12-16-2020

Investigating the Role of Hydration and DNA Dynamic Alterations in DNA Recognition by A Heterocyclic Diamidine

Van Ha
Georgia State University, tuongvanhale@gmail.com

INVESTIGATING THE ROLE OF HYDRATION AND DNA DYNAMIC ALTERATIONS IN DNA RECOGNITION BY A HETEROCYCLIC DIAMIDINE

by

VAN HA

Under the Direction of Gregory M. K. Poon, PhD

ABