transitions are described here.

4.3 Material and Methods

Molecular dynamics (MD) simulations for 3.0 μs were performed without any restraints on the DB2277−AAGATA complex. All energy minimizations and molecular dynamics (MD) simulations were conducted on the PMEMD CUDA module of the Assisted Model Building with an Energy Refinement (AMBER14) program (29). The initial structure of the DB2277−DNA complex used in MD simulations was based on the restrained MD NMR structure [Protein Data Bank (PDB) entry 6AST] (7).

To conduct the molecular dynamics simulations of the DB2277 complex in the AMBER package, all the parameters of the molecules are required. Amber contains the parameters for DNA; however, the parameters of DB2277 used for MD simulation must be determined, and the values are reported in detail below. The procedure used is similar to that described by the Cheatham group (30) and our group for minor groove binders (31)

Force fields are key foundations of molecular mechanics with the fundamental assumption of Born-Oppenheimer approximation (32), where the energy of a system is described as a function of nuclear coordinates only. The accuracy of force fields further depends on the validity of

additional assumptions; additivity and transferability (32). Additivity corresponds to the expression of the potential energy of any system as a sum of different potentials and transferability refers to the development of the potential energy functions on a small set of molecules which can be applied to a much wider range of molecules with similar chemical groups. However, there is the limited availability of parameters for organic molecules in AMBER force fields (33). The equation to represent the additive form of force fields is as follows:

$$E_{total} = \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2}\left[1 \pm \cos(n\phi - \gamma)\right] + \sum_{\text{non-bonded}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}}\right] \quad (1)$$

The force field involving separation of potential energy terms in the equation 1 is empirical and provides useful insight into the determination of molecular structure and dynamics. The energy terms used in the functional form of the force field are described below:

The first term in the equation 1 represents the bond stretching and is denoted with a harmonic potential. K_r represents the force constant of the bond (generally very high), is an indication of the amount of energy required to stretch or compress a chemical bond significantly. r_0 represents the reference bond length, defined by the value of the bond length when all other terms in the potential energy function are zero.

The second term in the above equation 1 represents the change in potential energy in relation to deformation of bond angles. K_φ represents the force constant for the bond angle with values typically less than for bond stretching and is an indication that it takes less energy for a bond angle to deviate from its reference value.

The third term in the above equation 1 is referred to as the torsional term and is represented as the change in the potential energy of the molecular system as a function of the rotation about each dihedral angle. The energies involved here are significantly lower than for bond stretching and angle bending. V_n represents the barrier height, n represents the number of minima in the energy function (multiplicity) and γ is the phase factor, which determines the position of the minima.

The fourth term in the above equation 1 is represented by the non-bonded interactions involving atoms in the molecule separated by 3 or more bonds and between atoms in the different molecule. This term includes both the electrostatic interaction through the Coulombic potential, and the van der Waals interaction using the Lennard-Jones 12-6 potential. The van der Waals potential contains an attractive and a repulsive term. The attractive part represents the dispersion forces generated between instantaneous dipoles, which arise from fluctuations in electronic charge distributions in all molecules. The repulsive term reflects the observation that atoms repel each other below a certain distance (typically close to 0.3 nm). The electrostatic potential term is represented by the sum of electrostatic potentials generated by charges placed on atomic nuclei, i.e., partial atomic charges where r_{ij} is the distance between the nuclei i and j.

The foundation of classical force fields assumes that that similar chemical groups in different molecules interact in the same way (34). Therefore, the development of a force field is based on reproducing the energy surfaces for a set of small molecules that are typically well-characterized and contain the functional groups that occur in all biomolecules. This assumption gives a force field very broad applicability. Thus, force field parametrization expansion for new atom and bond types plays a very important role in molecular mechanics. For the internal terms in

the force field function, equilibrium bond length and equilibrium bond angles are mostly obtained from the experiment and high-level *ab-initio* calculations (33).

The first step in the generation of force field parameters includes specification of atom type of the molecule. Usage of more specific atom types helps in describing the chemical environment of molecule more accurately at the cost of burdening the parametrization (35). Selection of the consistent or appropriate charge approach is necessary to accurately fit the conformational and non-bonded energies in a transferable way. Default charge scheme used in the AMBER force fields (or GAFF parametrization), RESP "Restrained electrostatic potential" at HF/6-31G* works well with the small molecules (36-38).

4.3.1 *Force Field Parameters for DB2277*

Most of the parameters in the force field parameter files for the DB2277 molecule were derived from the existing set of bonds, angles and dihedrals for the similar atom types in parm99 and GAFF force fields (35, 39). The assignment of specifying atom types for the DB2277 molecule used in generating the force field parameters are shown in previously reported NMR structure of DB2277-DNA complex (7). Optimized geometry of DB2277 from Gaussian 09 (40) was used to obtain the equilibrium bond length for CK-CA and the equilibrium bond angle for CA-CA-N2. Torsional angles of DB2277 listed below play a significant role in defining its conformation and its interaction with DNA;

1) NB-CK-CA-CA and N2-CA-CA-CA: Obtained from previously reported parameters for DAPI and DB921 molecules (30, 31).

2) N*-CK-CA-CA, CT-OS-CA-CA, CT-OS-CA-NC CA-CT-OS-CA, CA-CA-CT-H1 and CA-CA-CT-OS: Obtained from previously reported parameters for DB2277 in DB2277-DNA NMR structure (70).

Since the dihedral angles, CA-CA-CT-OS, CA-CT-OS-CA, CT-OS-CA-NC and N*-CK-CA-CA are not defined in GAFF/Parm99 force field, therefore these dihedrals were re-parametrized and were reported earlier. These dihedrals define the rotation around bonds connecting the phenyl group and aza-benzimidazole moiety and are crucial in predicting the dynamic rotational movements of the phenyl and amidine moieties. Spartan16 (41) was used to perform *ab-initio* calculations using DFT/6-31+G (d) (42) level of theory for defining the above-mentioned torsional angle parameters. Potential energy profiles were obtained for dihedral angles varying from 0 to 180 degrees (30, 31) and were fitted with the least square fitting in kaleidaGraph software using the following cosine function equation:

$\sum Vn/2\ [1 + \cos(n\Phi - \gamma)]$

where Vn is the torsional barrier, n is periodicity (0,1,2,3,4 or 6), and γ is phase angle. The frcmod file (force field modification) modified with these newly parametrized dihedral angles of DB2277 can be found in the previously published NMR structure of DB2277-DNA complex.

The AMBER14 package was used to equilibrate the DB2277−DNA complex system using "ff14SB" and ff14SB/chi/ez (i.e., parmbsc0_chiOL4_ezOL1) force field modifications for DNA (43) Significant improvement in the development of empirical force fields and MD simulation methods has helped to describe the structure, energetics, and dynamics of nucleic acids more accurately (44−46) The parm94−99 force field was parametrized for simulation on the 1−10 ns time scale, but severe distortions of DNA structure were found in 50 ns simulations due to irreversible transitions in α and γ torsion angles. Thus, parmbsc0 (39) contains the reparametrized α and γ dihedral terms using the high-level quantum mechanical (QM) calculations on models of sugars and phosphates. Similarly, the χ distribution using relevant model systems was subsequently improved using various levels of theory in QM methods. The most tested χ

modifications are the "OL" modifications used in ff14SB. Therefore, the combination of ff14, ε/ζOL1, and χOL4, which includes ε/ζOL1 and χOL4 modifications to the ff14 force field, is suitable to use for DNA (47).

MD simulation of the complex was performed under explicit solvation conditions using a truncated octahedron periodic box filled with TIP3P water. The box wall was extended to a distance of 10.0 Å from any solute atom in each dimension. The system was neutralized by sodium cations using the LEAP module of AMBER. The particle mesh Ewald method with a 10.0 Å cutoff was used for the correct treatment of electrostatic interactions. During the equilibration, the SHAKE procedure was applied to constrain all bonds involving hydrogen atoms to their correct values. The system was relaxed with 1000 steps of energy minimization in the initial phase of equilibration. Then the system was heated from 0 to 300 K over 10 ps under constant-volume conditions. A production run on the system was subsequently performed for 3.0 μs for the DB2277−DNA complex under NPT conditions (isothermal, isobaric ensemble). For the detailed investigation of the dynamic motions of the phenyl rings of DB2277, the time interval for each frame was decreased to 0.25 ps during the 1020−1021 ns period of the MD simulation. Helical analysis of the DNA was done using the CURVES+ module (48) and analysis of trajectories was performed using various scripts in the cpptraj module (49). All of the molecular graphics were generated using VMD 1.9 and UCSF Chimera molecular visualization software (50−52).

4.4 Results and Discussion

NOESY spectra obtained for the DB2277−AAGATA complex in D_2O buffer show degeneracy in the chemical shift (δ) observed for all B1/B1′, B2/B2′, B5/B5′, and B6/B6′ phenyl proton pair signals of DB2277 as labeled in **Figure 4.1** (7). Observed averaging of chemical shifts of phenyl proton signals could be caused by rapid phenyl rotation or by an unlikely complete chemical shift

overlap of all four sets of phenyl protons on the NMR chemical shift time scale. NMR data reported by Embrey and Searle et al. for Hoechst 33258 indicate that the similar overlap observed for phenyl protons of that molecule is caused by rotational averaging of the signals (53, 54). Intramolecular exchange rate calculations from EXSY NMR between the two opposite binding orientations of DB2277 in the DB2277−AAGTTT complex (5′−3′ and 3′− 5′) at a 1:1 binding ratio are 6.8 s^{-1} at 303 K and 2.8 s^{-1} at 285 K (5). The half-life observed for the exchange between major and minor binding species is around 100 ms as determined by NMR, whereas the half-life for dissociation of the DB2277− AAGTTT complex is in the range of 30−60 s as determined by surface plasmon resonance. Therefore, the long rate of dissociation of DB2277 from the complex allows microscopic exchange between major and minor binding species of DB2277 while bound to DNA (5). It is also clearly evident from all of these observations that internal phenyl dynamics of DB2277 in the bound state are much faster than the exchange of DB2277 between binding orientations, end to end rotation, in the more symmetric minor groove complexes. With the -AAGATA-sequence, only one favored bound form of DB2277 is observed, and this allowed determination of the DB2277− DNA complex structure (7).

The initial conformation of the DB2277−AAGATA complex for MD simulations was based on the restrained NMR structure (7). The equilibrated structure obtained after 2 ns of MD with restraints obtained from NMR was used as an initial conformation for MD (**Figure 4.1**) (7). The three aromatic groups of DB2277 were initially very close to one plane, while the amidine groups are twisted ∼30° out of the phenyl plane due to a steric clash of protons on the phenyls and amidines (7).

The DNA sequence used in the restrained NMR structure was retained in the MD calculations, and the DB2277 complex is shown in a schematic graphic in **Figure 4.1**. Because of

the labeling the DB2277 atoms in the 3.0 μs MD simulations, it was obvious that the phenyl at the top of **Figure 4.1** (Ph1), attached to the flexible -CH_2O-, had rotated 180° several times while no rotation of Ph2 was observed in this time span. The two phenyl-amidine groups of DB2277 are both H-bonded to T-O2 groups at the floor of the minor groove; however, the Ph1 phenyl-amidine (**Figure 4.1**) is coupled through an interfacial water that is an integral part of the complex, and the MD results show that it is very dynamic (**Figure 4.2**). The observation that the phenyl amidine rotates 180° during the MD simulation is the first observation of this type for a small molecule bound to DNA and allows us to evaluate the mechanism of flipping.

The lower phenyl (Ph2), linked to BI, is constrained in the complex through an optimum H-bond of the inner-facing -NH_2 group of the amidine (Am2) with T7-O2, as well as strong van der Waals interactions with the walls and floor groups of the minor groove (**Figure 4.3**). Because of these constraints, no 180° rotational motion was observed for phenyl Ph2 throughout the MD simulations. Beginning and ending views of the flipped structures of Ph1 (C1 and C2 point out in frame 1218, while C3 and C4 point out in frame 3763) and the constant orientation of Ph2 (C17 and C19 are pointed out) are seen in **Figure 4.3A**. Clearly, the dynamics of these two phenyls are very different. The average values calculated for torsional angles Ø2 and α2 (**Figure 4.3B** and **Table 4.1**) show no major fluctuations from the standard geometry of the conformation in the DB2277−DNA complex with Ø2 close to 17° and α2 at 43° (average). **Figure 4.4** shows major fluctuations in torsional angles for Ph1, and these are discussed below.

4.4.1 *Dynamic Internal Motions of Phenyl Ph1*

Ph1 was investigated in detail to understand the interplay of water, DB2277, and DNA in the flipping motion of this phenyl. The 180° rotation of Ph1 attached to -OCH_2- is observed 16 times through the MD simulation. To analyze the flipping mechanism of the Ph1 phenyl, the MD

simulation frames related to a typical flip, 180° rotation, at ∼1 μs in the simulations, of Ph1 were observed in detail. Similar mechanisms of phenyl rotation in DB2277, however, were observed for all phenyl rotations in the simulation. Interestingly, the rotations observed for phenyl Ph1 were observed to happen in <1 ns, and the results shown here are captured for the rotation of Ph1 in 1 ns time period between 1020 and 1021 ns of the MD simulation.

For a high-resolution investigation of the rotational motion of Ph1, the time interval for each frame was decreased from our standard, 2 ps, to 0.25 ps. Up to frame 1218 of the observed nanosecond, the DB2277−AAGATA complex (**Figure 4.2**) stays close to the initial low-energy conformation that is similar to the published NMR structure of the complex (PDB entry 6AST) (7). In this conformation, N2 of the planar amidine, Am1, points above the plane of the aromatic groups of DB2277 (**Figure 4.2**) to facilitate the water-mediated H-bond with O2 of T16 (N2−H···O−H···O−T16). C1 and C2 Ph1 phenyl atoms (**Figure 4.2**) project out of the minor groove in this structure.

As a buildup to full phenyl rotations, substantial dynamic changes in Ø1 and α1 torsion angles of Ph1 phenyl are observed within a few picoseconds, i.e., around the 304.5−310 ps range, which relate to frames starting from frame 1218 and ending at frame ∼1240 (highlighted in orange bars in **Figure 4.4** and shown in **Figure 4.2**) while maintaining H-bonds of the Am1 group in the optimum range with DNA (**Table 4.1**). A large change in the α1 torsional angle observed at the 305 ps time step in the 1 ns flip region, i.e., from ∼40° to ∼137° (averaged), is highlighted by a red arrow in **Table 4.1** and **Figure 4.4**. Concerted flexible twists of the highly dynamic Ø1 and α1 torsions, in this time range, play a key role in maintaining the water-mediated H-bond contact with DNA that must be released before a 180° rotation can occur. The perpendicular position of Ph1 with respect to other aromatic rings of DB2277 caused by dynamic thermal motions of the atoms

at 309.75 ps in frame 1238 is shown in **Figure 4.2** with the large phenyl Ø1 twist of −73° and Am close to ∼150°. Highly dynamic states of Ø1 and α1 torsion angles of Ph1 between 305.25 and ∼800 ps are evident in **Table 4.1** and **Figure 4.4**. This dynamic range is well illustrated by frames 1238, 1660, and 2700 in **Figure 4.2**, where a water-mediated H-bond contact of Am1 with T16-O2 (T16-O2···H−O···H−N-Am1) helps to stabilize the complex energetics in frames 1660 and 1238. The configuration of the DB2277−DNA complex in frame 2700 is stabilized by a dynamic H-bond water network of the amidine and T-O2 in the minor groove of DNA (**Figure 4.2B**). These large motions of Ph1 clearly set it up for a complete rotation, while no such motions are observed for Ph2 (**Figures 4.3 and 4.4**).

4.4.2 *Role of Am1 Twist in the Ph1 Flip*

Role of Am1 Twist in the Ph1 Flip. Dynamic internal motions of the Am1 group play a vital role in stabilizing Ph1 in DB2277. The α1 torsion angle of amidine Am1 remains stable around −50° for almost 500 ps in the 1 ns trajectory as reported in **Table 4.1** (highlighted in green) and **Figure 4.4** before a transition in amidine group rotation. As a result of the large torsional fluctuations, the H-bond to the interfacial water is lost (**Figure 4.5A**) and Ph1 rapidly flips so that C3, C4, and N2 are pointed out of the minor groove in frame 3244 as compared to the orthogonal position of Ph1 relative to other aromatic moieties in the molecule in frame 1238 as shown in **Figure 4.2**. The total time span from frame 1238 to 3244 is around 500 ps. In the conformation shown in frame 3238 in **Figure 4.5**, N1 points inside the minor groove but is placed below the plane of Ph1 with no possible H-bond or water-mediated contact with DNA. Although the three aromatic rings of DB2277 are roughly planar in this structure, there is no H-bond stabilization of the Am1 group, and this structure has an energy that is higher than the optimum energy. The Am1 amidine rotates to a position in the plane of Ph1 in frame 3244, and a water molecule forms an H-bond with T16-O2;

however, there is still no H-bond contact of amidine Am1 with DNA (**Figure 4.5**). Significant rotation of the α1 torsion angle of Ph1 of DB2277 [frames 3244−3248 (**Figure 4.5**)] helps to place Am1 in an optimal position to form a water-mediated H-bond contact of Am1 with T16-O2. Between frames 3244 and 3248 (**Figure 4.5**), the structure moves to a lower energy by forming the water-mediated H-bond contact of the amidine group with T16-O2 (T16-O2···2.2 Å···H-O···1.9 Å···H−N-Am1). The three aromatics are close to planar, and water-mediated H-bond contacts help to stabilize this conformation. The dramatic transition of the α1 torsion angle from approximately −50° to approximately 40° within 0.5 ps (frames 3244 and 3245) is evident in the reported 810.75−811.25 ps time span marked with a red arrow in Table 1. This large twist observed in α1 of the Am1 group (**Table 4.1** and **Figure 4.4**) facilitates the full 180° rotation in Ph1 as seen in frame 3763 at 940.75 ps (**Figure 4.3A**) which is quite similar to frame 3248 (**Figure 4.5**). C3 and C4 atoms of the flipped phenyl point out of the minor groove after rotation in **Figure 4.5**. This low-energy con-formation of DB2277 is similar to the initial structure and remains stable for several hundred nanoseconds when Ph1 undergoes a reverse 180° rotation back to the initial state. Two key H-bonds responsible for the specificity and stability of the DB2277−DNA complex (**Figure 4.3A**) remain in an optimal range in the flipped phenyl conformations of DB2277 during rotation of Ph1: G5-NH_2 with N of aza-BI (3.2 Å) and C14-O2 (cytidine at position 14) with NH-aza-BI (3.2 Å). The lower phenyl Ph2 remains tightly bound in the minor groove of DNA with an H-bond contact of Am2-NH with T7-O2 (2.9 Å).

The variations in the relative motions of the base pairs in the DB2277−DNA complex over the 1 ns trajectory of the Ph1 flip are represented in terms of their helical parameters (minor groove width, roll, and twist angles) displayed in **Figure 4.6**. Dynamic motions of the minor groove as the bound DB2277 phenyl-amidine undergoes large torsional fluctuations cause increases in the

minor groove width of DNA (**Figure 4.6**). Thus, transient breathing motions of the complex assist in the dynamic flipping of Ph1. Helical twist calculations for DNA are consistent with the published helical analysis of the NMR structure of the DB2277−DNA complex.

In summary, we have shown how the flexible -OCH_2- linker in DB2277 provides freedom for the Ph1 phenyl to project away from the floor of the minor groove of DNA. The flexibility of -OCH_2- Ph1 helps DB2277 to track along the minor groove curvature by forming indirect and dynamic water-mediated H-bond contacts of Am1 with the bases at the floor of the minor groove. The flexibility and water-mediated contact allow Ph1 to rotate several times in a microsecond so that we are able to observe the rotation in a bound compound for the first time. During these flips of Ph1, the aza-benzimidazole H-bonds to G-NH_2 and C-O2 (cytidine) of the central GC bp, which account for the mixed sequence recognition by DB2277, and H-bonds of the amidine of Ph2 remain quite constant. Neither the aza-BI nor Ph2 shows the large dynamic motions as seen with Ph1 or a 180° rotation.

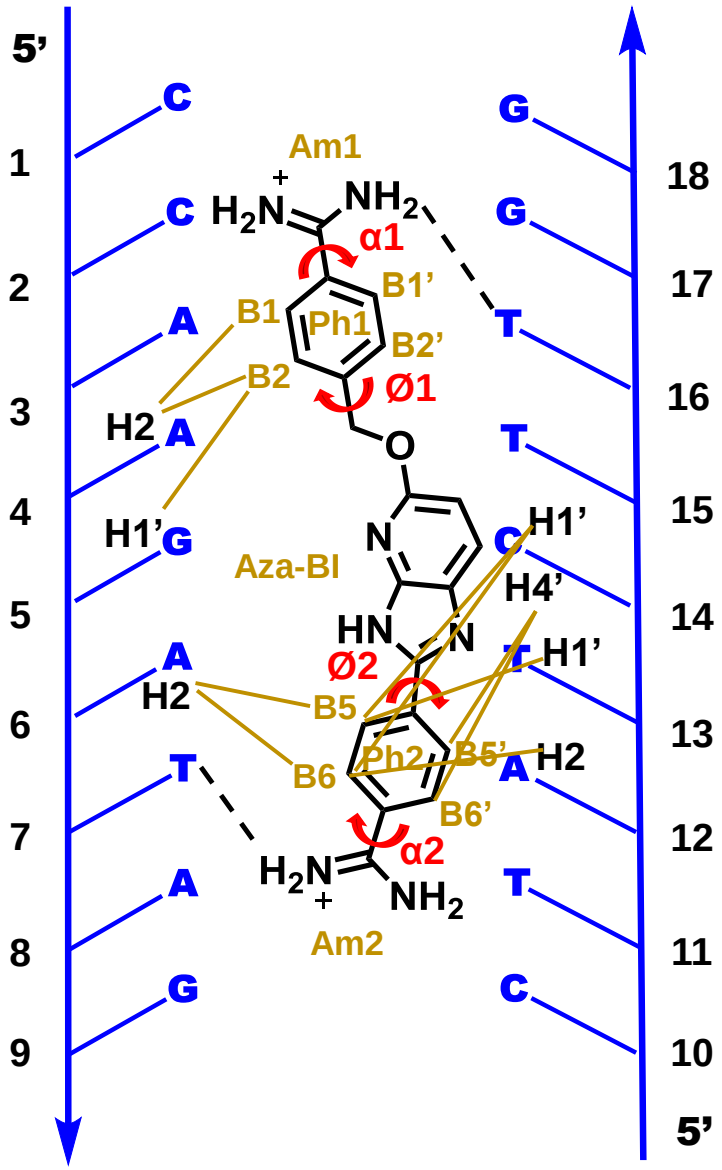


Figure 4.1 Molecular structure of DB2277 with the target DNA sequence.
Abbreviations for aromatic rings are listed along with their torsion angles. Brown lines represent the nuclear Overhauser effect interactions of DB2277 protons with specific DNA protons in the minor groove (7). Dashed lines illustrate the H-bond interactions of amidines with T-O2 in the minor groove of DNA.

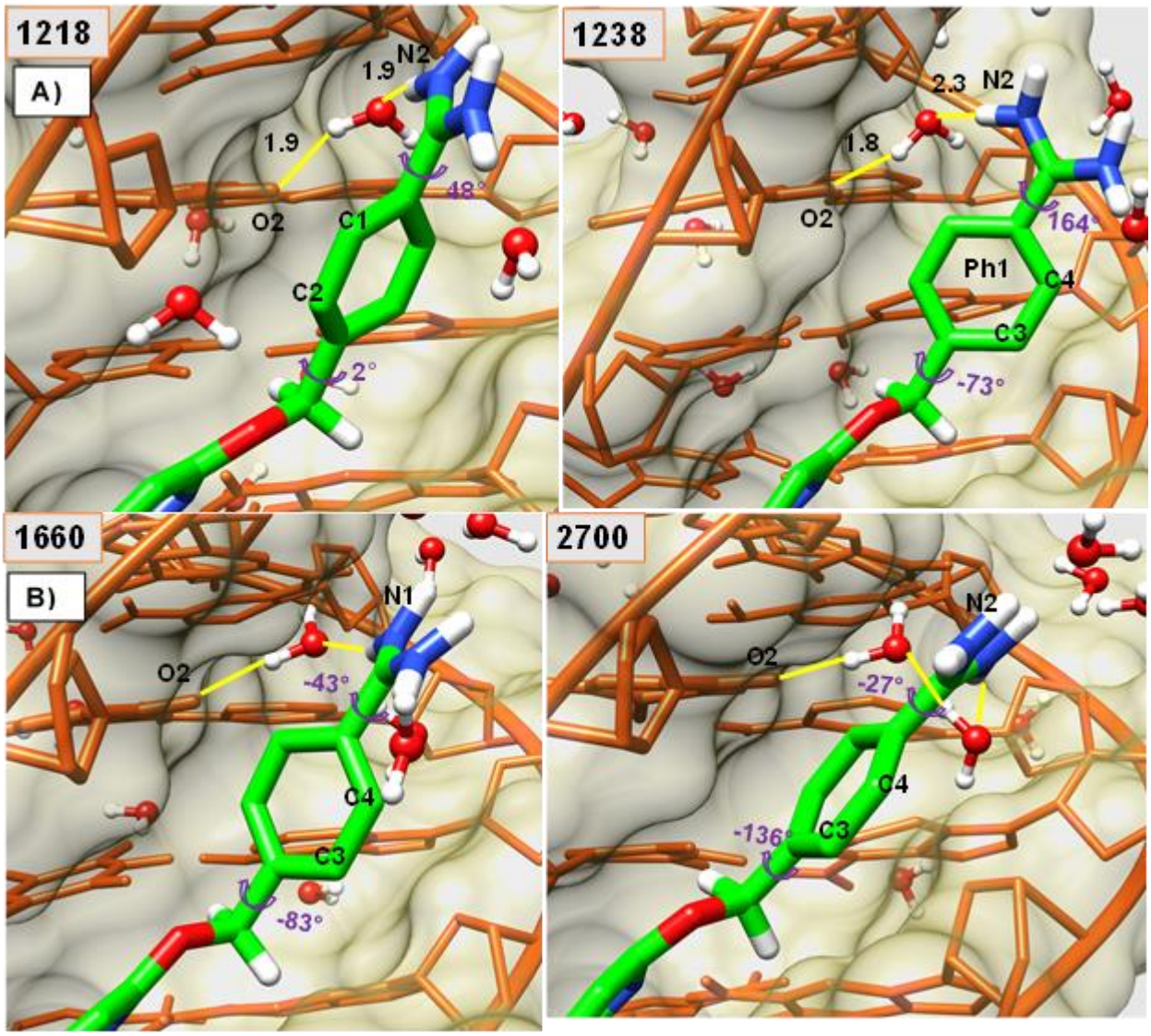


Figure 4.2 Minor groove views of the DB2277−DNA complex.

Ø1 and α1 torsion angles are shown at several key times (frame number at the top left) along with water-mediated H-bonds (yellow, angstroms). Panel A shows larger twists for α1 and Ø1, via comparison of frame 1218 with the orthogonal position of Ph1 relative to other aromatic groups of bound DB2277 in frame 1238. Starting at frame 1218 with C1 and C2 pointed out in panel A and end at the frame 2700 with C3 and C4 pointed out in panel B. The DNA backbone is represented as orange sticks with khaki space fill. The DB2277 molecule is shown as sticks (green, C). Water molecules are shown as balls and sticks.

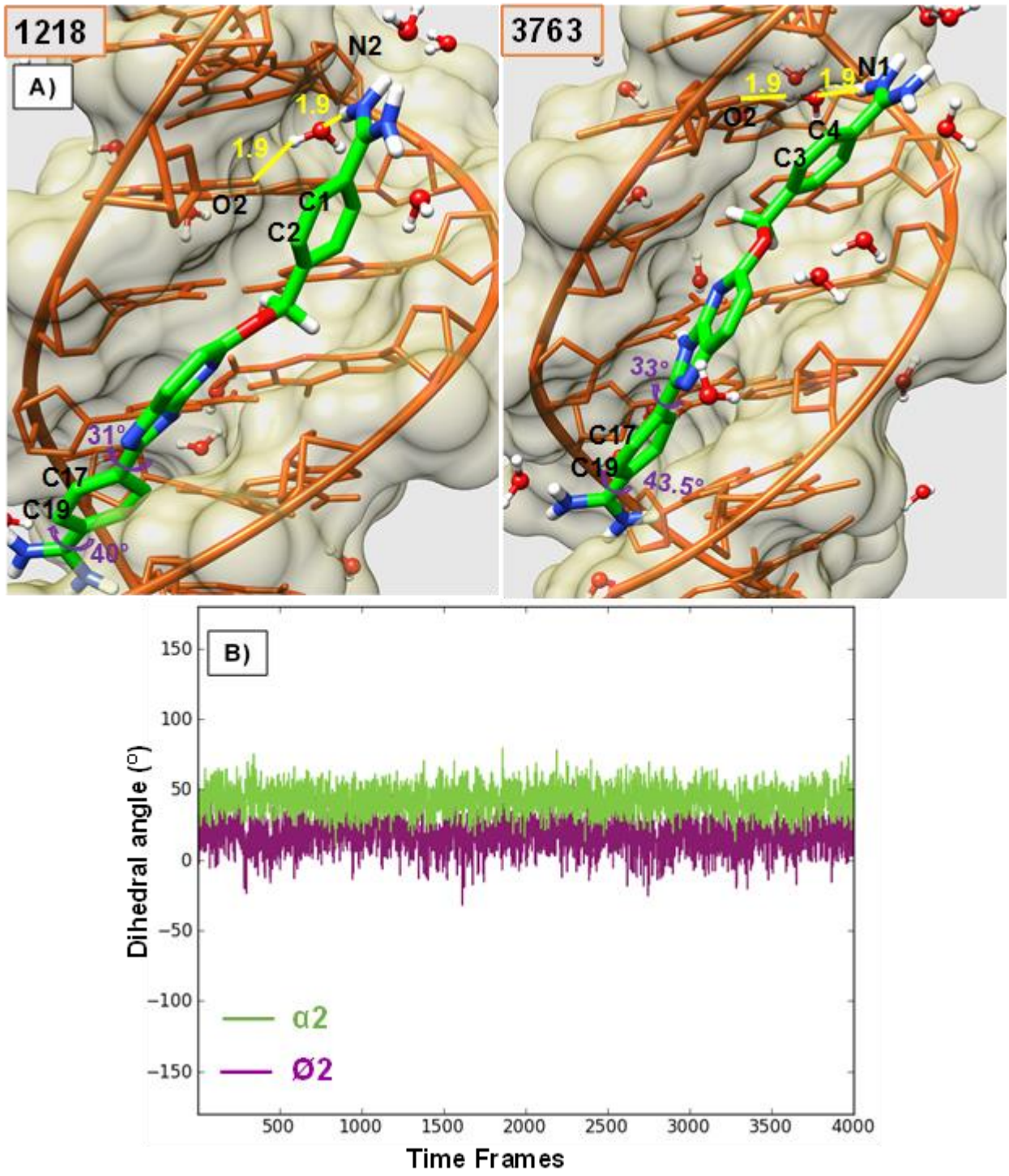


Figure 4.3 Minor groove views of the DB2277−DNA complex with torsion angle plot for Ph2 phenyl.

(A) Minor groove views of the DB2277−DNA complex with reported Ø2 and α2 torsion angles of phenyl Ph2 and amidine Am2, respectively, for the initial frame and the end frame for a 1 ns trajectory of the Ph1 flip. Colors as in Figure 4.2. (B) Torsion angle plot of Ø2 and α2 angles of Ph2 in the bound DB2277 molecule for a 1 ns trajectory. The time interval for each frame in the plot is 0.25 ps of a 1 ns trajectory (1020− 1021 ns of the 3 μs MD simulation).

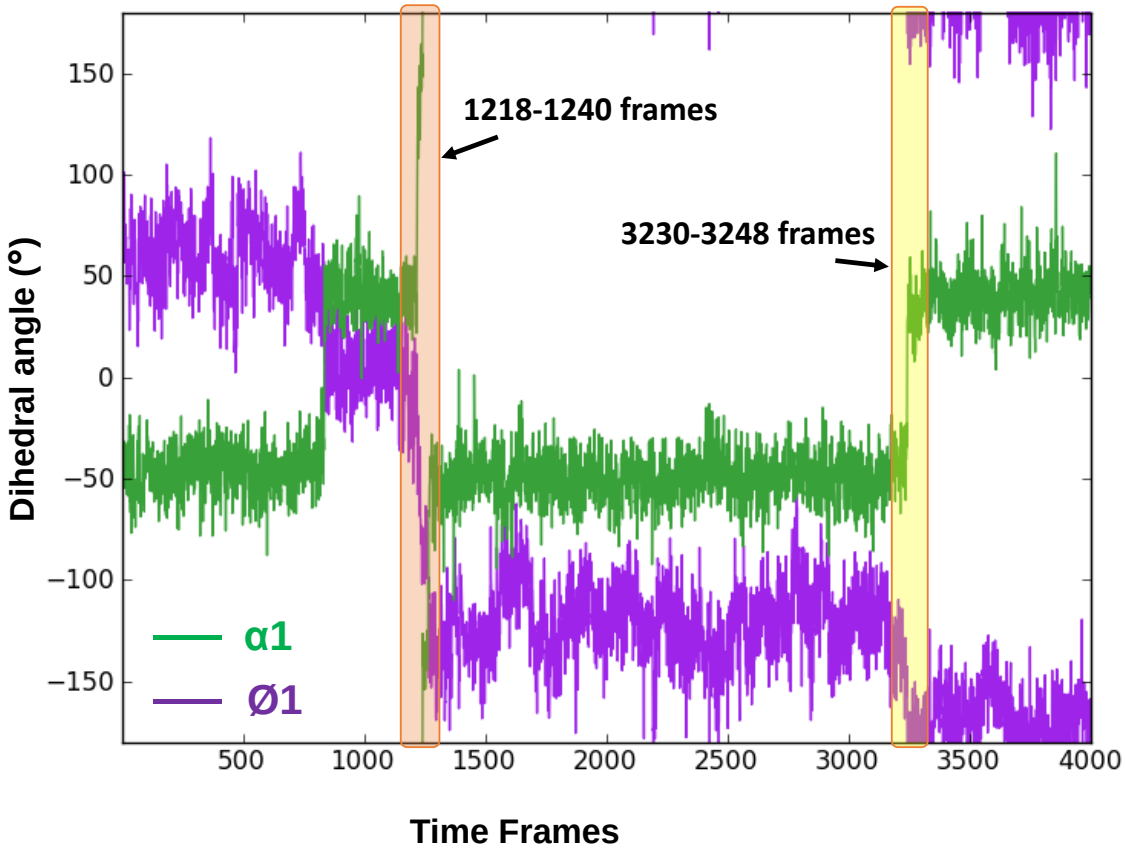


Figure 4.4 Torsion angle plot of Ø1 and α1 angles of phenyl Ph1 and amidine Am1, respectively, of the bound DB2277 molecule.

1 ns trajectory of the Ph1 flip. Major transitions of Ø1 and α1 torsions are highlighted and illustrated in time frames in Figures 4.2 and 4.5. The time interval for each frame in the plot is 0.25 ps of a 1 ns trajectory (1020−1021 ns of the 3 μs MD simulation).

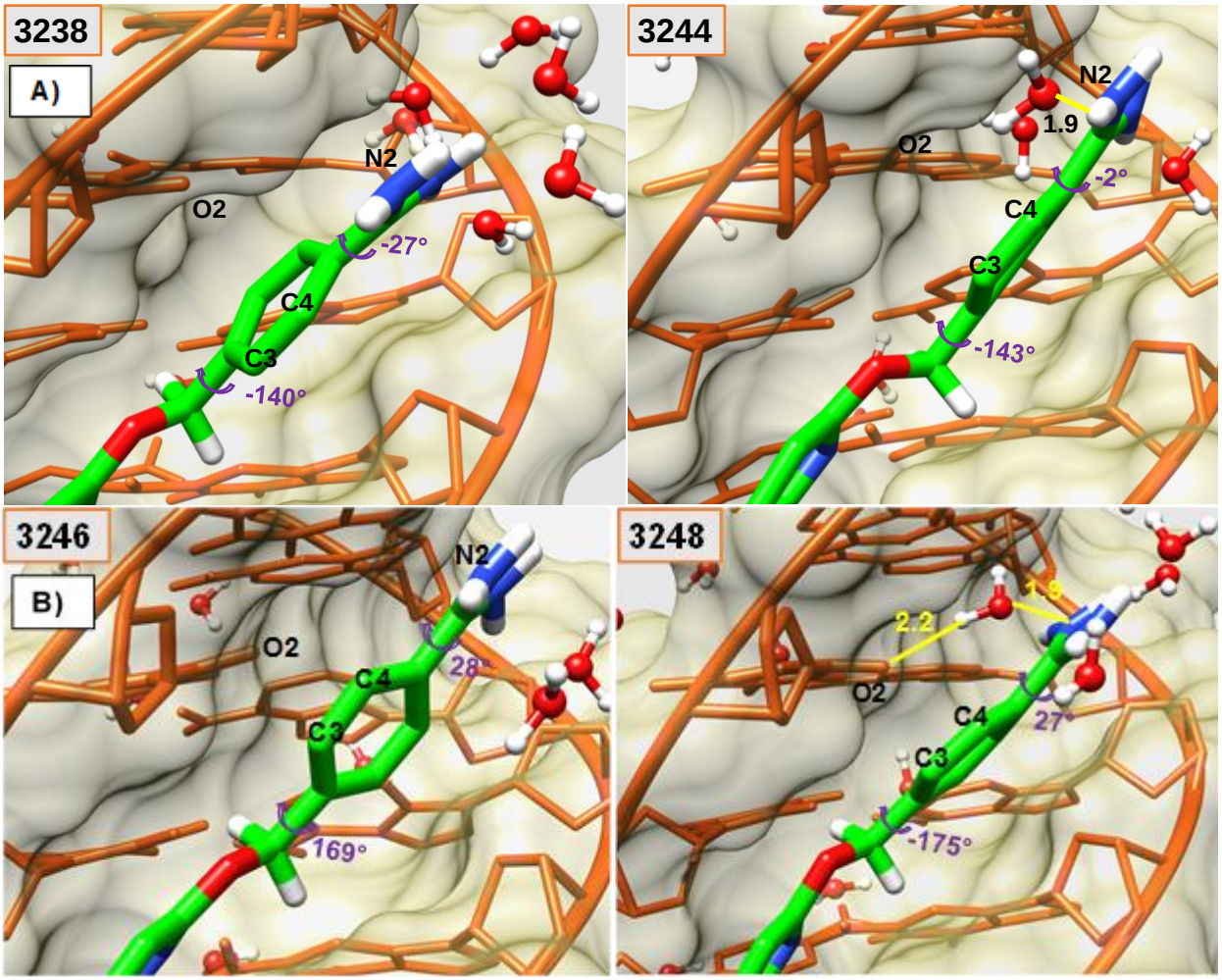


Figure 4.5 Minor groove views of the DB2277−DNA complex capturing the concerted motions of water molecules and Am1 twist to facilitate the Ph1 flip.
Frames captured within a few picoseconds (frames 3238−3248). Flipped Ph1 in complex with C3 and C4 atoms pointed out of the minor groove. (B) The water-mediated H-bond network of the Am1 group with the minor groove of DNA helps to stabilize the structure in frame 3248 as it approaches the minimum energy compared to other frames. Water-mediated H-bonds (yellow) between T16-O2 and NH of Am1 at the floor of the minor groove are reported in angstroms in different frames. Ø1 and α1 torsion angles are also reported. Colors as in Figure 4.2.

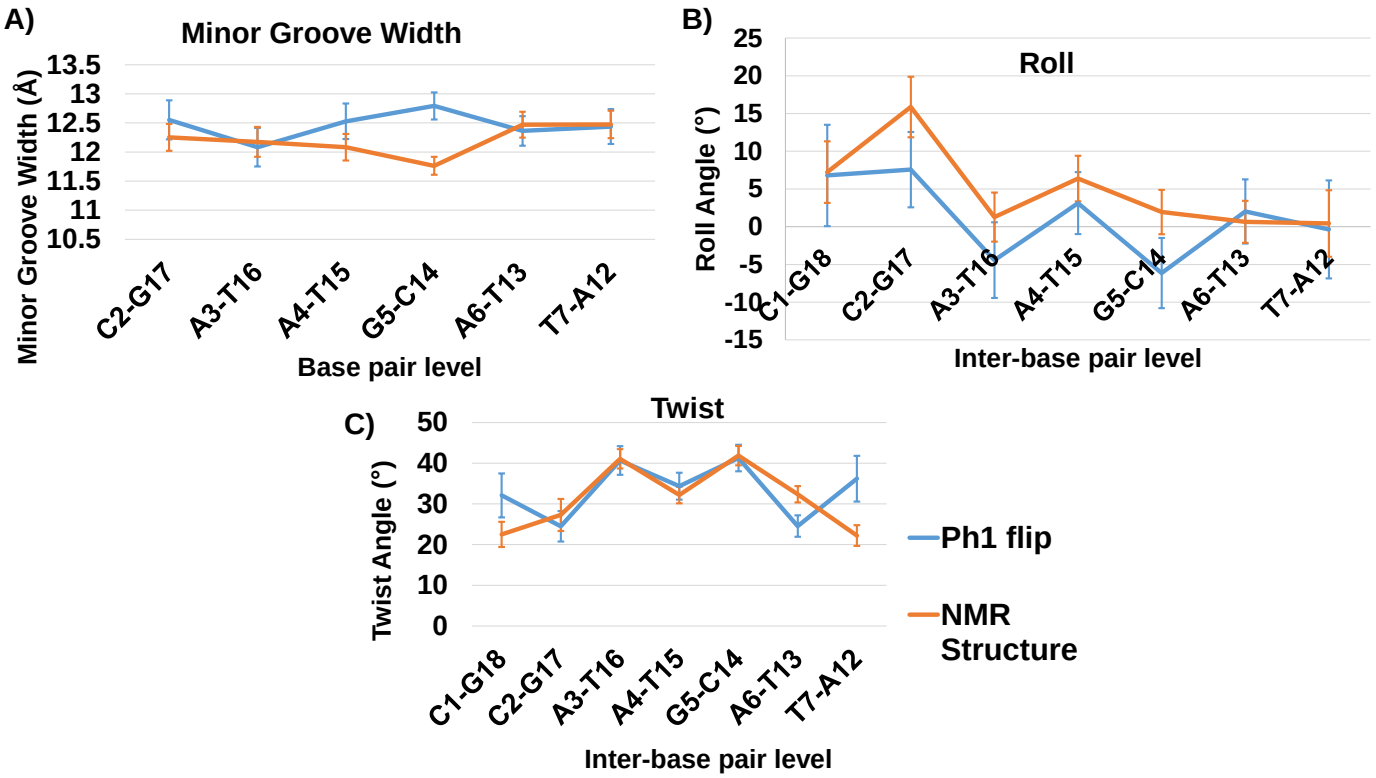


Figure 4.6 Minor groove width, Roll angles and Twist angles of bound DNA in comparison to the NMR structure of DB2277-DNA complex.
A) Minor groove width (in Å) of bound DNA for 1-ns trajectory (blue) in comparison to the minor groove width of DNA in the NMR structure of DB2277-DNA complex (orange). B) Roll angles and C) Twist angles at inter-base pair level for the bound DNA in 1ns Ph1-flip trajectory compared to roll and twist angles of DNA in NMR structure of DB2277-DNA complex (orange). Standard deviation calculated from all time steps in 1-ns trajectory of Ph1 flip.

Table 4.1 Phenyl and Amidine Torsion Angles Ø1, Ø2, α1, and α2 of the Bound DB2277 Molecule during the 1 ns Trajectory of the Ph1 Flip[a]

Time (ps) (Frames)	Ø1 (°)	α1 (°)	Ø2 (°)	α2 (°)
0-208.75 (0-838)	60 ± 18	-46 ± 11	17 ± 11	42 ± 9
209.5-304.25 (838-1217)	4 ± 16	40 ± 11		
305.25-310.25 (1221-1241)	-80 ± 144 (highly dynamic)	137 ± 17		
316.75-810.75 (1267-3243)		-50 ± 12		
811.25-1000 (3245-4000)		40 ± 12		

[a]*Frame numbers are shown in parentheses below the time steps (in picoseconds) where each frame is 0.25 ps. The 0 ps point of the flip period points at the 1020 ns point of the 3 μs MD simulations. Significant shifts in the α1 value are marked with red arrows, and the time range that shows a stable α1 value is highlighted in green. The standard deviation is calculated using all the frames within the time range mentioned in the table. Ø2 and α2 stay in a limited structure for the entire simulation. Torsional plots using these values are also shown in Figures 4.3 and 4.4.*

4.5 References

1) Chai, Y., Paul, A., Rettig, M., Wilson, W. D., and Boykin, D. W. (2014) Design and Synthesis of Heterocyclic Cations for Specific DNA Recognition: From AT-Rich to Mixed-Base-Pair DNA Sequences. J. Org. Chem. 79, 852−866.

2) Paul, A., Chai, Y., Boykin, D. W., and Wilson, W. D. (2015) Understanding Mixed Sequence DNA Recognition by Novel Designed Compounds: The Kinetic and Thermodynamic Behavior of Azabenzimidazole Diamidines. Biochemistry 54, 577−587.

3) Paul, A., Nanjunda, R., Kumar, A., Laughlin, S., Nhili, R., Depauw, S., Deuser, S. S., Chai, Y., Chaudhary, A. S., David-Cordonnier, M. H., Boykin, D. W., and Wilson, W. D. (2015) Mixed up minor groove binders: Convincing A.T specific compounds to recognize a G.C base pair. Bioorg. Med. Chem. Lett. 25, 4927−4932.

4) Guo, P., Paul, A., Kumar, A., Farahat, A. A., Kumar, D., Wang, S., Boykin, D. W., and Wilson, W. D. (2016) The Thiophene ″Sigma-Hole″ as a Concept for Preorganized, Specific Recognition of GC Base Pairs in the DNA Minor Groove. Chem. - Eur. J. 22, 15404− 15412.

5) Harika, N. K., Paul, A., Stroeva, E., Chai, Y., Boykin, D. W., Germann, M. W., and Wilson, W. D. (2016) Imino proton NMR guides the reprogramming of A*T specific minor groove binders for mixed base pair recognition. Nucleic Acids Res. 44, 4519−4527.

6) Guo, P., Paul, A., Kumar, A., Harika, N. K., Wang, S., Farahat, A. A., Boykin, D. W., and Wilson, W. D. (2017) A modular design for minor groove binding and recognition of mixed base pair sequences of DNA. Chem. Commun. 53, 10406−10409.

7) Harika, N. K., Germann, M. W., and Wilson, W. D. (2017) First Structure of a Designed Minor Groove Binding Heterocyclic Cation that Specifically Recognizes Mixed DNA Base Pair Sequences. Chem. - Eur. J. 23, 17612−17620.

8) Paul, A., Kumar, A., Nanjunda, R., Farahat, A. A., Boykin, D. W., and Wilson, W. D. (2017) Systematic synthetic and biophysical development of mixed sequence DNA binding agents. Org. Biomol. Chem. 15, 827−835.

9) Goodsell, D., and Dickerson, R. E. (1986) Isohelical analysis of DNA groove-binding drugs. J. Med. Chem. 29, 727−733.

10) Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., and Dickerson, R. E. (1997) Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. Structure 5, 1033−1046.

11) Kissinger, K., Krowicki, K., Dabrowiak, J. C., and Lown, J. W. (1987) Molecular recognition between oligopeptides and nucleic acids. Monocationic imidazole lexitropsins that display enhanced GC sequence dependent DNA binding. Biochemistry 26, 5590−5595.

12) Goodsell, D. S., Ng, H. L., Kopka, M., Lown, J. W., and Dickerson, R. E. (1995) Structure of a dicationic monoimidazole lexitropsin bound to DNA. Biochemistry 34, 16654−16661.

13) Antony-Debre,́I., Paul, A., Leite, J., Mitchell, K., Kim, H. M., Carvajal, L. A., Todorova, T. I., Huang, K., Kumar, A., Farahat, A. A., Bartholdy, B., Narayanagari, S.-R., Chen, J., Ambesi-Impiombato, A., Ferrando, A. A., Mantzaris, I., Gavathiotis, E., Verma, A., Will, B., Boykin, D. W., Wilson, W. D., Poon, G. M. K., and Steidl, U. (2017) Pharmacological inhibition of the transcription factor PU.1 in leukemia. J. Clin. Invest. 127, 4297−4313.

14) Campbell, N. H., Evans, D. A., Lee, M. P., Parkinson, G. N., and Neidle, S. (2006) Targeting the DNA minor groove with fused ring dicationic compounds: comparison of in silico screening and a high-resolution crystal structure. Bioorg. Med. Chem. Lett. 16, 15−19.

15) Joubert, A., Sun, X. W., Johansson, E., Bailly, C., Mann, J., and Neidle, S. (2003) Sequence-selective targeting of long stretches of the DNA minor groove by a novel dimeric bis-benzimidazole. Biochemistry 42, 5984−5992.

16) Neidle, S. (2001) DNA minor-groove recognition by small molecules. Nat. Prod. Rep. 18, 291−309.

17) Rettig, M., Germann, M. W., Ismail, M. A., Batista-Parra, A., Munde, M., Boykin, D. W., and Wilson, W. D. (2012) Microscopic Rearrangement of Bound Minor Groove Binders Detected by NMR. J. Phys. Chem. B 116, 5620−5627.

18) Glass, L. S., Nguyen, B., Goodwin, K. D., Dardonville, C., Wilson, W. D., Long, E. C., and Georgiadis, M. M. (2009) Crystal Structure of a Trypanocidal 4,4′-Bis(imidazolinylamino)-diphenylamine Bound to DNA. Biochemistry 48, 5943−5952.

19) Lane, A. N., Jenkins, T. C., and Frenkiel, T. A. (1997) Hydration and solution structure of d(CGCAAATTTGCG)2 and its complex with propamidine from NMR and molecular modelling. Biochim. Biophys. Acta, Gene Struct. Expression 1350, 205−220.

20) Lane, A. N., Jenkins, T. C., Brown, T., and Neidle, S. (1991) Interaction of berenil with the EcoRI dodecamer d-(CGCGAATTCGCG)2 in solution studied by NMR. Biochemistry 30, 1372−1385.

21) Degtyareva, N. N., Wallace, B. D., Bryant, A. R., Loo, K. M., and Petty, J. T. (2007) Hydration changes accompanying the binding of minor groove ligands with DNA. Biophys. J. 92, 959−965.

22) Spitzer, G. M., Fuchs, J. E., Markt, P., Kirchmair, J., Wellenzohn, B., Langer, T., and Liedl, K. R. (2008) Sequence-specific positions of water molecules at the interface between DNA and minor groove binders. ChemPhysChem 9, 2766−2771.

23) Nguyen, B., Neidle, S., and Wilson, W. D. (2009) A Role for Water Molecules in DNA−Ligand Minor Groove Recognition. Acc. Chem. Res. 42, 11−21.

24) Savelyev, A., and MacKerell, A. D., Jr. (2015) Differential Deformability of the DNA Minor Groove and Altered BI/BII Backbone Conformational Equilibrium by the Monovalent Ions Li(+), Na(+), K(+), and Rb(+) via Water-Mediated Hydrogen Bonding. J. Chem. Theory Comput. 11, 4473−4485.

25) Lee, M., Krowicki, K., Shea, R. G., Lown, J. W., and Pon, R. T. (1989) Molecular recognition between oligopeptides and nucleic acids. Specificity of binding of a monocationic bis-furan lexitropsin to DNA deduced from footprinting and ^{1}H NMR studies. J. Mol. Recognit. 2, 84−93.

26) Kawamoto, Y., Bando, T., and Sugiyama, H. (2018) Sequence-specific DNA binding Pyrrole-imidazole polyamides and their applications. Bioorg. Med. Chem. 26, 1393−1411.

27) Dervan, P. B., Doss, R. M., and Marques, M. A. (2005) Programmable DNA Binding Oligomers for Control of Transcription. Curr. Med. Chem.: Anti-Cancer Agents 5, 373−387.

28) Wemmer, D. E., and Dervan, P. B. (1997) Targeting the minor groove of DNA. Curr. Opin. Struct. Biol. 7, 355−361.

29) Case, D. A., Babin, V., and Berryman, J. T. (2014) Amber14.

30) Spackova, N.a., Cheatham, T. E., Ryjacek, F., Lankas, F., van Meervelt, L., Hobza, P., and Sponer, J. (2003) Molecular Dynamics Simulations and Thermodynamics Analysis of DNA−Drug Com-plexes. Minor Groove Binding between 4‘, 6-Diamidino-2-phenyl-indole and DNA Duplexes in Solution. J. Am. Chem. Soc. 125, 1759− 1769.

31) Athri, P., and Wilson, W. D. (2009) ″Molecular Dynamics of Water-Mediated Interactions of a Linear Benzimidazole−Biphenyl Diamidine with the DNA Minor Groove.″. J. Am. Chem. Soc. 131(22), 7618−7625.

32) Leach, A. (2009) Molecular modelling: principles and applications, Pearson Prentice Hall.

33) Wang, J., Wolf, R. M.; Caldwell, J. W., Kollman, P. A. and Case, D. A. "Development and testing of a general AMBER force field". *J. Comput. Chem.*, 25, 2004, 1157-1174.

34) L. Monticelli and D. Peter Tieleman, Force Fields for Classical Molecular Dynamics, 2013, p. 197-213.

35) Wang, J., Wang, W., Kollman P. A.; Case, D. A. "Automatic atom type and bond type perception in molecular mechanical calculations". *J. Mol. Graph. Model.* 25, 2006, 247260.

36) Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem*. 97, 10269-10280.

37) Cornell, W.D., Cieplak, P., Bayly, C.I. and Kollmann, P.A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc*., 115, 9620-9631.

38) Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5(2):129–145.

39) Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/ gamma conformers. Biophys. J. 92, 3817−3829.

40) Gaussian 09, Revision B.01 Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT (2009) by M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al.

41) Computational Chemistry, David Young, Wiley-Interscience, 2001. Appendix A. A.1.6 pg 330, Spartan.

42) Orio, M., Pantazis, D.A. and Neese, F. (2009) Density functional theory. *Photosynth. Res.* 102, 443-453.

43) Galindo-Murillo, R., Robertson, J. C., Zgarbova, M., Sponer, J., Otyepka, M., Jurecka, P., and Cheatham, T. E. (2016) Assessing the Current State of Amber Force Field Modifications for DNA. J. Chem. Theory Comput. 12, 4114−4127.

44) Auffinger, P., and Westhof, E. (1998) Simulations of the molecular dynamics of nucleic acids. Curr. Opin. Struct. Biol. 8, 227− 236.

45) Beveridge, D. L., and McConnell, K. J. (2000) Nucleic acids: theory and computer simulation, Y2K. Curr. Opin. Struct. Biol. 10, 182−196.

46) Cheatham, T. E., III, and Kollman, P. A. (2000) Molecular dynamics simulation of nucleic acids. Annu. Rev. Phys. Chem. 51, 435− 471.

39) Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/ gamma conformers. Biophys. J. 92, 3817−3829.

47) Cheatham, T. E., 3rd, and Case, D. A. (2013) Twenty-five years of nucleic acid simulations. Biopolymers 99, 969−977.

48) Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res. 37 (17), 5917−5929.

49) Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput. 9, 3084−3095.

50) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. J. Mol. Graphics 14, 33−38.

51) Couch, G. S., Hendrix, D. K., and Ferrin, T. E. (2006) Nucleic acid visualization with UCSF Chimera. Nucleic Acids Res. 34, No. e29.

52) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605−1612.

53) Embrey, K. J., Searle, M. S., and Craik, D. J. (1993) Interaction of Hoechst 33258 with the minor groove of the A+T-rich DNA duplex d(GGTAATTACC)2 studied in solution by NMR spectros-copy. Eur. J. Biochem. 211, 437−447.

54) Searle, M. S., and Embrey, K. J. (1990) Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ^{1}H NMR spectroscopy. Nucleic Acids Res. 18, 3753−3762.

5 CONCLUSIONS

In this dissertation, the molecular recognition of the mixed base pair (bp) sequences of DNA by small molecules has been systematically investigated. DB2277, a novel heterocyclic cation specifically recognizes a single GC bp flanked by AT sequences in the minor groove of DNA. The NMR spectroscopy was used to illustrate the highly dynamic nature of DB2277 binding with AAAGTTT type of mixed bp sequences. The strong binding efficiency allows the compound to form two reverse-oriented microstructures in the minor groove of DNA of relatively symmetric DNAs such as AAAGTTT. Comparison of NMR exchange rate dynamics with the much longer dissociation off-rate from surface plasmon resonance (SPR) has shown that DB2277 rapidly flips back and forth between the two opposite minor groove orientations before its complete dissociation from the target DNA. Examination of over-thirty mixed bp DNA hairpin sequences using NMR and thermal melting resulted in the finding of a unique DNA sequence with AAGATA binding site that showed a single and selective binding orientation with DB2277.

This dissertation has also illustrated the first solution structure of the heterocyclic cation, DB2277, with an AT/GC mixed DNA sequences in a 1:1 complex. The molecular basis of the specific recognition of a single GC bp in AAGATA binding site by DB2277 has been well described in the high-resolution NMR structure of the complex. The structural insight into the specific interactions involved in the DB2277-DNA complex was obtained using NOE (Nuclear Over Hauser effect) restraints from NMR in conjunction with molecular dynamics (MD) simulations. The NMR-restrained MD structure has specified the strong interactions of G-NH and C-O2 of the central bp in mixed-sequence DNA with aza-N and aza BI -NH of the aza-BI (aza-benzimidazole) group of DB2277, respectively. These two key interactions indicate how a small molecule like DB2277 is able to selective recognize mixed bp DNA sequences.

Extended MD calculations on the NMR structure of the DB2277-DNA complex reported further stabilization of the complex through a dynamic hydration network and interfacial water molecules that play a vital role in mediating the interactions between amidinium protons and DNA bps. The flexibility imparted to the phenyl, which is linked to the heterocyclic core of the DB2277 via a flexible linker (-OCH_2- group), and interfacial water molecules allow this phenyl to undergo 180° rotations in the bound compound in the multi-microsecond MD simulations. This is the first time that phenyl rotations have been reported in the DNA bound state of a compound along with the mechanistic details on the rotation. Structural and dynamic aspects of DB2277 in complex with the mixed bp DNA sequences studied here are very useful in designing additional compounds to bring more diversity in the sequence-specific DNA recognition field. Therefore, designing additional heterocyclic compounds to target a wide range of sequence-specific DNAs will be very useful in extending the use of small molecules in various therapeutic applications for future use.

APPENDICES

Appendix A. Modular Design for Minor Groove Binding and Recognition of Mixed Base Pair Sequences of DNA

Pu Guo, Ananya Paul, Arvind Kumar, Narinder K. Harika, Siming Wang, Abdelbasset A. Farahat, David W. Boykin, and W. David Wilson*

Guo, P., et al. *Chemical Communications.* **2017** 53, 10406-10409.

My contribution to this project was Molecular dynamics simulations and its analysis, and writing.

Appendix A.1 Abstract

The design and synthesis of compounds that target mixed, AT/GC, DNA sequences is described. The design concept connects two N-methyl-benzimidazole-thiophene single GC recognition units with a flexible linker that lets the compound fit the shape and twist of the DNA minor groove while covering a full turn of the double helix.

Appendix A.2 Introduction

Small, synthetic molecules that selectively target biological macromolecules in cells and induce specific responses, such as changes in gene expression, are a central goal of biomolecular compound design and synthesis research as well as therapeutic development (1-8). While there is significant progress in this area with proteins, there is limited progress in the variety of compound designs to target mixed base pair (bp) DNA sequences. Approaches for compounds that can target the DNA component of DNA-protein complexes, for example, would help remove this block to progress and provide an important step forward in the area (9-12). In order to effectively target DNA in cells, compounds with appropriate physical properties must be designed along with the development of synthetic approaches that are reasonable in cost and effort.

Heterocyclic-cation, minor-groove binders that can selectively bind to DNA have become a valuable resource in therapeutics, and compounds that successfully act on specific cancers, parasitic organisms and bacteria are now in animal and clinical testing as well as therapeutic use (1,13-15). These successful agents have provided proof of concept that selective and functional recognition of the DNA minor groove with a set of modules, combined in different ways for different sequences, is possible. The relatively simple agents available, however, primarily recognize only A·T bp sequences and have off-target effects (16).

Our approach to overcome this limitation is design and preparation of modular minor-groove, sequence-specific compounds that can recognize G·C in addition to A·T bp in longer DNA sequences than most current agents. The design strategy starts with heterocyclic cationic molecular modules that have effective DNA binding affinity in AT sequences, cell and nuclear uptake, reasonable syntheses, relatively low human toxicity and a range of physical properties to increase the chances of finding active compounds for the treatment of a variety of diseases (11). Using this

approach a thiophene-*N*-methylbenzimidazole (thiophene-*N*-MeBI) compound (DB2429, Figure A.1A) was designed according to the "σ-hole" preorganizing principle for macromolecular targeting (17). This effect is based on the presence of low-lying C−S σ* orbitals on the thiophene S that possess positive electrostatic potential for interaction with electron donating atoms such as the unsubstituted N in *N*-MeBI (18,19). Such rational control of the conformation of small molecules is a key concept of optimized molecular design for targeting macromolecules and can lead to significantly improved binding affinity and specificity that is independent of compound−DNA contacts. Compounds with furan or pyridine in place of the thiophene in DB2429 bind with less affinity and selectivity in support of the role of σ-hole interactions with the thiophene (17).

The critical question at this point, which is addressed in this report for the first time, is how to extend relatively simple compounds, such as DB2429, to selectively recognize longer, more complex DNA sequences with additional G·C bp. Our approach combines modular GC recognition units to prepare agents that can selectively recognize two G·C bps in DNA sequences such as $(A/T)_3$-(G/C)-$(A/T)_n$-(G/C)-$(A/T)_3$ where "n" can be 0-5 bps (Figure A.1). For the initial test of this approach, a flexible linker, –O-$(CH_2)_3$-O-, for DB2429-type GC recognition molecular modules was used. The design strategy includes enough flexibility to match the shape and curvature of the DNA minor groove and an appropriate length with modules spaced to bind to the GAAAC DNA sequence and a full turn of the DNA helix.

Our first transcription factor (TF)-DNA complex for targeting was the PU.1 TF that is of critical importance in the development of acute myeloid leukemia (AML). Compounds that bind AT or single G·C bp sequences have been successful at targeting the PU.1 promoter and inhibiting AML in cells (20). To increase the effectiveness of the TF-promoter DNA targeting approach, we

wish to target additional TFs, such as the Blimp-1 TF protein that impairs T cell function and is correlated with AML progression (21). Blimp-1 binds to a consensus GAAAC/G promoter sequence, (22,23) and agents that bind strongly to that sequence have the potential to inhibit the TF and provide new anti-AML activity. Scheme A.1 shows a retrosynthetic analysis which allows modular assembly of small molecules designed to bind to DNA sequences with two G·C bps. The target bis-amidino-*N*-methylbenzimidazolethiophenes are prepared from amidino-*N*-methylphenylenediamines by coupling with the appropriate bis-thiophenecarboxaldehydes in the presence of selected oxidizing agents. The required bis-aldehydes are obtained by linking the protected 5-(4-hydroxyphenyl) thiophene-2-carboxaldehdyes with 1, 3-dibromopropane employing standard Williamson Ether synthesis methodology. The protected 5-(4-hydroxyphenyl) thiophene-2-carboxaldehdyes are readily accessible using standard Suzuki coupling reactions between various substituted bromothiophenes and 4-hydroxyphenylboronic acid. A set of related compounds prepared using this approach are shown in Figure A.1B. Their synthesis routes are shown in Supporting Information, Scheme A.2. The resulting agents are particularly striking because relatively simple changes in chemical groups in AT specific compounds, the addition of specific H-bond accepting groups for G·C bp recognition, converts them into compounds (Figure A.1B) that strongly and specifically recognize mixed bps sequences of DNA much more strongly than the original AT sequences. This is the first example where heterocyclic cations linked in this manner have been systematically designed and prepared to recognize G·C bp in complex DNA sequences.

Appendix A.3 Material and Methods

All commercial reagents were used without purification. Melting points were determined on a Mel-Temp 3.0 melting point apparatus, and are uncorrected. TLC analysis was carried out on silica

gel 60 F254 precoated aluminum sheets using UV light for detection. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker 400 MHz spectrometer using the indicated solvents. Mass spectra were obtained from the Georgia State University Mass Spectrometry Laboratory, Atlanta, GA. The compounds reported as salts contain waters of hydration and in each case a water signal was noted in the ^{1}H-NMR spectra. Elemental analyses were performed by Atlantic Microlab Inc., Norcross, GA.

In the DNA thermal melting (T_m), circular dichroism (CD), fluorescence anisotropy, and electrospray ionization mass spectrometry (ESI-MS) experiments, the hairpin oligomer sequences were used as shown in Table A.1. In SPR experiments, 5′-biotin labeled hairpin DNA oligomers were used. All DNA oligomers were obtained from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) with reverse-phase HPLC purification and mass spectrometry characterization.

The buffer used in T_m, CD, and fluorescence experiments was 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4 (TNE 100). The biosensor-surface plasmon resonance (SPR) experiments were performed in filtered, degassed TNE 100 with 0.05% (v/v) surfactant P20. 50 mM ammonium acetate buffer with 10% MeOH was used in ESI-MS experiments.

Appendix A.3.1 UV-vis Thermal Melting (Tm)

DNA thermal melting experiments were performed on a Cary 300 Bio UV-vis spectrophotometer (Varian). The concentration of each hairpin DNA sequence was 3 μM in TNE 100 using 1 cm quartz cuvettes. The solutions of DNA and ligands were tested with the ratio of 2:1 [ligand] : [DNA]. All samples were increased to 95 °C and cooled down to 25 °C slowly before each experiment. The spectrophotometer was set at 260 nm with a 0.5 °C/min increase beginning at 25 °C, which is below the DNA melting temperature and ending above it at 95 °C. The absorbance of the buffer was subtracted, and a graph of normalized absorbance versus temperature was created

using KaleidaGraph 4.0 software. The ΔT_m values were calculated using a combination of the derivative function and estimation from the normalized graphs.

Appendix A.3.2 Biosensor-Surface Plasmon Resonance (SPR)

SPR measurements were performed with a four-channel Biacore T200 optical biosensor system (GE Healthcare, Inc., Piscataway, NJ). A streptavidin-derivatized (SA) CM5 sensor chip was prepared for use by conditioning with a series of 180 s injections of 1 M NaCl in 50 mM NaOH (activation buffer) followed by extensive washing with HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.05% P20, pH 7.4). Biotinylated-DNA samples (AAAATTTT, AAAAGTTTT, GAAG, GAAAC, GAAAAC, and GAAAAAC hairpins, Table A.1) of 25-30 nM were prepared in HBS buffer and immobilized on the flow cell surface by noncovalent capture as previously described (24). Flow cell 1 was left blank as a reference, while flow cells 2−4 were immobilized separately by manual injection of biotinylated-DNA stock solutions (flow rate of 1 μL/min) until the desired amount of DNA response units (RU) was obtained (250−300 RU). Ligand solutions were prepared with degassed and filtered TNE 100 with 0.05% (v/v) surfactant P20 by serial dilutions from a concentrated stock solution. Typically, a series of different ligand concentrations (2 nM to 500 nM) were injected over the DNA sensor chip at a flow rate of 100 μL/min for 180 s, followed by buffer flow for ligand dissociation (600−1800 s). After each cycle, the sensor chip surface was regenerated with a 10 mM glycine solution (pH 2.5) for 30 s followed by multiple buffer injections to yield a stable baseline for the following cycles. RU_{obs} was plotted as a function of free ligand concentration (C_{free}), and the equilibrium binding constants (K_A) were determined either with a one-site binding model, where $r = (RU_{obs}/RU_{max})$ represents the moles of bound compound/mol of DNA hairpin duplex and K is macroscopic binding constant.

$$\boldsymbol{r = K * C_{free}/1 + K * C_{free}} \qquad \textbf{(1)}$$

RU_{max} can be used as a fitting parameter, and the obtained value compared to the predicted maximal response per bound ligand can also be used to independently evaluate the stoichiometry (25). Kinetic analyses were performed by globally fitting the binding results for the entire concentration series using a standard 1:1 kinetic model with integrated mass transport-limited binding parameters as described previously (26).

Appendix A.3.3 Fluorescence anisotropy

Fluorescence spectra were recorded on a Cary Eclipse Spectrophotometer, with excitation and emission slit width typically fixed at 10, 10 nm. The free compound solutions at different concentrations were prepared in TNE 100, and DNA sequence aliquots were added from a concentrated stock. Titration spectra were collected after allowing an incubation time of 10 min. The excitation wavelength is based on molecular absorbance from UV-vis spectroscopy. Emission spectra were monitored at the fluorescence excitation wavelength at 25 °C.

Fluorescence anisotropy (r) measures the extent of polarization of the fluorescence emission of a system in solution when excited with polarized light. It is directly related to the rotational diffusion of the system and is low for small and flexible molecules (higher depolarization), and increases when larger complexes are formed (slower rotational movement, smaller depolarization) (27). Therefore, the change in fluorescence anisotropy can be used to monitor a binding reaction (28). This is particularly convenient for systems where no or small changes in fluorescence intensity is observed between the free and bound states. The fluorescence anisotropy is defined in the equation:

$$\boldsymbol{r = \frac{Ivv - GIvh}{Ivv + 2GIvh}}\ \boldsymbol{, G = \frac{Ihv}{Ihh}} \qquad \textbf{(2)}$$

where I_{vv} is the fluorescence emission intensity measured for vertically polarized excitation and vertically polarized emission. I_{vh} is the intensity measured for vertically polarized excitation and horizontally polarized emission, and G is a correction factor.

Depending on the parameters achieved from the previous step, add an appropriate concentration of ligands until the total intensity of emission wavelength is ~ 400. Record that concentration. Place blank cuvette containing TNE 100 in the instrument. Collect the spectrum using parameters obtained from the previous part, double check the emission spectra. The polarizer (Agilent Technologies, Manual Polarizer Accessory) was installed to the system and set the excitation and emission wavelength of compound same as the previous scan spectrum. The slit width fixed same as [10, 10 nm] and the average time of reading was set as 50 s. Read I_{vv}, I_{vh} of the buffer as a blank sample. The I_{hh}, I_{hv} of ligand only solution with the concentration same as the record was read to calculate the G of ligand. The I_{vv} and I_{vh} of each titration sample including ligand only were read to be made into scatter plot and fit the data to get the K_D value in KaleidaGraph 4.0 software.

Appendix A.3.4 Circular Dichroism (CD)

Circular dichroism experiments were performed on a Jasco J-810 CD spectrometer in 1 cm quartz cuvette at 25 °C. A buffer scan as a baseline was collected first in the same cuvette and subtracted from the scan of following samples. The hairpin DNA sequence GAAAC or AAAATTTT (5 μM),Table A.1, in TNE 100 was added to the cuvette prior to the titration experiments and then the compound was added to the DNA solution and incubated for 10 min to achieve equilibrium binding for the DNA-ligand complex formation. For each titration point, four spectra were averaged from 500 to 220 nm wavelength with scan speed 50 nm/min, with a response time of 1 s. Baseline-subtracted graphs were created using the KaleidaGraph 4.0 software.

Appendix A.3.5 Competition Electrospray Ionization Mass Spectrometry (ESI-MS)

Electrospray Ionization Mass Spectrometry (ESI-MS) analyses were performed on a Waters Q-TOF micro Mass Spectrometer (Waters Corporate, Milford, MA) equipped with an electrospray ionization source (ESI) in a negative ion mode. DNA sequences AAAATTTT, AAAAGTTTT, GAAC, GAAAC, GAAAAC and GAAAAAC, Table A.1, for ESI-MS experiments were purified by dialyzing it in 50 mM ammonium acetate buffer (pH 6.7) at 4 ºC with 3x buffer exchange. Test samples were prepared in 50 mM ammonium acetate with 10% v/v methanol at pH 6.7 and introduced into the ion source through direct infusion at 5 µl/min flow rate. The competitive experiments were done by mixing a ligand and DNAs with different sequences at different ratios. The instrument parameters were typically as follows: capillary voltage of 2800 V, sample cone voltage of 30 V, extraction cone voltage of 1.0 V, desolvation temperature of 70 °C, and source temperature of 100 °C. Nitrogen was used as nebulizing and drying gas. A multiply charged spectra were acquired through a full scan analysis at mass range from 300-2500 Da and then deconvoluted to the spectra presented. MassLynx 4.1 software was used for data acquisition and deconvolution.

Appendix A.3.6 Ab-Initio Calculations and Molecular Dynamic (MD) Simulation

Optimization and electrostatic potential calculations were performed for the DB2528 molecule using DFT/B3LYP theory with the 6-31+G* basis set in Gaussian 09 (Gaussian, Inc., 2009, Wallingford, CT) with Gauss-view 5.09 (29). Partial charges were derived using the RESP fitting method (Restrained Electrostatic potential) (30,31).

AMBER 14 (Assisted Model Building with Energy Refinement) software suite was used to perform molecular dynamic (MD) simulations using “ff14SB” and ff14SB/chi/ez (i.e., parmbsc0_chiOL4_ezOL1) force field modifications for DNA (32). Canonical *B*-form *ds*[(5’-CCAAAGAAACTTTGG-3’)(5’-CCAAAGTTTCTTTGG-3’)] DNA was built in Nucleic Acid

Builder (NAB) tool in AMBER. AMBER preparation and force field parameter files required to run molecular dynamic simulations for DB2528 molecule were produced using ANTECHAMBER (33). Specific atom types assigned for DB2528 molecule were adapted from the ff99 force field. Most of the force field parameters for DB2528 molecule were derived from the existing set of bonds, angles and dihedrals for the similar atom types in parm99 and GAFF force fields. Some dihedral angle parameters were obtained from previously reported parametrized data (34,35).

AutoDock Vina program was used to dock the DB2528 in the minor groove of DNA to obtain the initial structure for DB2528-DNA complex (36). MD simulations were performed in explicit solvation conditions where the DNA-DB2528 complex was placed in a truncated octahedron box filled with TIP3P water using xleap program in AMBER. Sodium ions were used to neutralize the system. A 10 Å cutoff was applied on all van der Waals interactions. The MD simulation was carried out using the Sander module with SHAKE algorithm applied to constrain all bonds. Initially, the system was relaxed with 500 steps of steepest-descent energy minimization. The temperature of the system was then increased from 0 K to 310 K for over 10 ps under constant-volume conditions. In the final step, the production run on the system was subsequently performed for 300 ns under NPT (constant-pressure) conditions. Coordinate file of DB2528-DNA complex along with water molecules in proximity is also attached (terminal bp is not included due to fraying at the ends of DNA).

Appendix A.4 Results and Discussion

Thermal melting (T_m) provides a rapid, evaluation screen of compounds with selected DNA sequences (Table A.1, A.2) (37,38) Binding to DNA gives an increase in the melting temperature that is related to binding affinity. Hairpin DNA oligomers, which have monomolecular melting transitions, were chosen for the screen. DB2429 (Figure A.1A) shows our approach of design of

single G recognition compounds and for linking these modular units to recognize more complex sequences. DB2528 and analogs are the first test of this concept and T_m results are shown in Table A.2. DB2528, which combines two DB2429 units with a flexible –O-$(CH_2)_3$-O- linker (Figure A.1B), does not bind well to a single G segment (AAAAGTTTT, ΔT_m = 2℃) but binds with a ΔT_m of 9℃ with the target two G·C bps sequence, GAAAC. It has lower T_m with longer, GAAAAC and GAAAAAC sequences (6℃). It is very encouraging that it shows no significant binding with an all AT sequence (ΔT_m = 1℃), and closer G·C bps (GC, GAC, GAAC), all with ΔT_m <2℃ (Table A.1, A.2, Figure A.2). The ΔT_m for GAAAC and CAAAG are essentially identical but the less symmetrical sequence GAAAG gives a 3℃ lower ΔT_m (Figure A.2). Modification of the amidine (DB2604 and 2614) or thiophene (DB2612 and 2614) groups can enhance interactions with the minor groove in cases where self-association limits DNA interactions, but the additions did not enhance binding with DB2528 (Figure A.1B, Table A.2). The replacement of *N*-MeBI in DB2528 with benzimidazole (BI) gives a more classical minor groove binder that shows strong binding with the AT sequences and weaker binding to the G sequences (not shown). This result clearly confirms the crucial importance of *N*-MeBI to the G·C bp recognition. Circular dichroism (CD) results with DB2528 and the GAAAC sequence show a strong positive induced CD peak at the compound absorption wavelength (Figure A.3). This result clearly indicates that the compound is a minor groove binder. Addition of a pure AT sequence to DB2528 results in very little spectral change in agreement with quite weak binding to pure AT sequence.

Biosensor-SPR methods provide an excellent way to quantitatively evaluate the interaction of small molecules with immobilized biomolecules (24). SPR provides sensitive, real-time progress of the binding reaction with binding affinity, kinetics, and stoichiometry of complex formation (25,39) Based on the T_m results, the interactions of DB2528 were evaluated by SPR

with G·C bp sequences (Table A.3). As can be seen in Figure A.4, DB2528 binds strongly with GAAAC and global kinetics fitting yielded a single binding site with a K_D of 5 nm, a rapid on-rate (k_a = 2.83±0.5×10^6 $M^{-1}s^{-1}$) and a very slow off-rate (k_d = 1.7±0.4×10^{-2} s^{-1}). The association and dissociation of DB2528 with the shorter GAAG sequence are quite fast with a high K_D = 149 nM. Pure AT and single G sequences were also studied by SPR, and it is encouraging that DB2528 has no detectable binding with either of them. These results indicate excellent selectivity and strong binding for a two G sequence that has a specific distance between the two G·C bps. As a test of the SPR binding results, the K_D was determined by a fluorescence anisotropy titration of the compound with the GAAAC DNA, and the K_D value is in agreement with the results from SPR (Table A.3, and Figure A.5) and validates the SPR results for binding affinity. Clearly, the combination of thiophene-*N*-MeBI and a flexible –O-$(CH_2)_3$-O- linker creates an effective module for strong two G·C bps specific recognition of a full turn of the DNA helix.

Competition ESI-mass spectrometry (MS) provides a direct analysis of relative binding affinity, stoichiometry and specificity for small molecule DNA complexes (40,41). The use of several DNA sequences simultaneously mixed with a compound creates a competitive binding environment for direct comparison of DNA interactions. The competition ESI-MS analysis results of DB2528 with DNA sequences AAAATTTT, AAAAGTTTT, and GAAAC, are shown in Figure A.6 A, B. The upper plot shows three peaks for the three DNA sequences. On addition of DB2528, the peak for GAAAC (9773) decreases with the simultaneous appearance of a new peak at *m/z*=10510 that is characteristic of a 1:1 GAAAC-DB2528 complex (Figure A.6B). There is no appearance of any complex peak with the other DNA sequences. Results for competition ESI-MS analysis of DB2528 with sequences GAAC, GAAAC, GAAAAC, and GAAAAAC are shown in Figure A.6 C, D to test the sequence specificity. The upper plot shows four peaks for the four

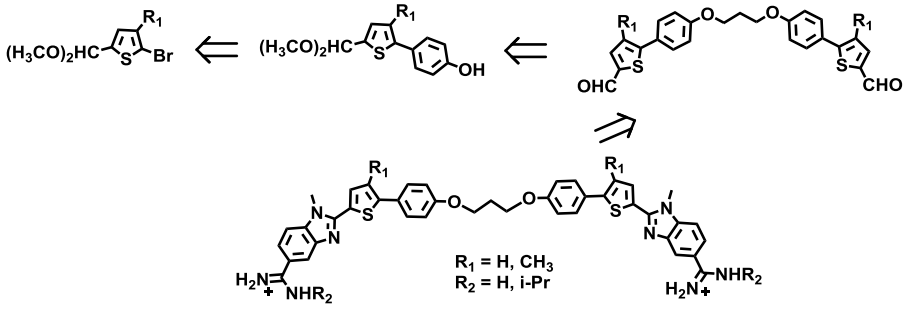
sequences. With the addition of DB2528, only the peak for GAAAC (9773) decreases with the simultaneous appearance of a new peak at *m/z*=10510 for a 1:1 GAAAC-DB2528 complex. There is no significant complex peak with any of the other tested sequences. In agreement with the binding results, the competition ESI-MS results clearly indicate that DB2528 binds with GAAAC very strongly and specifically in a 1:1 complex.

DB2528 was designed to H-bond with two G·C bps separated by three A·T bp and for terminal amidines H-bonds with AT. The compound binds strongly and specifically to this sequence but how does it fit to the minor groove and A·T/G·C bps? To answer this question, the molecular dynamics structures for the DB2528 complex with the minor groove of AGAAACT (Figure A.7) were determined with the AMBER 14 software suite and the ff99 force field. Force constants for DB2528 were added and the structure determined as previously described (34,35,42). The full view in Figure A.7A and Figure A.7A show DB2528 is able to match the curvature of the DNA minor groove and can cover a full turn of the double helix. The same view is shown in space filling model in Figure A.8B and shows the excellent contacts and van der Waals interactions that the compound makes with the minor groove molecular walls. A view of the complex structure at the upper part of the model is shown in Figure A.7B. This view shows strong H-bond between the N of *N*-MeBI and the G-NH that points out into the groove. There is also an H-bond between the amidine-NH that points to the floor of the groove and a T=O. The view in Figure A.7C shows the same H-bond pattern with the other end of the complex. The amidine-*N*-MeBI-thiophene-phenyl-O- can easily be seen tracking along the minor groove. Again, there are strong H-bonds from the G-NH to the N of *N*-MeBI and from the amidine to a T=O. As can be seen, both amidines are strongly hydrated and this certainly contributes to the energetics of binding and the fit to the groove

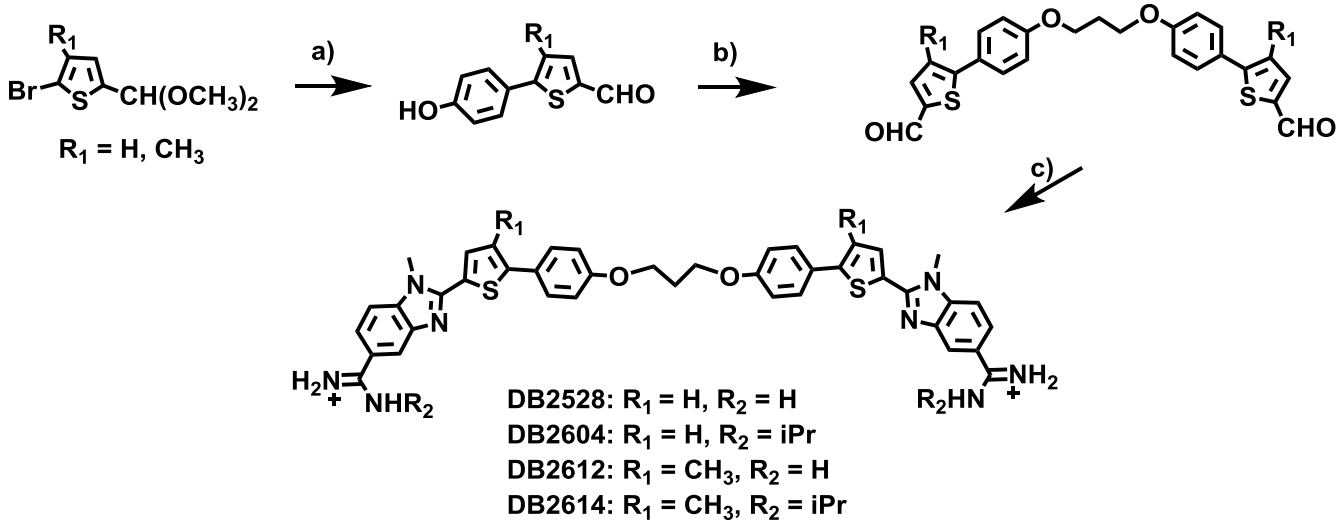
(Figure A.7A and Figure A.8). The compound curves around the groove and there is an exact indexing of two *N*-MeBI-G-NH and amidine-A·T bp.

In conclusion, the initial goal of this work was to link previously designed single G·C bp binding modules with a flexible linker for recognition of two G·C bps in the core sequence AGAAACT that has functional significance. With optimized design, synthesis and analysis a linked heterocyclic-diamidine has been obtained that has an excellent match to the DNA minor groove shape and registry of the compound with DNA functional groups. These results show, for the first time, what can be accomplished with modular compounds of this type in selective targeting of complex DNA. The compound has synthesis that is reasonable in cost and time and offers a promising route to develop a broad array of modular agents for control of gene expression.

This work was supported by National Institutes of Health Grant GM111749 to WDW and DWB.



Scheme A.1 Retrosynthetic analysis for bis-amidine-N-Methylbenzimidazole-thiophene compounds from bis-thiophenecarboxaldehydes that are prepared by linking arylthiophenecarboxaldehydes. Full synthetic details are in Supplementary Materials.



Scheme A.2. Synthesis of two G·C bps binders. Reagents and conditions: a) i. 4-hydroxyphenyl-boronic acid, $Pd(PPh_3)_4$, 2 M aqueous K_2CO_3, EtOH, reflux ii. HCl, H_2O. b) dibromoalkane, K_2CO_3, DMF, 80 °C. c) 3-amino-4-(methylamino)benzamidine or 3-amino-4-(methylamino)-N-isopropylbenzamidine, 1,4-benzoquinone, DMF, EtOH, reflux.

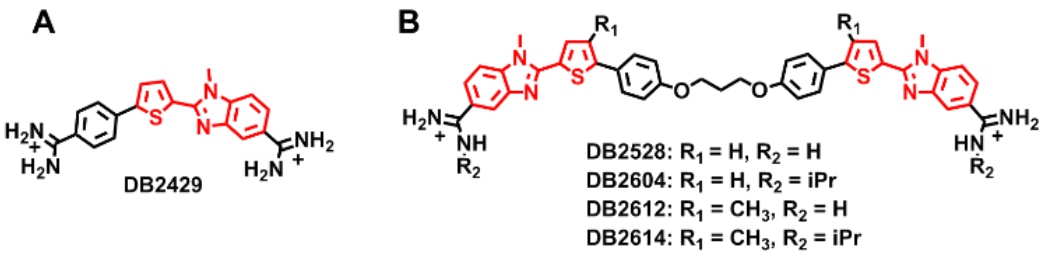


Figure. A.1 A) The compound DB2429 with best binding affinity and selectivity for a single G·C bp sequence found in our initial research. B) Chemical structure of new compounds designed to link two GC recognition modules. Red: G·C bp recognition module.

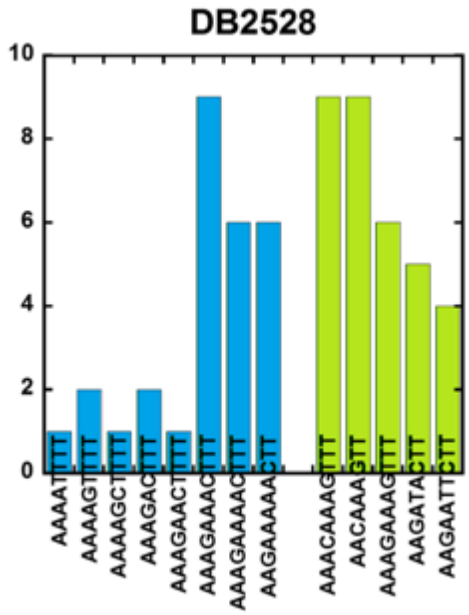


Figure A.2. Comparisons of thermal melting results (ΔT_m, °C) of DB2528 with pure AT and mixed DNA sequences. ΔT_m = T_m (the complex) - T_m (the native DNA). 3 µM DNA sequences were studied in TNE 100 with the ratio of 2:1 [ligand]:[DNA]. An average of two independent experiments with a reproducibility of ±0.5 °C.

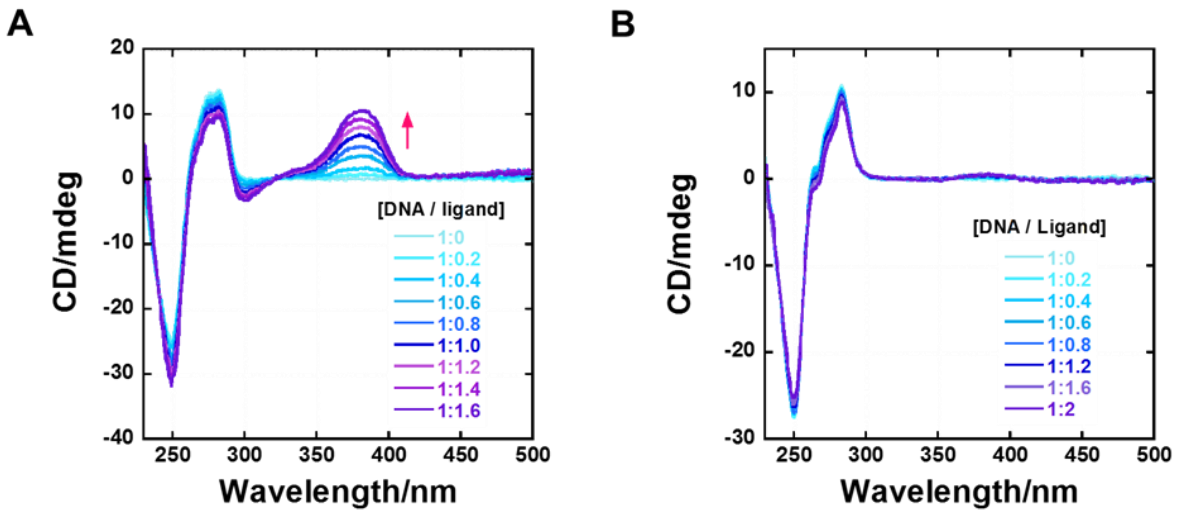


Figure A.3. Circular dichroism titration spectra of DB2528. DB2528 was titrated into 5 μM sequence A) AAAGAAACTTT; B) AAAATTTT (Table A.1). Arrow indicates the ligand induced spectral changes in DNA during titration.

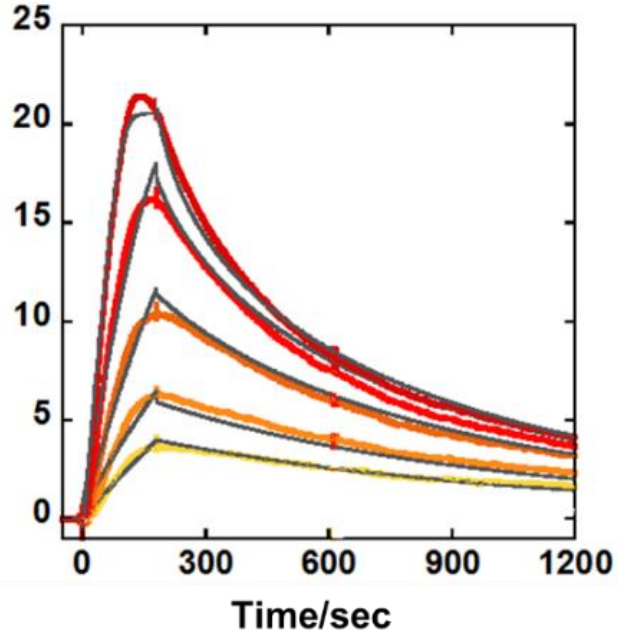


Figure. A.4 Representative SPR sensorgrams for DB2528 in the presence of GAAAC DNA, concentrations of DB2528 from bottom to top are 10, 15, 30, 50, and 100 nM. The solid black lines are best-fit values for the global kinetic fitting of the results with a single site function.

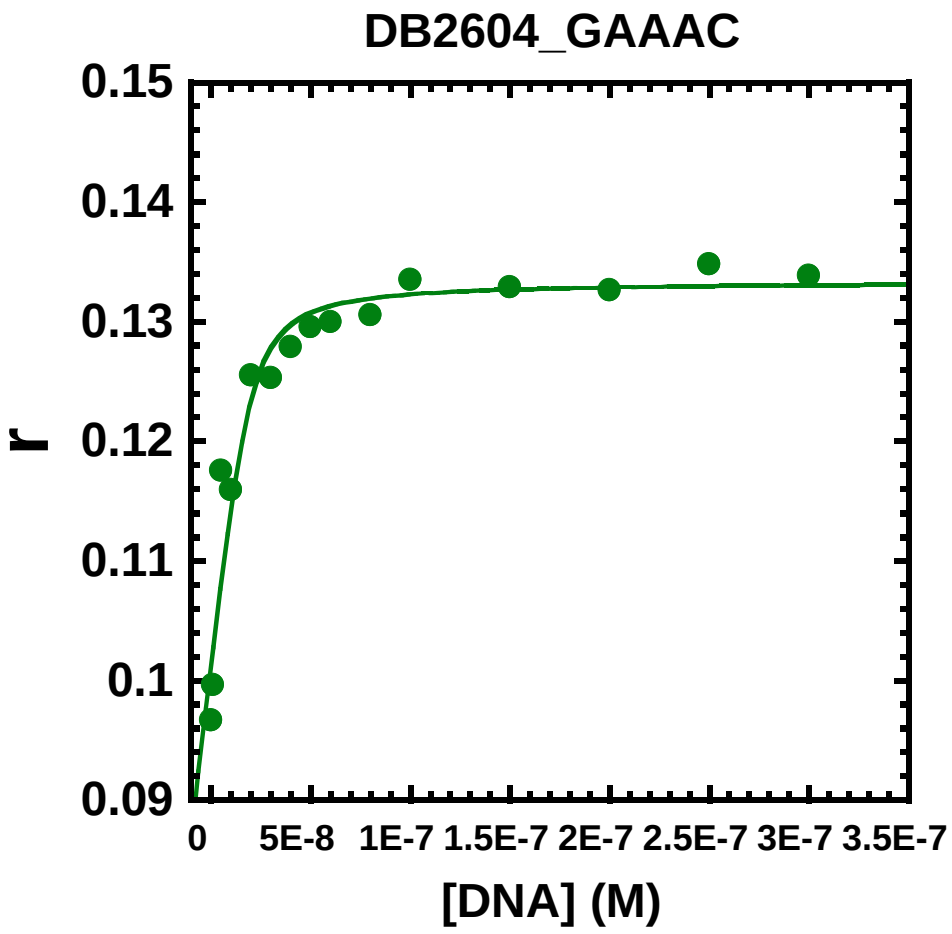


Figure A.5. Fluorescence Anisotropy (FA) binding curve between 20 nM DB2604 and sequence GAAAC in TNE 100. Ex_{max} = 365 nm; Em_{max} = 468 nm; slit width: [10, 10 nm]. The FA collection concentration points are 0, 1, 5, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300 nM.

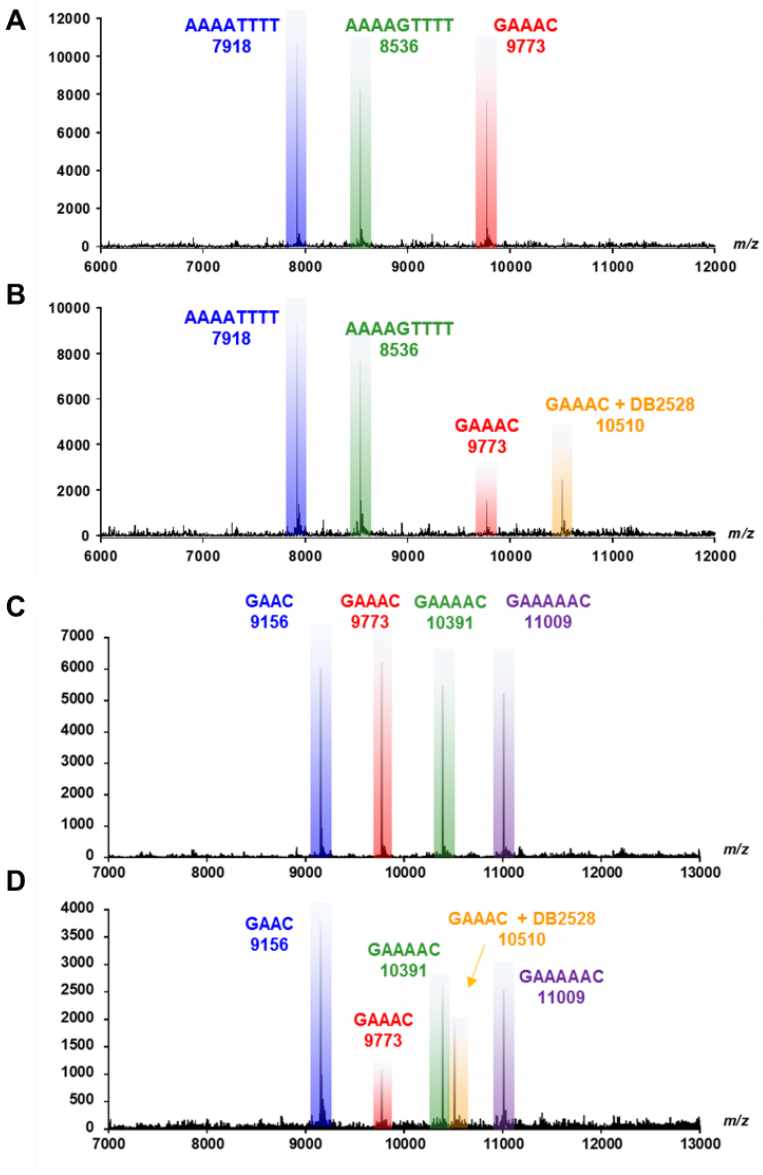


Figure. A.6 ESI-MS negative mode spectra of the competition binding of sequences A), B) AAAATTTT, AAAAGTTTT and GAAAC(10 µM each) ; C), D) GAAC, GAAAC, GAAAAC and GAAAAAC (10 µM each); with 40 µM DB2528 in buffer (50 mm ammonium acetate with 10% methanol(v/v), pH 6.8). A), C) The ESI-MS spectra of free DNA mixture. B), D) The ESI-MS spectra of DNA mixture with DB2528. The ESI-MS results shown here are deconvoluted spectra and molecular weights are shown with each peak. Full DNA sequences are in Table A.1.

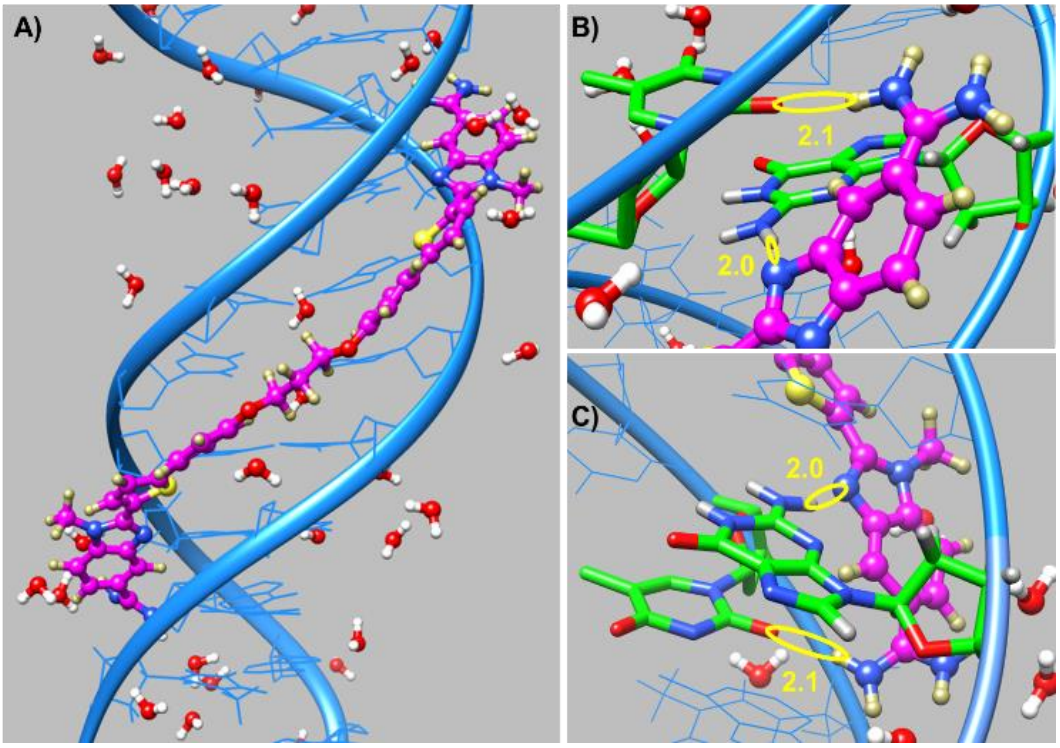


Figure. A.7. A) Minor groove view of the DB2528-DNA complex with proximal water molecules. B) Hydrogen bonding the upper G-NH and an opposite strand T-O2 with N of N-MeBI and NH of an amidine of DB2528 respectively are shown in yellow circles (in Å). C) View of hydrogen bonding of the lower G-NH and an opposite strand T-O2 with N of N-MeBI and NH of an amidine of DB2528, respectively is shown in yellow circles. Upper and lower are with respect to the model in A. DNA is represented in ribbon style (in blue) whereas G and T involved in H-bonding are shown in stick representation (B and C in green). DB2528 is shown in the ball and stick representation (B and C in magenta).

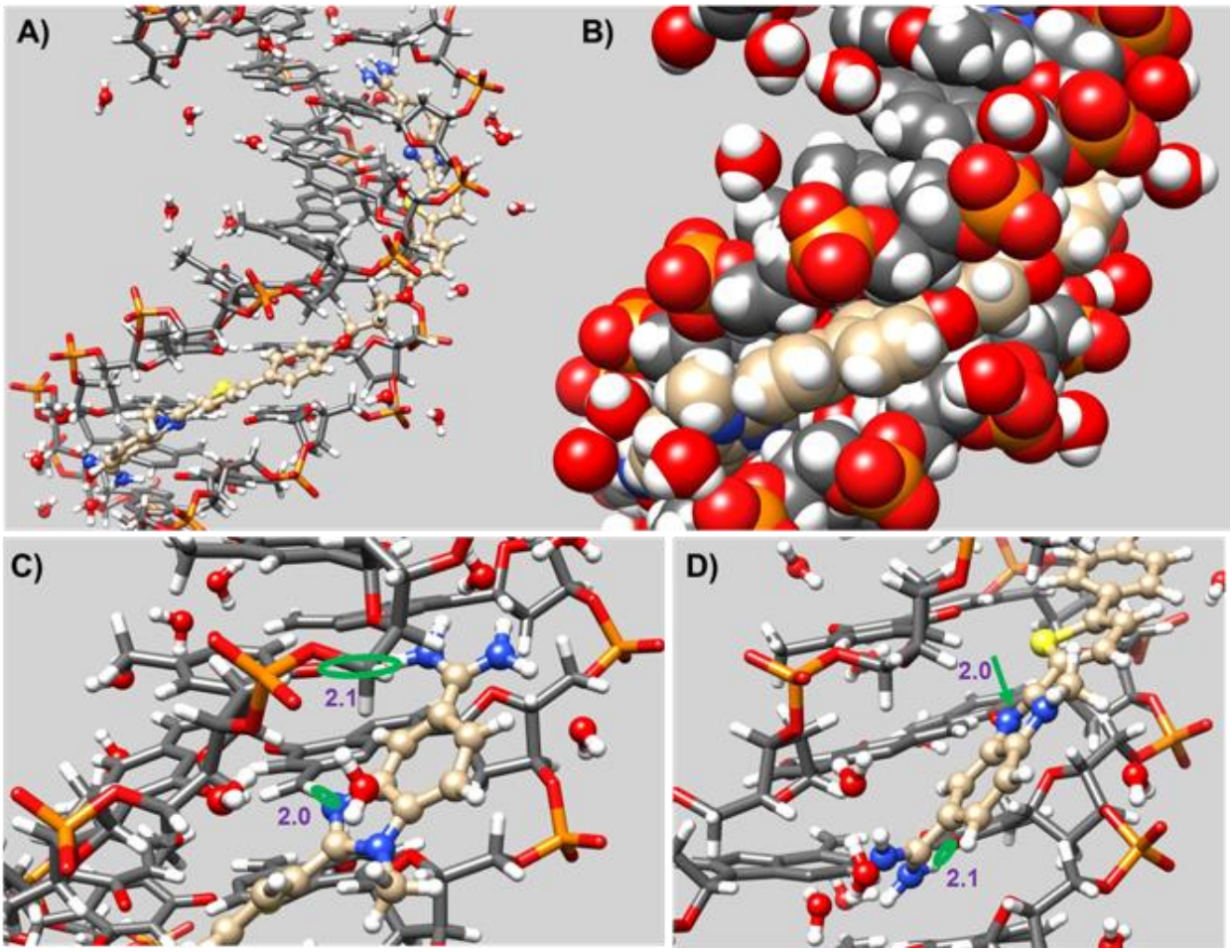


Figure A.8: A) Full view of the DB2528-DNA complex along with the water molecules in vicinity of complex. B) Space fill representation of DB2528-DNA complex. C) Upper part of the complex with H-bond of G6-NH and T26-O2 of DNA with N of N-MeBI and NH of amidine of DB2528 respectively is market in green (in Å). D) Lower part of the complex with H-bond interaction of G21-NH and T11-O2 of DNA with N of N-MeBI and NH of amidine respectively, is marked in the green (in Å).

Table A.1. All sequences used in this paper. [a]

Name	Sequences
AAAATTTT	5'-CCAAAATTTTGCTCTCAAAATTTTGG-3'
AAAAGTTTT	5'-CCAAAAGTTTTGCTCTCAAAACTTTTGG-3'
GC	5'-CCAAAAGCTTTTGCTCTCAAAAGCTTTTGG-3'
GAC	5'-CCAAAGACTTTGCTCTCAAAGTCTTTGG-3'
GAAC	5'-CCAAAGAACTTTGCTCTCAAAGTTCTTTGG-3'
GAAG	5'-CCAAAGAAGTTTGCTCTCAAACTTCTTTGG-3'
GAAAC	5'-CCAAAGAAACTTTGCTCTCAAAGTTTCTTTGG-3'
GAAAG	5'-CCAAAGAAAGTTTGCTCTCAAACTTTCTTTGG-3'
GATAC	5'-CCAAGATACTTGCTCTCAAGTATCTTGG-3'
CAAAG-1	5'-CCAAACAAAGTTTGCTCTCAAACTTTGTTTGG-3'
CAAAG-2	5'-CCAACAAAGTTGCTCTCAACTTTGTTGG-3'
GAAAAC	5'-CCAAAGAAAACTTTGCTCTCAAAGTTTTCTTTGG-3'
GAATTC	5'-CCAAGAATTCTTGCTCTCAAGAATTCTTGG-3'
GAAAAAC-1	5'-CCAAGAAAAACTTGCTCTCAAGTTTTTCTTGG-3'
GAAAAAC-2	5'-CCAAAGAAAAACTTTGCTCTCAAAGTTTTTCTTTGG-3'

[a]*The loop circle is shown underlined. The 5'-biotin labeled sequences were used in SPR experiments.*

Table A.2. Thermal melting studies ΔT_m (℃) of studied compounds with pure AT and mixed DNA sequences.[a]

ΔT_m (℃) **DNA T_m (℃)**[b]	**AAAA TTTT** **64**	**AAAA G TTTT** **67**	**AAAA GC TTTT** **69**	**AAA GA C TTT** **70**	**AAA GAA C TTT** **71**	**AAA GAAA C TTT** **69**	**AAA GAAAA C TTT** **70**	**AA GAAAAA C TT** **70**
DB2528	**1**	**2**	**1**	**2**	**1**	**9**	**6**	**6**
DB2604	**2**	**3**	**1**	**2**	**2**	**9**	**6**	**6**
DB2612	**1**	**<1**	**1**	**1**	**1**	**2**	**1**	**1**
DB2614	**2**	**2**	**1**	**1**	**2**	**6**	**5**	**5**

[a]*$\Delta T_m = T_m$ (the complex)-T_m (the native DNA). 3 μM DNA sequences were tested in TNE 100 with the ratio of 2:1 [ligand] : [DNA]. An average of two independent experiments with a reproducibility of ±0.5 ℃. Full DNA sequences can be seen in Table A.1.*

[b]*The values are the melting temperature of DNAs only used to calculate the relative data.*

Table A.3. Summary of binding affinity (K_D, nM) for the interaction of all test compounds with biotin-labeled DNA sequences using biosensor-SPR method. [a]

K_D (nM)	AAAA TTTT	AAAA G TTTT	AAA GAAAC TTT	AAA GAAAAC TTT	AA GAAAAAC TT	AAA GAAG TTT
DB2528	**NB**	**NB**	**5**	**6**	**--**	**149**
DB2604	**150**	**--**	**12**	**--**	**30**	**--**
DB2612	**--**	**--**	**50**	**--**	**100**	**--**
DB2614	**--**	**--**	**85**	**--**	**98**	**--**

[a]*All the results in this table were investigated in TNE 100 with 0.05% P20 at a 100 µL min^{-1} flow rate. "--" experiment not done, "NB" no binding. The listed binding affinities are an average of two independent experiments carried out with two different sensor chips and the values are reproducible within 10% experimental errors.*

Appendix A.5 References

1. S. Neidle, D. E. Thurston, *Nat. Rev. Cancer* 2005, **5**(4), 285–296.

2. L. A. Howell, M. Searcey, *ChemBioChem* 2009, **10**(13), 2139–2143.

3. J. Mantaj, P. J. M. Jackson, K. M. Rahman, D. E. Thurston, *Angew. Chem. Int. Ed.* 2017, **56**(2), 462-488.

4. J. B. Chaires, *Biopolymers* 2015, **103**(9), 473–479.

5. G. Savreux-Lenglet, S. Depauw, M. H. David-Cordonnier, *Int. J. Mol. Sci.* 2015, **16**(11):26555-26581.

6. M. P. Barrett, C. G. Gemmell, C. J. Suckling, *Pharmacol. Ther.* 2013, **139**(1):12-23.

7. C. Dardonville , J. J. N. Martínez, *Curr. Med. Chem.* 2017, **24**, DOI: 10.2174/0929867324666170623091522

8. H. Aikawa, A. Yano and K. Nakatani, *Org. Biomol. Chem.* 2017, **15**(6), 1313–1316.

9. J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vázquez, J. L. Mascareñas, *Nat. Commun.* 2013, **4**, 1874.

10. G. He, A. Tolic, J. K. Bashkin, G. M. K. Poon, *Nucleic Acids Res.* 2015, **43**(8), 4322-4331.

11. M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson, G. M. K. Poon, *Nucleic Acids Res.* 2014, **42**(2), 1379–1390.

12. C. R. Millan, F. J. Acosta-Reyes, L. Lagartera, G. U. Ebiloma, L. Lemgruber, J. J. Nué Martínez, N. Saperas, C. Dardonville, H. P. de Koning, J. L. Campos, *Nucleic Acids Res.* 2017, DOI 10.1093/nar/gkx521.

13. J. S. Kang, P. B. Dervan, *Q. Rev. Biophys.* 2015, **48**(4), 453–464.

14. X. Cai, P. J. Gray, D. D. Von Hoff, *Cancer Treat. Rev.* 2009, **35**(5), 437–450.

15. C. Lin, R. I. Mathad, Z. Zhang, N. Sidell, D. Yang, *Nucleic Acids Res.* 2014, **42**(9), 6012–6024.

16. R. Nanjunda, W. D. Wilson, in *Curr. Protoc. Nucleic Acid Chem.*, 2012, Chapter 8, Unit8.8.

17. P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin, W. D. Wilson, *Chem. Eur. J.* 2016, **22**(43), 15404–15412.

18. B. R. Beno, K.-S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *J. Med. Chem.* 2015, **58**(11), 4383–4438.

19. D. Tilly, F. Chevallier, F. Mongin, *Synthesis* 2016, **48**(2), 184–199.

20. I. Antony-Debré, A. Paul, J. Leite, K. Mitchell, H. M. Kim, L. A. Carvajal, T. I. Todorova, K. Huang, A. Kumar, A. A. Farahat, B. Bartholdy, S.-R. Narayanagari, J. Chen, A. Ambesi-Impiombato, A. A. Ferrando, I. Mantzaris, E. Gavathiotis, A. Verma, B. Will, D. W. Boykin, W. D. Wilson, G. M. K. Poon, U. Steidl, *J. Clin. Invest.* 2017, In Press, Manuscript No.: 92504-JCI-RG-1

21. L. Zhu, Y. Kong, J. Zhang, D. F. Claxton, W. C. Ehmann, W. B. Rybka, N. D. Palmisiano, M. Wang, B. Jia, M. Bayerl, T. D. Schell, R. J. Hohl, H. Zeng, H. Zheng, *J. Hematol. Oncol.* 2017, **10**(124), 1-13.

22. T. C. Kuo, K. L. Calame, *J. Immunol.* 2004, **173**(9), 5556–5563.

23. S. Deng, T. Yuan, X. Cheng, R. Jian, J. Jiang, *Mol. Biol. Rep.* 2010, **37**(8), 3747–3755.

24. R. Nanjunda, M. Munde, Y. Liu, W. Wilson, in (Ed.: Y. Wanunu), Methods for Studying Nucleic Acid/Drug Interactions, 2011, pp. 92–122.

25. Y. Liu, A. Kumar, S. Depauw, R. Nhili, M.-H. David-Cordonnier, M. P. Lee, M. A. Ismail, A. A. Farahat, M. Say, S. Chackal-Catoen, A. Batista-Parra, S. Neidle, D. W. Boykin, W. D. Wilson, *J. Am. Chem. Soc.* 2011, **133**(26), 10171–10183.

26. M. Munde, A. Kumar, R. Nhili, S. Depauw, M.-H. David-Cordonnier, M. A. Ismail, C. E. Stephens, A. A. Farahat, A. Batista-Parra, D. W. Boykin, et al., *J. Mol. Biol.* **2010**, *402*, 847–864.

27. W. A. Lea, A. Simeonov, *Expert Opinion on Drug Discovery* **2010**, *6*, 17–32.

28. A. G. Kozlov, R. Galletto, T. M. Lohman, in *Single-Stranded DNA Binding Proteins*, Humana Press, Totowa, NJ, **2012**, pp. 55–83.

29. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, *Gaussian 09, Revision B. 01. Gaussian 09, Revision B.01. Wallingford CT: Gaussian*, **2009**.

30. C. I. Bayly, P. Cieplak, W. Cornell, *J. Phys. Chem.* **1993**, *97*, 10269–10280.

31. U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129–145.

32. D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, *Amber 14*, NYU Press, **2014**.

33. J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

34. P. Athri, W. D. Wilson, *J. Am. Chem. Soc.* 2009, **131**(22), 7618–7625.

35. N. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza, J. Šponer, *J. Am. Chem. Soc.* 2003, **125**(7), 1759–1769.

36. O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455–461.

37. M. Munde, A. Kumar, P. Peixoto, S. Depauw, M. A. Ismail, A. A. Farahat, A. Paul, M. V. Say, M.-H. David-Cordonnier, D. W. Boykin, W. D. Wilson, *Biochemistry* 2014, **53**(7), 1218–1227.

38. X. Shi, J. B. Chaires, *Nucleic Acids Res.* 2006, **34**(2), e14.

39. R. D. Taylor, S. Asamitsu, T. Takenaka, M. Yamamoto, K. Hashiya, Y. Kawamoto, T. Bando, H. Nagase, H. Sugiyama, *Chem. Eur. J.* 2014, **20**(5), 1310–1317.

40. S. Laughlin, S. Wang, A. Kumar, A. A. Farahat, D. W. Boykin, W. D. Wilson, *Chem. Eur. J.* 2015, **21**(14), 5528–5539.

41. A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M.-H. David-Cordonnier, D. W. Boykin, W. D. Wilson, *Bioorg. Med. Chem. Lett.* 2015, **25**(21), 4927–4932.

42. S. Laughlin-Toth, E. K. Carter, I. Ivanov, W. D. Wilson, *Nucleic Acids Res.* 2017, **45**(3), 1297–1306.

Appendix B. Compound Shape Effects in Minor Groove Binding Affinity and Specificity for Mixed Sequence DNA

Pu Guo, Abdelbasset A. Farahat, Ananya Paul, Narinder K. Harika, David W. Boykin*, W. David Wilson*

Guo, P., et al. *Journal of the American Chemical Society*. **2018** *140*, 14761-14769.

My contribution to this project was includes molecular dynamics simulations and its analysis.

Appendix B.1 Abstract

AT specific heterocyclic cations that bind in the DNA duplex minor groove have had major successes as cell and nuclear stains and as therapeutic agents which can effectively enter human cells. Expanding the DNA sequence recognition capability of the minor groove compounds could also expand their therapeutic targets and have an impact in many areas, such as modulation of transcription factor biological activity. Success in the design of mixed sequence binding

compounds has been achieved with *N*-methylbenzimidazole (*N*-MeBI) thiophenes which are preorganized by a σ-hole interaction to fit the shape of the DNA minor groove and H-bond to the –NH of G·C base pairs that projects into the minor groove. Initial compounds bind strongly to a single G·C base pair in an AT context with a specificity ratio of 50 (K_d AT-GC/K_d AT) or less and this is somewhat low for biological use. We felt that modifications of compound shape could be used to probe local DNA microstructure in target mixed base pair sequences of DNA and potentially improve the compound binding selectivity. Modifications were made by increasing the size of the benzimidazole *N*-substituent, for example, by using *N*-isobutyl instead of *N*-methyl, and by changing the molecular twist by introducing substitutions at specific positions on the aromatic core of the compounds. In both cases we have been able to achieve a dramatic increase in binding specificity, including no detectible binding to pure AT sequences, without a significant loss in affinity to mixed base pair target sequences.

Appendix B.2 Introduction

Design and preparation of agents that can recognize mixed base pair (bp) sequences of DNA containing combinations of A·T and G·C bps have been of long term interest. Starting in the mid-1980s, this took a step forward with work in the Dickerson and Lown laboratories to redesign the minor groove binding polyamides netropsin (Nt.) and distamycin (Dst.), which specifically recognizing AT bp, into synthetic compounds that could also bind to a G·C bp (1,2). They proposed doing this by replacing one or more pyrrole groups in Nt and Dst with an imidazole group. The extra N in imidazole relative to pyrrole allows the imidazole group to form an H-bond with the G-NH_2 in the minor groove of a G·C bp (1-3). The proposal attracted attention in several additional laboratories and success in GC recognition was achieved by connecting two recognition units to make a hairpin polyamide (4-6). Although several groups have worked with these synthetic

polyamides, solution difficulties, aggregation, and poor cell uptake have limited their applications (7-9).

In contrast to netropsin (Nt) and distamycin (Dst), other types of AT specific heterocyclic cations, such as, Hoechst dyes, furamidine, DAPI, pentamidine, as well as numerous compounds that covalently attach to the minor groove, have had major successes as cell / nuclear stains and therapeutic agents (10-15). Interestingly, however, no systematic effort to redesign these agents into mixed bps recognition compounds, as with the polyamides, has been made. We felt that these promising and successful agents could potentially be redesigned to have significantly expanded use with broadened sequence recognition capability. We have now had considerable success in the design of agents that can recognize G·C bps as well as AT (16-18). Such new agents would have broad applications, for example, in targeting transcription factors to modulate gene expression (19,20).

For genes controlling critical cell functions, modifications of transcription factor (TF) activity, either directly or indirectly through steps such as enhancer activity, diseases such as leukemia or other cancers can result (20-22). Given the variety of functions and diseases controlled by TFs, a strong interest in targeting and controlling this activity with small molecules has developed (20). The problem is that TFs have evolved to bind high molecular weight nucleic acids, and not, in general, small molecules (23,24). This has made finding binding sites on TFs to interact strongly and specifically with small molecules, a difficult task. As a result, TFs are often defined as “undruggable” (23). Our concept is to target TF-DNA complexes by binding of designed agents to specific DNA sequences. To do this, we must design new types of DNA binding agents with broader sequence recognition capabilities than the classical AT-specific agents.

Incorporating the *N*-MeBI thiophene module into the heterocyclic cations, as in **Figure B.1**, initially gave DB2429 which specifically binds to a target G·C bp with flanking AT bps (18). DB2429 was expanded to DB2457 better cover the AT bps that flank the central GC. These compounds are a significant step forward in our molecular design and synthesis project for recognition of mixed bp DNA sequences (16,17,25). Analysis of the importance of the *N*-MeBI-thiophene module to binding specificity and selectively clearly shows that both the *N*-MeBI and thiophene are necessary for a strong, specific interaction (18). Replacing the thiophene with a furan or pyridine, for example, reduced both binding strength and specificity. The same was true for replacing the *N*-MeBI with a benzoxazole group which in principle has the same H-bonding capability as *N*-MeBI. Compounds with the *N*-Me substitution replaced by an NH group bind much more strongly to pure AT DNA sequences than to any sequence with a G·C bp (18,26). The remarkable reversal of binding specificity with a relatively small change in the overall compound structure is an important observation in new ideas for the design of agents to recognize specific DNA sequences.

The specificity ratio for binding to the single G·C bp sequence versus the pure AT sequence is about a factor of 10 for DB2429 while it is about 50 for the larger DB2457 which binds more strongly to all sequences (18). As described above, the strong decrease of binding affinity and selectivity for most structural changes with DB2429 and DB2457 raised the question of whether the selectivity ratio of these structures could be enhanced by different types of substitutions. Because of the critical importance of selectivity in targeting DNA, we have investigated new modifications of these initial compounds that have the potential to increase the selectivity. Initial synthetic modifications and DNA binding studies have been conducted with derivatives of the larger, stronger binding compound, DB2457. From the first crystal structures of a minor groove

binder to extensive structures of heterocyclic diamidines by the Neidle group, the wider minor groove in GC versus AT sequences has been observed (27). Based on these observations, the concept for the compound modifications is to target DNA microstructural variations, such as the wide minor groove in G·C bp containing regions. We reasoned that compounds with increased bulk might favor the wider, mixed bp sequences over pure AT. To test this idea the following general changes in the compound structure and functional groups were made with the goal of increasing specificity: (i) by increasing compound bulk primarily by replacing the *N*-Me group of *N*-MeBI with groups of increased size and (ii) by modifying the overall twist in the compound linked aromatic core structure by adding appropriate substituents (**Figure B.1**).

As we move to design new compounds to inhibit specific transcription factors, two things are very important: the selectivity ratio and the ability of the compounds to enter different cell types. At least some substituent changes at the *N*-BI position should have positive effects on both of these critical factors. We were able to successfully synthesize compounds with a wide variety of substituents at the *N*-BI position and three compounds with substituents on the aromatic groups of DB2457 that modified the effective steric bulk of the compounds. With these modifications, we were able to make significant improvements in compound specificity without a decrease in binding affinity, no increase in K_D.

Appendix B.3 Compound Design

The first set of compounds that were prepared had modifications at the *N-BI* position of DB2457 (**Table B.1** and **B.2**). We wanted to increase the length and complexity of the alkyl group, for example, methyl – ethyl – isopropyl (*i*-Pr) – isobutyl – neopentyl substituents. There are also linear groups with more polar substituents as in DB2728. To significantly change the shape of the alkyl group, compounds with cyclobutyl (DB2726), cyclopentyl (DB2714) and cyclohexyl (DB2727)

substituents were prepared. Finally, the chemistry of the substituent was changed dramatically with different phenyl (Ph) substituents on the BI group, *N*-Ph (DB2740) and *N*-ortho-methyl phenyl (DB2747). A control compound without an *N*-BI substituent, a benzimidazole, was also prepared. Such compounds have *N*-H groups pointed into the DNA minor groove and found to have very strong selectivity for pure AT sequences, relative to the compounds described above (26). Two compounds, with an *N*-MeBI and an *N*-*i*-Pr-BI were prepared with the thiophene S replaced by Se to evaluate the effect of size and changes in polarizability at that position (28,29). Four additional compounds were prepared with substitutions designed to increase the twist of the aromatic rings of the core structure. DB2754 has a Me on the thiophene, DB2753 has a Me on the Ph attached to the BI while DB2759 has a chloro and DB2762 has a trifluoromethyl group adjacent to the amidine on the same phenyl ring. We were unable to synthesize a compound with a Me adjacent to the amidine. Brief synthetic details are given here for the synthesis of the new compounds along with **Scheme B.1**. Full synthetic details for all compounds are given in the Supporting Information.

Appendix B.4 Materials and Methods

In the DNA thermal melting (T_m), circular dichroism (CD) experiments, the hairpin oligomer sequences were used as shown in **Figure B.1**. In SPR experiments, 5′-biotin labeled hairpin DNA oligomers were used. All DNA oligomers were obtained from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) with reverse-phase HPLC purification and mass spectrometry characterization.

The buffer used in T_m and CD experiments was 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4 (TNE 100). The biosensor-surface plasmon resonance (SPR) experiments were performed in filtered, degassed TNE 100 with 0.05% (v/v) surfactant P20.

Appendix B.4.1 UV-vis Thermal Melting (Tm)

DNA thermal melting experiments were performed on a Cary 300 Bio UV-vis spectrophotometer (Varian). The concentration of each hairpin DNA sequence was 3 μM in TNE 100 using 1 cm quartz cuvettes. The solutions of DNA and ligands were tested with the ratio of 2:1 [ligand] / [DNA]. All samples were increased to 95 °C and cooled down to 25 °C slowly before each experiment. The spectrophotometer was set at 260 nm with a 0.5 °C/min increase beginning at 25 °C, which is below the DNA melting temperature and ending above it at 95 °C. The absorbance of the buffer was subtracted, and a graph of normalized absorbance versus temperature was created using KaleidaGraph 4.0 software. The ΔT_m values were calculated using a combination of the derivative function and estimation from the normalized graphs.

Appendix B.4.2 Biosensor-Surface Plasmon Resonance (SPR)

SPR measurements were performed with a four-channel Biacore T200 optical biosensor system (GE Healthcare, Inc., Piscataway, NJ). A streptavidin-derivatized (SA) CM5 sensor chip was prepared for use by conditioning with a series of 180 s injections of 1 M NaCl in 50 mM NaOH (activation buffer) followed by extensive washing with HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.05% P20, pH 7.4). Biotinylated-DNA samples (AAATTT, AAAGTTTand AAAGCTTT hairpins, **Figure B.1**) of 25-30 nM were prepared in HBS buffer and immobilized on the flow cell surface by noncovalent capture. Flow cell 1 was left blank as a reference, while flow cells 2−4 were immobilized separately by manual injection of biotinylated-DNA stock solutions (flow rate of 1 μL/min) until the desired amount of DNA response units (RU) was obtained (250−300 RU). Ligand solutions were prepared with degassed and filtered TNE 100 with 0.05% (v/v) surfactant P20 by serial dilutions from a concentrated stock solution. Typically, a series of different ligand concentrations (2 nM to 500 nM) were injected over the DNA sensor

chip at a flow rate of 100 µL/min for 180 s, followed by buffer flow for ligand dissociation (600−1800 s). After each cycle, the sensor chip surface was regenerated with a 10 mM glycine solution (pH 2.5) for 30 s followed by multiple buffer injections to yield a stable baseline for the following cycles. RU_{obs} was plotted as a function of free ligand concentration (C_{free}), and the equilibrium binding constants (K_A) were determined either with a one-site binding model, where $r = (RU_{obs}/RU_{max})$ represents the moles of bound compound/mol of DNA hairpin duplex and K is macroscopic binding constant.

$$\boldsymbol{r = K * C_{free}/1 + K * C_{free}} \qquad \textbf{(1)}$$

RU_{max} can be used as a fitting parameter, and the obtained value compared to the predicted maximal response per bound ligand can also be used to independently evaluate the stoichiometry. Kinetic analyses were performed by globally fitting the binding results for the entire concentration series using a standard 1:1 kinetic model with integrated mass transport-limited binding parameters as described previously (30,31).

Appendix B.4.3 Circular Dichroism (CD)

Circular dichroism experiments were performed on a Jasco J-810 CD and Jasco J-1500 CD spectrometer in 1 cm quartz cuvette at 25 °C. A buffer scan as a baseline was collected first in the same cuvette and subtracted from the scan of following samples. The hairpin DNA sequence AAAGTTT (5 µM), **Figure B.1**, in TNE 100 was added to the cuvette prior to the titration experiments and then the compound was added to the DNA solution and incubated for 10 min to achieve equilibrium binding for the DNA-ligand complex formation. For each titration point, four spectra were averaged from 500 to 230 nm wavelength with scan speed of 50 nm/min, with a response time of 1 s. Baseline-subtracted graphs were created using the KaleidaGraph 4.0 software.

Appendix B.4.4 Structural Calculations

Molecular torsional angle map calculations of the compounds were performed in the Spartan'16 software. The "constrain dihedral" command was used with selected compounds to restrain four atoms to define the middle rotation bond and two terminal bonds which formed the dihedral as the calculation targets. The calculation range was set from 0° to 100° or 180° through 11 or 19 steps. Calculations were carried out with the energy profile method at ground state with density functional B3LYP/6-31G* in vacuum. After the calculations, the relative energy (rel. E) (kJ/mol) was displayed in a spreadsheet. The torsional angle map can be created with the constraint of the torsional angle as the X-axis and rel. E as the Y-axis by using KaleidaGraph 4.0 software.

Appendix B.4.5 Ab-initio Calculations and Molecular Dynamic (MD) Simulation

Optimization and electrostatic potential calculations were performed for the DB2708 molecule using DFT/B3LYP theory with the 6-31+G* basis set in Gaussian 09 (Gaussian, Inc., 2009, Wallingford, CT) with Gauss-view 5.09 (32). Partial charges were derived using the RESP fitting method (Restrained Electrostatic potential) (33,34). AMBER 14 (Assisted Model Building with Energy Refinement) software suite was used to perform molecular dynamic (MD) simulations (35). Canonical *B*-form *ds*[(5'-CCAAAGTTTGG-3')(5'-CCAAACTTTGG-3')] DNA was built in Nucleic Acid Builder (NAB) tool in AMBER. AMBER preparation and force field parameter files required to run molecular dynamic simulations for DB2708 molecule were produced using ANTECHAMBER (36). Specific atom types assigned for DB2708 molecule were adapted from the ff99 force field. Most of the force field parameters for DB2708 molecule were derived from the existing set of bonds, angles and dihedrals for the similar atom types in parm99 and GAFF force fields (25). Some dihedral angle parameters were obtained from previously reported parametrized data (37,38). Parameters of DB2708 in frcmod file are also listed.

AutoDock Vina program was used to dock the DB2528 in the minor groove of DNA to obtain the initial structure for DB2708-DNA complex (39). MD simulations were performed in explicit solvation conditions where the DNA-DB2708 complex was placed in a truncated octahedron box filled with TIP3P water using xleap program in AMBER. Sodium ions were used to neutralize the system. A 10 Å cutoff was applied on all van der Waals interactions. The MD simulation was carried out using the Sander module with SHAKE algorithm applied to constrain all bonds. Initially, the system was relaxed with 500 steps of steepest-descent energy minimization. The temperature of the system was then increased from 0 K to 310 K for over 10 ps under constant-volume conditions. In the final step, the production run on the system was subsequently performed for 500 ns under NPT (constant-pressure) conditions. Coordinate file of DB2708-DNA complex along with water molecules in proximity is also attached.

Appendix B.5 Results and Discussion

Appendix B.5.1Chemistry

The general synthetic approach to the target diamidine compounds **7** is outlined in Scheme 1. Nucleophilic aromatic substitution reactions between 4-bromo-1-fluoro-2-nitrobenzene **1** and various aliphatic and aromatic amines yields the 4-bromo-*N*-alkyl and *N*-aryl-2-nitrobenzenes **2** in reasonable yields (41-79%), which ultimately leads to the *N*-alkyl and *N*-aryl benzimidazoles. The nucleophilic aromatic substitution reaction with aliphatic amines was achieved at room temperature in ethanol whereas the less basic aryl amines required heating in dimethylacetamide in the presence of Cs_2CO_3 at 160^{o}C. Suzuki-Miyaura coupling of 4-cyanophenylboronic acids with the various 4-bromo-*N*-alkyl and *N*-aryl-2-nitrobenzenes under standard conditions provides the nitro biphenyl analogues **3** in good yields (51-79%). The arylphenylene diamines **4** required for formation of the *N*-substituted benzimidazoles were obtained by stannous chloride reduction of

the nitro groups of **3** and were subsequently allowed to react directly without characterization. Sodium metabisulfite mediated the oxidative condensation and cyclization of the phenylene diamines **4** with the previously reported (18) 5-(cyanophenyl)-2-formylthiophenes **5** to yield the bis-nitriles **6** in acceptable yields (39-76%). Lithium bis(trimethylsilyl)amide in THF reacts with the bis-nitriles **6** at room temperature to yield, after acidic work-up, the diamidines **7** generally in acceptable yields (30-79%).

Appendix B.5.2 DNA thermal melting: Screening for mixed sequence binding

Changes in DNA thermal melting temperature (T_m) provide a rapid way for initial ranking of compounds for binding affinity with different DNA sequences. We have used the three selected sequences shown in **Figure B.1** for comparative experiments. Most heterocyclic cations prepared to this time have been pure AT sequence binders and a DNA with the binding site -AAATTT- was used to test for pure AT sequence binding affinity. The target DNA of primary interest has a single G·C bp with the binding sequence –AAAGTTT-. The third sequence with two G·C bps was also used to test for binding affinity and selectivity versus AAAGTTT.

The reference compound, DB2457, with an *N*-Me substituent was previously reported (18) and has preferential binding to the single G·C bp sequence, the ΔT_m is 14℃, with lower ΔT_m and weaker binding to the pure AT and two G·C bps sequences, 6 and 5℃, respectively. The compounds with the *N*-ethyl and *N*-*i*-Pr substituents had similar binding to AAAGTTT and weaker binding to the pure AT sequence for improved selectivity as desired. At the level of evaluation of T_m, most of the other derivatives at the *N*-X position had very weak binding to the pure AT sequence but variable binding to the single and double G·C bps sequences. The *t*-butyl derivative, for example, had significantly weaker binding to AAATTT but slightly weaker binding to AAAGTTT. DB2753 and DB2754, methyl-substituted aromatic compounds, had much weaker

binding to all sequences. A –Cl substituent ortho to an amidine group, DB2759, on the other hand, gave stronger binding to AAAGTTT than the aromatic methyl compounds and very low binding to the pure AT sequence. A trifluoromethyl substituent in the same position had significantly weaker binding to both the single G·C bp and pure AT sequences. Two compounds were prepared with S in thiophene replaced with Se, but they showed no significant improvement in K_D for AAAGTTT or in selectivity. The compounds will be of a long term interest, however, in cell uptake studies.

Appendix B.5.3 Biosensor-SPR: Methods for quantitative binding

A biacore sensorchip (CM5) was functionalized with streptavidin and used to immobilize 5'-biotin labeled DNA sequences (**Figure B.1**) in flow cells 2-4 and flow cell 1 was left as a blank, for background subtraction. With different compounds in the flow solution, we were able to determine comparative binding constants for all of the derivatives (**Table B.2**). Example sensorgrams are shown for representative compounds with the different DNAs in **Figure B.2**. With the AAATTT and AAAGCTTT sequences, all compounds showed significantly weaker AAATTT binding than with AAAGTTT with generally fast on and off rates. No kinetics and limited K_D values could be accurately determined for those DNA sequences. With the AAAGTTT sequence, however, excellent sensorgrams were obtained at low concentrations for most derivatives (**Figure B.2**). Kinetics fits with a one site model were excellent and allowed a determination of K_D values (**Table B.2**). Kinetics fits were generally required to determine K_D values since many sensorgrams did not reach a steady-state level, especially at the lower concentrations.

DB2429, the original compound (**Figure B.1**), has a K_D for AAAGTTT of 50 nM with 10 fold higher KD value for AAATTT and 20 fold higher K_D value for AAAGCTTT (18). The parent *N*-Me compound in this study, DB2457, has a 10-fold lower K_D value for AAAGTTT of 4 nM

with 50 fold higher K_D values for AAATTT and AAAGCTTT relative to DB2429. The *N*-ethyl and *N*-*i*-Pr compound have similar K_D values for single G·C bp sequence but a higher value for AAATTT with a selectivity ratio of K_D s of near 250. They have similar selectivity for GC to the *N*-Me compound. The *N*-isobutyl (DB2711) has slightly weaker binding to the single G·C bp sequence (K_D = 13 nM) but high selectivity with no measurable binding to AAATTT and AAAGCTTT sequences. This compound with its high selectivity in binding validates the idea of *N*-substitution to enhance selectivity, a key feature for designing any new potential drug molecules in this series. The *n*-butyl compound (DB2715) has a 10 nM K_D with AAAGTTT and a 100 times weaker value to AAAGCTTT. Increasing the size of the *N*-alkyl substituent to *N*-neopentyl (DB2718), decreased affinity for binding to the AAAGTTT sequence with a 31 nM K_D and lower selectivity to the GC sequence. Surprisingly, when an O atom was introduced into the alkyl group (DB2728 and DB2764), similar binding to the single G·C bp sequence was observed, relative to the butyl compound, but with stronger binding to the GC sequence. Several cyclic alkyl substituted compounds were studied with quite variable selectivity ratios. The *N*-cyclopentyl compound (DB2714) has slightly weaker binding to the single G·C bp sequence (K_D = 9 nM) but with high selectivity, near 150 for GC and negligible binding to AAATTT that was too weak to evaluate accurately (see **Figure B.2**). The cyclobutyl compound, DB2726, with a relatively small cyclic substituent, however, has lower selectivity and behaves in a similar manner to the *i*-Pr derivative. DB2727 and DB2738 with larger cyclic alkyl groups also have weaker binding to the single G·C bp sequence but improved selectivity. The two *N*-aromatic compounds have excellent binding to AAAGTTT with no detectable binding to AAATTT but surprisingly strong binding to the AAAGCTTT sequence. According to the selectivity ratios shown in **Table B.2**, DB2711 and DB2759 have the best binding selectivity for the sequence AAAGTTT comparing to AAATTT

and AAAGCTTT and maintain relatively strong binding to AAAGTTT (K_D < 15nM). Moreover, DB2714 and DB2715 show AAAGTTT selectivity ratios all more than 100 fold higher for AAAGCTTT. These four compounds with the modifications based on DB2457 have good binding with AAAGTTT and excellent sequence specificity. Excitingly, except for the *N*-Me substituted compound DB2457 and DB2673, other compounds show very high selectivity (at least 250 fold) for AAAGTTT comparing to the pure AT sequence.

As with the T_m results, the Se compounds did not have the best K_D for AAAGTTT or the best selectivity. The four compounds with substitutions on the aromatic core system and an *N*-*i*-Pr substituent have weaker binding than the parent *N*-*i*-Pr but no detectible binding to the AAATTT and GC sequences. These results make the chloro derivative, DB2759, and the *N*-isobutyl, DB2711 very attractive for further development since they have relatively favorable binding to AAAGTTT with a K_D of around 14 nM and high selectivity with no detectible binding to the pure AT or two G·C bps sequences.

The increasing size of the *N*-alkyl groups results in similar or decreasing binding affinity with sequence AAAGTTT. Fortunately, the binding to reference pure AT and AAAGCTTT sequences decreased to a larger extent with many of the compounds (**Table B.2**). A moderate size of *N*-substituted groups can give the best target sequence binding and selectivity. Since the micro-environment such as width and depth of DNA minor groove change according to the different A·T and G·C bp combinations, (40-42) the size and stereo structure of binding ligands can strongly affect sequence specificity. Regarding the width of the minor groove of three studied sequences, the order is AAAGCTTT > AAAGTTT > AAATTT at the center sequence (40-42). Moderately sized substitutions such as i-Pr, cyclopentyl can target the single G·C bp sequence specifically. The aromatic substitutions may match the two G·C bps sequence microstructure. Although they

bind to AAAGTTT strongly but in many cases also with stronger binding to AAAGCTTT. Each of the compounds can also induce change in the minor groove structure of each DNA and at this time an experimental analysis is the best way to find the optimum compound for affinity and selectivity. The most important finding, however, is that modification of DB2457 can dramatically increase binding selectivity.

Appendix B.5.4 Biosensor-SPR Kinetics

A selection of compounds with a range of K_D values and a variety of chemical substitutions and structures was used to evaluate the kinetics of association, k_a, and dissociation, k_d, for the AAAGTTT sequence. As can be seen in **Figure B.2**, the binding kinetics were generally too rapid to fit accurately with the AAATTT and AAAGCTTT sequences. The kinetics results for AAAGTTT are summarized in **Table B.3** and **Figure B.3**. For the *N*-substituted compounds with an unsubstituted aromatic system, the K_D values with AAAGTTT in **Table B.3** cover a narrow range, from around 4 to 13 nM with only two compounds above 10 nM. The kinetics values also cover a relatively small range for k_a, from around 2 to 5 x 10^6 $m^{-1}s^{-1}$. The k_d values cover a somewhat larger range, from 0.008 to 0.05 s^{-1} but as seen in **Figure B.3**, a general correlation of the sets of values accounts for the small variation in K_D. The *N*-methyl (DB2457) and *N*-Ph (DB2740) compounds have similar low k_a values, around 2 x 10^6 $m^{-1}s^{-1}$. They also have similar low k_d values that give them strong binding affinity. Within the compounds in **Table B.3**, DB2715 with *n*-butyl and DB2711 with isobutyl represent the fastest association and disassociation rates, 4.6 x 10^6 $m^{-1}s^{-1}$ and 0.0444 s^{-1}, 4.1 x 10^6 $m^{-1}s^{-1}$ and 0.0516 s^{-1}, respectively. Moreover, they show the weakest binding (12.6 and 9.6 nM) to AAAGTTT among the compounds with an unsubstituted aromatic system. These results suggest that the butyl substitutions interfere with binding in the minor groove giving them fast but weak association and rapid dissociation with a relatively high

K_D. DB2764 with substitution of 2-methoxyethyl, which is the approximately same length as *n*-butyl (DB2715), shows a similar K_D value, 11 nM. DB2708 with *N*-*i*-Pr has similar on and off-rates as the parent, DB2457, to result in the same binding strength with AAAGTTT. DB2728 (3-methoxypropyl), is one methylene longer than DB2764 and shows a faster on-rate than DB2764. It has nearly the same off-rate and slightly stronger binding ability with single G·C bp sequence. The Se *N*-*i*-Pr derivative DB2712 has a k_a value similar to the thiophene but a higher k_d and as a result, a K_D that is around 2 fold weaker. Clearly, the larger, more polarizable Se offer no advantage. The cyclic alkyl substitution, cyclopentyl (DB2714), has moderate kinetics and K_D values (**Table B.3**). A methyl-substituted phenyl, *N*-*i*-Pr derivative, DB2753, has much weaker binding, K_D > 100nM, that is due to a very low k_a of 0.073 x 10^6 $m^{-1}s^{-1}$. The k_d for this compound, however, is in the range of other compounds in **Table B.3**. The core aromatic system of significantly twisted due to repulsion between the –methyl substituent and the adjacent phenyl ring. To bind into the minor groove, the twist must be reduced at an energy cost that slows association with lowering of k_a and an increase in K_D. Once bound it appears that the compound can form interactions similar to the compounds with an unsubstituted aromatic system to give k_d in the same range.

To help compare the kinetics data of the selected compounds, the relationship between on-rates (k_a) and off-rates (k_d) is plotted in **Figure B.3**. Green dash lines show the binding affinities (K_D). The parent compound DB2457 and the strongest binding compounds, DB2740 and DB2708, are located in a K_D line with a slope around 4 nM. DB2708 with *N*-*i*-Pr substitution has slightly faster association and disassociation because of its optimum structure matching the AAAGTTT minor groove. We can clearly see six compounds with different, but similar sized *N*-alkyl

substitutions that are distributed around the 10 nM green dash line. The on-rates and off-rates are all in a limited range.

Appendix B.5.5 CD binding mode

CD titration experiments are an effective and convenient method of evaluating the binding mode and the saturation limit for compounds binding with DNA sequences. CD spectra monitor the asymmetric environment of the compounds binding to DNA and therefore can be used to obtain information on the binding mode (43). There are no CD signals for the free compounds but on the addition of the compounds into DNA, substantial positive CD signals (ICD) arose in the absorption region between 300 and 450 nm. These positive ICD signals indicate a minor groove binding mode by these ligands, as expected from their structures. A monotonic increase of ICD signals at the range of 330 to 440 nm was observed upon an incremental addition of DB2708, DB2740, DB2759 and DB2754 to the AAAGTTT sequence. As can be seen from **Figure B.4**, all studied compounds from strong to moderate binders bind in the minor groove of the AAGTTT sequences with a 1:1 stoichiometry, in agreement with SPR results. In summary, the CD titration results confirm a minor groove binding mode for the compounds studied in **Figure B.1**.

Appendix B.5.6 Molecular Structure: Calculations

To help to understand the large differences in binding affinities between the control compound DB2708, with no substitutions on the core aromatic system, and the substituted compounds DB2753, DB2754, DB 2759 and DB2762, torsional angle maps were constructed in the *SPARTAN 16* software package (44). (**Figure B.5**). DB2708 has the lowest energy difference between the minimum energy torsional angle and 0° for both the Ph-amidine, the Ph-BI bond and the thiophene-Ph. DB2759 with a Ph-Cl substituent adjacent to the amidine has a slightly higher energy difference between the minimum energy torsional angle and 0° than the Ph-amidine bond

of DB2708 and this results in a slightly higher K_D. However, DB2762 with CF_3-Ph substitution shows a higher energy difference between the minimum energy torsional angle and 0°, up to 2.5 times higher than DB2708, and much weaker binding with the target sequence compared to DB2708. DB2753 with a methyl-Ph substituent adjacent to BI has the highest K_D value of the entire set of compounds and the highest energy difference between the minimum energy torsional angle and 0° for the Ph-BI bond. The torsional map of the thiophene-Ph bond of DB2708 shows 0° as the lowest energy state which is reasonable for the 5-atom and 6-atom ring contacts. DB2754 with a methyl-thiophene substituent adjacent to Ph-Am shows a higher K_D value and a higher energy difference for the thiophene-Ph. If we assume that the maximum torsional angle for optimum binding to DNA should be around 10° or less, these angles and energy differences are very important. To reduce the Ph-amidine bond to near 10° in DB2759 requires around 7 kJ/mol while to reduce the Ph-amidine bond to near 10° in DB2762 requires about twice as much energy. For the Ph-BI bond of DB2708 and DB2753, the energy of reducing to near 10° in DB2753 is about 5 times higher than with DB2708. It seems highly likely that these conformational energy costs are responsible for a least part of large binding constant differences for DB2708, DB2759, DB2762, DB2753, and DB2754. It should be noted that choosing another minimum torsional angle, such as 0° or 5°, leads to the same conclusions.

Appendix B.5.7 Molecular Structure: Molecular Dynamics

To help understand the basis of molecular recognition of DNA sequences with a single G·C bp in an AT context, molecular dynamics (MD) simulations for the representative compound, DB2708, were conducted in the AMBER 14 software suite (35). Force constants for DB2708 were determined as described in the methods section and added to the ff99 force field for the simulations. After 500ns of MD simulation, 25 PDB structures for the complex of DB2708 in the

AAAGTTT binding site (**Figure B.6**) were collected. Structures were evaluated in detail every 20ns (25 structures for 500ns) to determine what major features of the DB2708-DNA complex are primarily responsible for the excellent stability (K_D) and specificity ratios (K_D of AAAGTTT/ K_D of AAATTT). There are three optimum H-bonds in the complex. Both amidine groups form –N-H to T=O H-bonds (**Figure B.5B**) that are an average of 2.8Å in length (SD = 0.1Å). The amidines also form numerous highly dynamic H-bonds to water molecules that move in and out of the minor groove (45). These water molecules frequently also form H-bonds with AT bps at the floor of the minor groove and help link the compound to the specific binding site in the groove. The third strong H-bond is from the central G-NH that projects into the minor groove to the imidazole N in the BI group of DB2708, **Figure B.5C**, to account for much of the binding selectivity of DB2708. Additional selectivity in binding is provided by the –CH group of the six-member ring of BI that points into the minor groove. This -CH forms a dynamic close interaction with the –C=O of the dC base of the central G·C bp as well as to a T=O on an adjacent AT bp (**Figure B.5C**).

Additional direct stabilizing interactions are formed by –CH groups that point to the floor of the minor groove from the two phenyls of DB2708. These –CH groups form significant dynamic interactions with A-N3 groups on the bases at the floor of the groove (see **Figure B.5B** for an example). DB2708 tracks optimally along the minor groove with appropriate twist to match the minor groove curvature. Extensive interactions are formed by the conjugated aromatic-system of DB2708 with the sugar-phosphate walls of the minor groove. In addition to H-bonds with the AT bps, the amidines of DB2708 also form electrostatic interactions with the backbone phosphates that result in Na^+ release and an entropy increase that is another stabilizing component, along with water release from the minor groove, on complex formation. The sum of these numerous, relatively weak, stabilizing interactions results in strong binding of DB2708, a K_D of 4nM (**Table B.2**) with

a ΔG of binding of 11.5 kcal/mol. The *i*-Pr group of DB2708 forms a tight interaction with the DNA backbone in the G·C bp region. In an AAATTT sequence, the *i*-Pr group would have increased steric hindrance to increase the KD for AAATTT binding. The K_D for binding to an AAATTT sequence without the G·C bp, for example, increases from around 200nM with DB2457, the *N*-methyl derivative, to around 1000nM with the *i*-Pr compound DB2708 (**Table B.2**) to provide a significant increase in selectivity. Clearly, many of the *N*-X derivatives of this compound in **Table B.2** can match the minor groove microstructure in AAAGTTT better than with AAATTT. Simulation of additional *N*-X derivatives and more detailed analysis will be presented in a future paper.

Appendix B.6 Conclusions

Here we have synthesized a series of DNA sequence-specific recognition ligands based on the structure of DB2457 with *N*-MeBI-thiophenes which are preorganized by a σ-hole interaction to fit the shape of the DNA minor groove and H-bond to the –NH of G·C bps that project into the minor groove. In order to improve their potential biological applications, modifications have been made to the initial structure to increase the binding specificity. We increased the size of *N*-substitution of BI to match the DNA microstructure in target mixed base pair sequences and in successfully (**Table B.2**) improving the compound binding selectivity. We also changed the core structure twist of the compounds by introducing substitutions at specific positions on the aromatic core of the compounds. In both ways, we have been able to achieve a significant increase in binding specificity, including no detectable binding to pure AT sequences, without a significant loss of binding affinity to target mixed base pair sequences.

Various biophysical experiments have been performed to support our results: DNA thermal melting and biosensor-SPR were used to detect the binding affinity and kinetics between

compounds and target mixed base pair DNA sequences. The binding mode was determined by CD experiments along with the 1:1 minor groove binding. Structural calculations were performed to compare the different influence of modified groups at the same position of the compounds. Molecular dynamics was used to help explain the binding details and mechanism. Three H-bonds around the central G·C bp as BI(3)-C-H---O=dT(7), BI(3)-C-H---O=dC(15) and BI(4)-N---H-N-dG(6) help determine the strong binding affinity and high binding specificity of the representative *N*-*i*-Pr compound, DB2708, to an AAAGTTT sequence. These structural modifications provide ideas to produce significant improvements of the binding specificity of G·C bp sequences. The key compounds in this project should help improve the biological applications of small molecule DNA minor groove sequence-specific binders in projects such as the inhibition of transcription factors.

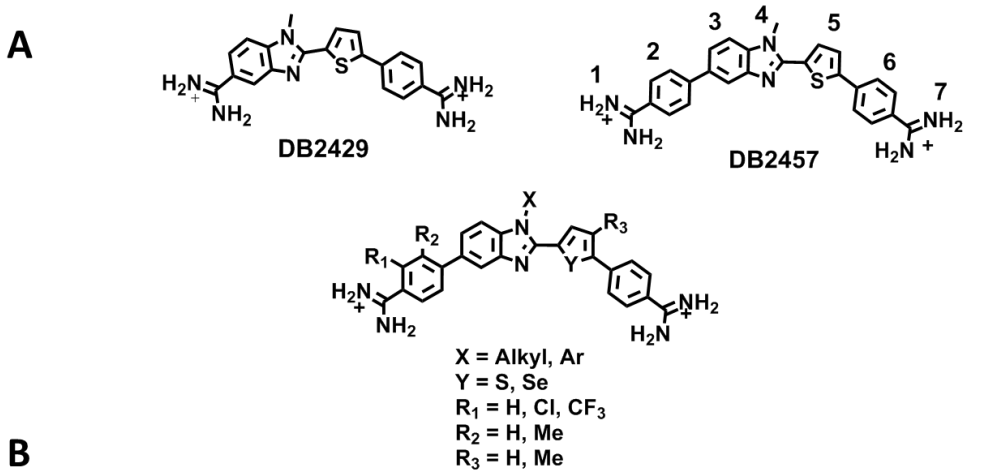


Figure B.1. A) Chemical structures of reported σ–hole single G·C bp binders and the modified compounds from DB2457 used in this study. B) The DNA sequences used in this study; DNA sequences used for SPR studies were labeled with 5'-biotin.

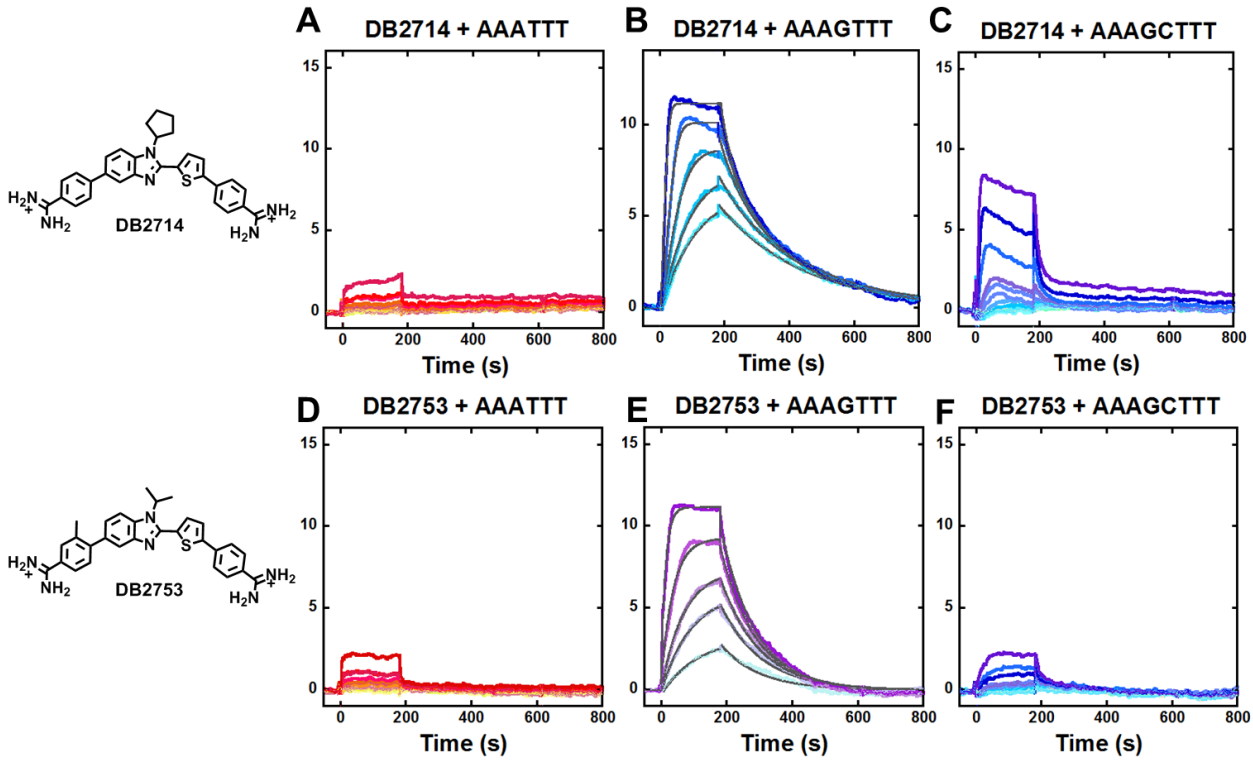


Figure B.2. Representative SPR sensorgrams for A-C, DB2714, D-F, DB2753 in the presence of AAATTT, AAAGTTT and AAAGCTTT hairpin DNAs. In A, C, D, F, the concentrations of sensorgrams is 2-500 nM of each compound from bottom to top. In B, the concentrations of DB2714 from bottom to top are 15, 20, 30, 50, and 100 nM; In E, the concentrations of DB2753 from bottom to top are 30, 70, 100, 200, and 500 nM. In B and E, the solid black lines are best-fit values for global kinetic fitting of the results with a single site function.

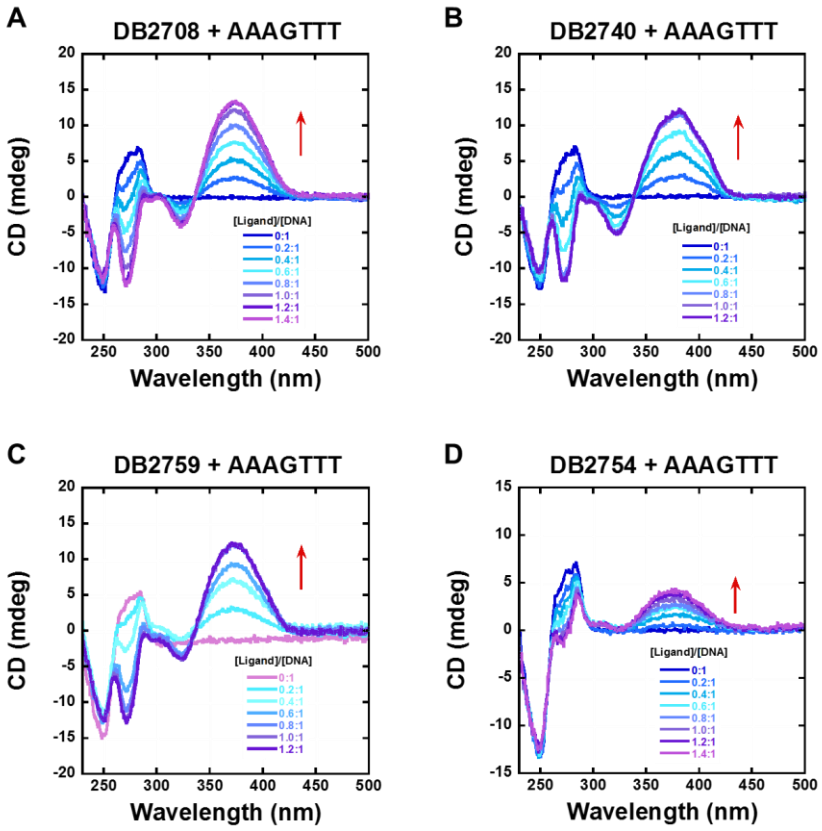


Figure B.3. Circular dichroism spectra for the titration of representative compounds, A) DB2708, B) DB2740, C) DB2759 and D) 2754 with a 5 µM AAAGTTT sequence in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) at 25℃. Arrows indicate the changes.

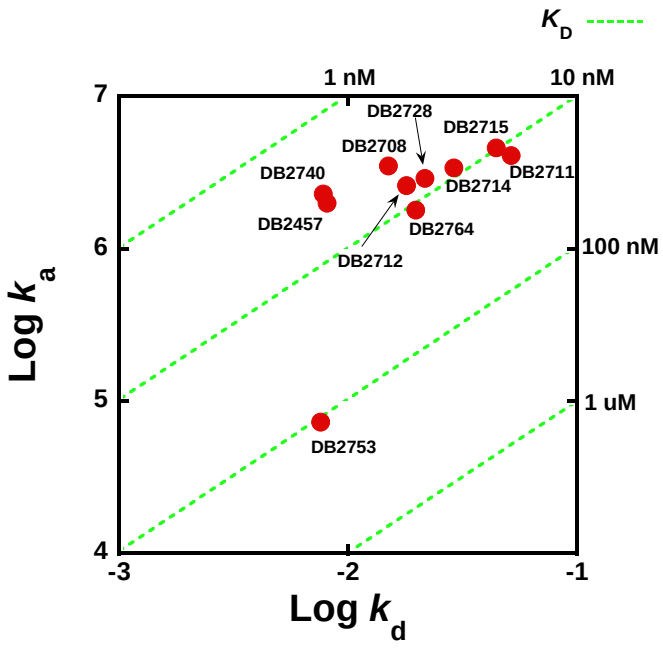


Figure B.4. Kinetic plot showing the relationship between on-rates (k_a), off-rates (k_d) and binding affinities (K_D) (Green dash line) represented on the diagonal axis for compounds in Table B.3, and determined by SPR. On-rates and off-rates and are plotted as logarithmic values.

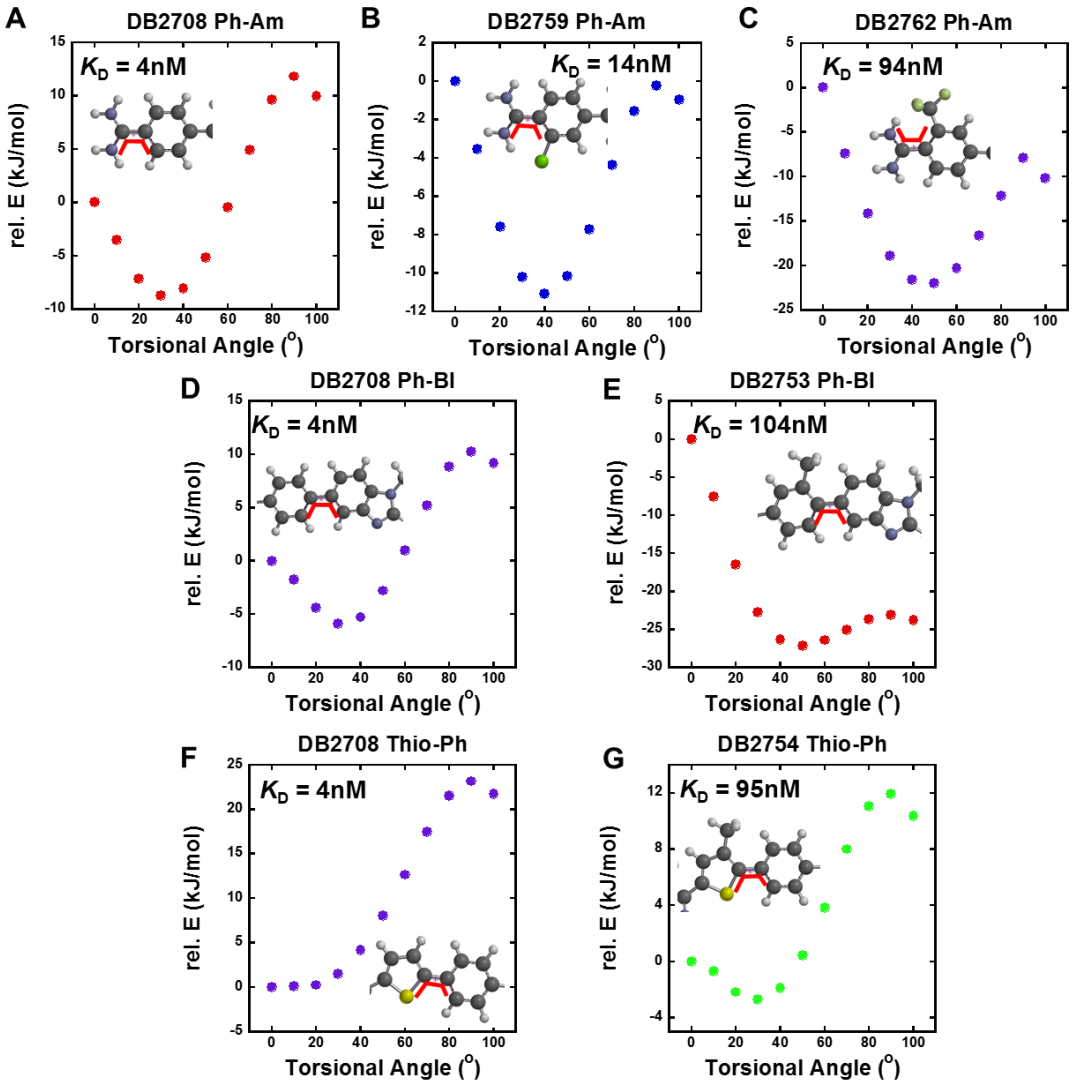


Figure B.5. (A, B, C) Torsional angle maps of the Ph-Am bond for DB2708, DB2759, and DB2762; and (D, E) of the Ph-BI bond for DB2708 and DB2753; and (F, G) of the Thiophene-Ph bond for DB2708 and DB2754. All calculations are performed at the B3LYP/6-31G* level of theory. The range of dihedral calculations is from 0°-100°. The scanned dihedral is shown as the bold red line at 0°in the compound structures.

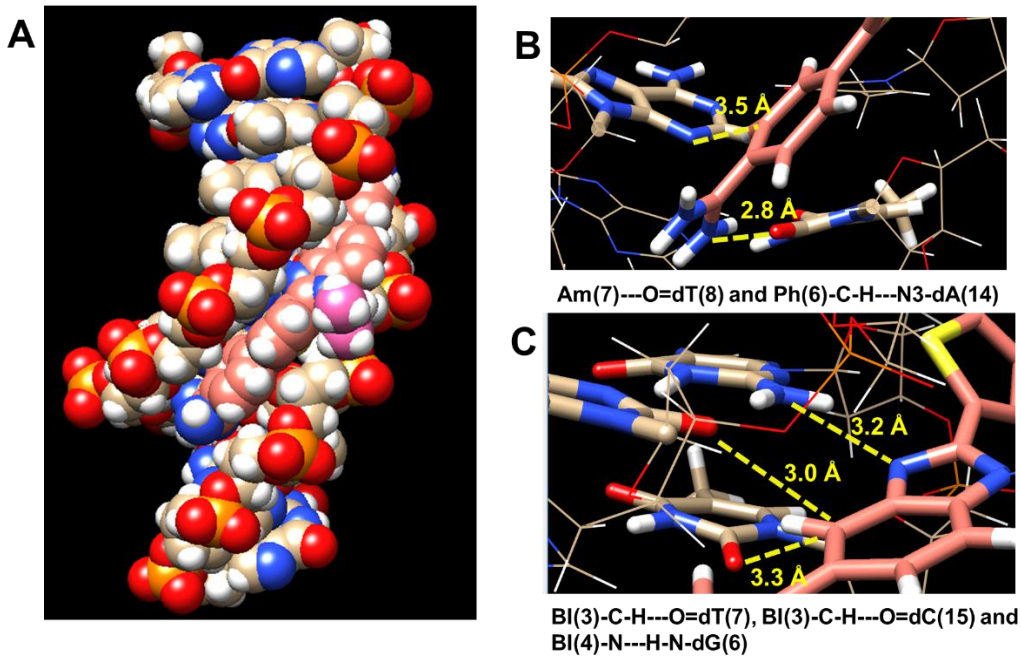
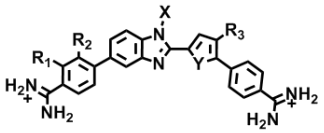
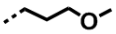


Figure B.6. Molecular Dynamics (MD) model of DB2708 with an AAAGTTT sequence: (A) A space-filling model viewed into the minor groove of the A3GT3 binding site with bound DB2708. The DNA bases are represented in tan-white-red-blue-yellow(C-H-O-N-P) color scheme and DB2708 is light pink-white-blue-yellow (C-H-N-S) color scheme with the N-i-Pr is facing out of the minor groove in a brighter-pink. The important interactions between different sections of DB2708-DNA are illustrated in (B) and (C). The terminal amidine group forms a strong hydrogen bond with the carbonyl group of dT(8) (yellow dashed lines), and another direct stabilizing interaction is observed from the Ph(6) –CH group which points to N3 of dA (14) at the floor of the minor groove (B). In (C), the imidazole (4)-N makes a strong hydrogen bond interaction with the exocyclic NH of dG(6). The BI(3)-C-H forms strong interactions with the carbonyl group of dT(7) in the minor groove and this unit also points towards the carbonyl group of dC(15) with close distance.

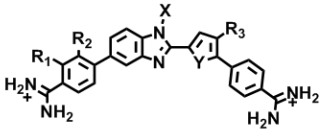
Table B.1. Thermal melting studies (ΔT_m, ℃) of DB2457 and analogues with pure AT and mixed DNA sequences. [a]

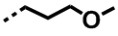


Compounds	R_1	R_2	R_3	X	Y	ΔT_m AAATTT (70℃)	ΔT_m AAAGTTT (66℃)	ΔT_m AAAGCTTT (67℃)
DB2457 (7a)	H	H	H	Me	S	6	14	5
DB2737 (7b)	H	H	H	Et	S	7	15	7
DB2708 (7c)	H	H	H	*i*-Pr	S	4	14	5
DB2711 (7d)	H	H	H		S	5	13	5
DB2718 (7e)	H	H	H		S	2	11	4
DB2715 (7f)	H	H	H		S	5	13	5
DB2728 (7g)	H	H	H		S	4	13	5
DB2764 (7h)	H	H	H		S	3	13	5
DB2726 (7i)	H	H	H		S	4	14	4
DB2714 (7j)	H	H	H		S	4	13	5
DB2727 (7k)	H	H	H		S	4	13	3
DB2738 (7l)	H	H	H		S	4	13	5
DB2740(7m)	H	H	H		S	5	17	7
DB2747 (7n)	H	H	H		S	7	16	9
DB2759 (7o)	Cl	H	H	*i*-Pr	S	3	12	3
DB2762 (7p)	CF_3	H	H	*i*-Pr	S	1	8	3
DB2753 (7q)	H	Me	H	*i*-Pr	S	1	3	2
DB2754 (7r)	H	H	Me	*i*-Pr	S	1	4	2
DB2673 (7s)	H	H	H	Me	Se	7	12	5
DB2712 (7t)	H	H	H	*i*-Pr	Se	5	13	6

[a] *ΔT_m = T_m (the complex) - T_m (the free DNA). 3 μM DNA sequences were studied in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) with the ratio of 2:1 [ligand]/[DNA]. An average of two independent experiments with a reproducibility of 0.5 ℃.*
Full DNA sequences: AAATTT: 5'-CCAAATTTGCCTCTGCAAATTTGG-3';
AAAGTTT: 5'-CCAAAGTTTGCTCTCAAACTTTGG-3';
AAAGCTTT: 5'-CCAAAGCTTTGCTCTCAAAGCTTTGG-3'.

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



Compounds	R_1	R_2	R_3	X	Y	K_D (nM) /Selectivity Ratio AAATTT	K_D (nM) AAAGTTT	K_D (nM) /Selectivity Ratio AAAGCTTT
DB2457 (7a)	H	H	H	Me	S	222 / 56	4	192 / 48
DB2737 (7b)	H	H	H	Et	S	750 / 250	3	145 / 48
DB2708 (7c)	H	H	H	*i*-Pr	S	1020 / 255	4	223 / 56
DB2711 (7d)	H	H	H		S	NB / --	13	NB / --
DB2718 (7e)	H	H	H		S	NB / --	31	690 / 22
DB2715 (7f)	H	H	H		S	NB / --	10	1080 / 108
DB2728 (7g)	H	H	H		S	NB / --	7	237 / 34
DB2764 (7h)	H	H	H		S	NB / --	11	347 / 32
DB2726 (7i)	H	H	H		S	NB / --	5	180 / 36
DB2714 (7j)	H	H	H		S	NB / --	9	1330 / 148
DB2727 (7k)	H	H	H		S	NB / --	14	388 / 28
DB2738 (7l)	H	H	H		S	NB / --	21	NB / --
DB2740(7m)	H	H	H		S	NB / --	3	107 / 36
DB2747 (7n)	H	H	H		S	NB / --	3	130 / 43
DB2759 (7o)	Cl	H	H	*i*-Pr	S	NB / --	14	NB / --
DB2762 (7p)	CF_3	H	H	*i*-Pr	S	NB / --	94	NB / --
DB2753 (7q)	H	Me	H	*i*-Pr	S	NB / --	104	NB / --
DB2754 (7r)	H	H	Me	*i*-Pr	S	NB / --	95	NB / --
DB2673 (7s)	H	H	H	Me	Se	650 / 77	9	300 / 33
DB2712 (7t)	H	H	H	*i*-Pr	Se	NB / --	7	290 / 41

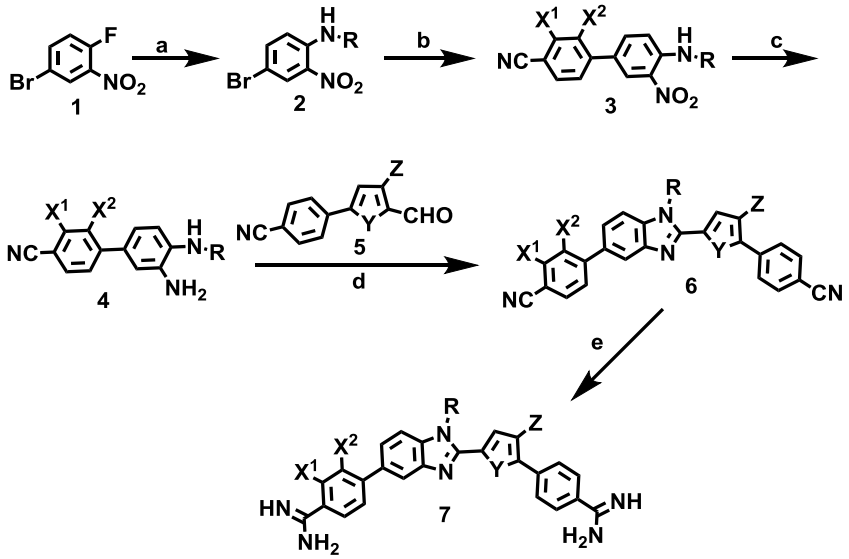
[a] *All the results in this table were investigated in Tris-HCl buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.05% P20, pH 7.4) at a 100 µL min^{-1} flow rate. NB means no binding. The listed binding affinities are an average of two independent experiments carried out with two different sensor chips and the values are reproducible within 10% experimental errors.*

Table B.3. Comparison of SPR Kinetics Results of Binding to the AAAGTTT DNA Sequence[a]

Compounds	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_D(M)
DB2457 (7a)	2.0 X 10^6	0.0081	4.1 X 10^{-9}
DB2708 (7c)	3.5 X 10^6	0.0150	4.3 X 10^{-9}
DB2711 (7d)	4.1 X 10^6	0.0516	12.6 X 10^{-9}
DB2715 (7f)	4.6 X 10^6	0.0444	9.6 X 10^{-9}
DB2728 (7g)	2.9 X 10^6	0.0217	7.4 X 10^{-9}
DB2764 (7h)	1.8 X 10^6	0.0198	11.0 X 10^{-9}
DB2714 (7j)	3.4 X 10^6	0.0290	8.5 X 10^{-9}
DB2740 (7m)	2.3 X 10^6	0.0078	3.4 X 10^{-9}
DB2753 (7q)	7.3 X 10^4	0.0076	104 X 10^{-9}
DB2712 (7t)	2.6 X 10^6	0.0180	7.0 X 10^{-9}

[a] Kinetics analysis was performed by global fitting with a 1:1 binding model. k_a is the association rate constant, while k_d is the dissociation rate constant. K_D was calculated through global fitting of the kinetic data obtained for various concentrations of the compounds; K_D is given by k_d / k_a.

Scheme B.1. The general synthetic approach to the target diamidine compounds



Reagents and conditions: a) RNH_2/EtOH or Cs_2CO_3/DMA; b) 4-cyanophenylboronic acid, $Pd(PPh_3)_4$, Na_2CO_3/H_2O, dioxane, reflux; c) $SnCl_2$/EtOH, reflux; d) $Na_2S_2O_5$/ DMSO, 130 ^oC; e) i- $LiN(TMS)_2$/THF, ii- HCl/EtOH.

Appendix B.7 References

(1) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183* (4), 553.

(2) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25* (23), 7408.

(3) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Mol. Biol.* **1998**, *5* (2), 104.

(4) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3* (6), 688.

(5) Bando, T.; Sugiyama, H. *Acc. Chem. Res.* **2006**, *39* (12), 935.

(6) Wang, S.; Nanjunda, R.; Aston, K.; Bashkin, J. K.; Wilson, W. D. *Biochemistry* **2012**, *51* (49), 9796.

(7) Nozeret, K.; Loll, F.; Cardoso, G. M.; Escudé, C.; Boutorine, A. S. *Biochimie* **2018**, *149*, 122.

(8) Hargrove, A. E.; Raskatov, J. A.; Meier, J. L.; Montgomery, D. C.; Dervan, P. B. *J. Med. Chem.* **2012**, *55* (11), 5425.

(9) Nishijima, S.; Shinohara, K.-I.; Bando, T.; Minoshima, M.; Kashiwazaki, G.; Sugiyama, H. *Bioorg. Med. Chem.* **2010**, *18* (2), 978.

(10) Chazotte, B. *Cold Spring Harb. Protoc.* **2011**, *2011* (1), pdb.prot5557.

(11) Giordani, F.; Munde, M.; Wilson, W. D.; Ismail, M. A.; Kumar, A.; Boykin, D. W.; Barrett, M. P. *Antimicrob. Agents Chemother.* **2014**, *58* (3), 1793.

(12) Wilson, W. D.; Nguyen, B.; Tanious, F. A.; Mathis, A.; Hall, J. E.; Stephens, C. E.; Boykin, D. W. *Curr. Med. Chem. Anticancer Agents* **2005**, *5* (4), 389.

(13) Paine, M. F.; Wang, M. Z.; Generaux, C. N.; Boykin, D. W.; Wilson, W. D.; De Koning, H. P.; Olson, C. A.; Pohlig, G.; Burri, C.; Brun, R.; Murilla, G. A.; Thuita, J. K.; Barrett, M. P.; Tidwell, R. R. *Curr. Opin. Investig. Drugs* **2010**, *11* (8), 876.

(14) Soeiro, M. N. C.; Werbovetz, K.; Boykin, D. W.; Wilson, W. D.; Wang, M. Z.; Hemphill, A. *Parasitology* **2013**, *140* (8), 929.

(15) Ming, X.; Ju, W.; Wu, H.; Tidwell, R. R.; Hall, J. E.; Thakker, D. R. *Drug Metab. Dispos.* **2009**, *37* (2), 424.

(16) Chai, Y.; Paul, A.; Rettig, M.; Wilson, W. D.; Boykin, D. W. *J. Org. Chem.* **2014**, *79* (3), 852.

(17) Paul, A.; Nanjunda, R.; Kumar, A.; Laughlin, S.; Nhili, R.; Depauw, S.; Deuser, S. S.; Chai, Y.; Chaudhary, A. S.; David-Cordonnier, M.-H.; Boykin, D. W.; Wilson, W. D. *Bioorg. Med. Chem. Lett.* **2015**, *25* (21), 4927.

(18) Guo, P.; Paul, A.; Kumar, A.; Farahat, A. A.; Kumar, D.; Wang, S.; Boykin, D. W.; Wilson, W. D. *Chem. Eur. J.* **2016**, *22* (43), 15404.

(19) Munde, M.; Wang, S.; Kumar, A.; Stephens, C. E.; Farahat, A. A.; Boykin, D. W.; Wilson, W. D.; Poon, G. M. K. *Nucleic Acids Res.* **2014**, *42* (2), 1379.

(20) Antony-Debré, I.; Paul, A.; Leite, J.; Mitchell, K.; Kim, H. M.; Carvajal, L. A.; Todorova, T. I.; Huang, K.; Kumar, A.; Farahat, A. A.; Bartholdy, B.; Narayanagari, S.-R.; Chen, J.; Ambesi-Impiombato, A.; Ferrando, A. A.; Mantzaris, I.; Gavathiotis, E.; Verma, A.; Will, B.; Boykin, D. W.; Wilson, W. D.; Poon, G. M. K.; Steidl, U. *J. Clin. Invest.* **2017**, *127* (12), 4297.

(21) Jiang, X.; Yang, Z. *OncoTargets Ther.* **2018**, *11*, 3533.

(22) Lambert, M.; Jambon, S.; Depauw, S.; David-Cordonnier, M.-H. *Molecules* **2018**, *23*

(6), 1479.
(23) Darnell, J. E. *Nat. Rev. Cancer* **2002**, *2* (10), 740.
(24) Koehler, A. N. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 331.
(25) Harika, N. K.; Germann, M. W.; Wilson, W. D. *Chem. Eur. J.* **2017**, *23* (69), 17612.
(26) Mallena, S.; Lee, M. P. H.; Bailly, C.; Neidle, S.; Kumar, A.; Boykin, D. W.; Wilson, W. D. *J. Am. Chem. Soc.* **2004**, *126* (42), 13659.
(27) Neidle, S. *Nat. Prod. Rep.* **2001**, *18* (3), 291.
(28) Steinmann, D.; Nauser, T.; Koppenol, W. H. *J. Org. Chem.* **2010**, *75* (19), 6696.
(29) Reich, H. J.; Hondal, R. J. *ACS Chem. Biol.* **2016**, *11* (4), 821.
(30) Liu, Y.; Kumar, A.; Depauw, S.; Nhili, R.; David-Cordonnier, M.-H.; Lee, M. P.; Ismail, M. A.; Farahat, A. A.; Say, M.; Chackal-Catoen, S.; Batista-Parra, A.; Neidle, S.; Boykin, D. W.; Wilson, W. D. *J. Am. Chem. Soc.* **2011**, *133* (26), 10171.
(31) Munde, M.; Kumar, A.; Nhili, R.; Depauw, S.; David-Cordonnier, M.-H.; Ismail, M. A.; Stephens, C. E.; Farahat, A. A.; Batista-Parra, A.; Boykin, D. W.; Wilson, W. D. *J. Mol. Biol.* **2010**, *402* (5), 847.
(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; GE Scuseria… - Inc.; Wallingford; CT; 2010. *Gaussian 09, revision b. 01, Gaussian*.
(33) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97* (40), 10269.
(34) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5* (2), 129.
(35) Case, D. A.; Babin, V.; Berryman, J.; Betz, R. M.; Cai, Q. *Amber 14*, NYU Press, **2014**
(36) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25* (2), 247.
(37) Athri, P.; Wilson, W. D. *J. Am. Chem. Soc.* **2009**, *131* (22), 7618.
(38) Špačková, N.; Cheatham, T. E.; Ryjáček, F.; Lankaš, F.; van Meervelt, L.; Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2003**, *125* (7), 1759.
(39) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2009**, *65*, 455.
(40) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. *Nature* **2009**, *461* (7268), 1248.
(41) Laughlin-Toth, S.; Carter, E. K.; Ivanov, I.; Wilson, W. D. *Nucleic Acids Res.* **2017**, *45* (3), 1297.
(42) Bishop, E. P.; Rohs, R.; Parker, S. C. J.; West, S. M.; Liu, P.; Mann, R. S.; Honig, B.; Tullius, T. D. *ACS Chem. Biol.* **2011**, *6* (12), 1314.
(43) Fornander, L. H.; Wu, L.; Billeter, M.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2013**, *117* (19), 5820.
(44) Hehre, W.; Ohlinger, S. *Spartan"16* Tutorial and User"S Guide. Wavefunction, Inc., **2016**
(45) Harika, N. K.; Wilson, W. D. *Biochemistry* **2018,** doi: 10.1021/acs.biochem.8b00647.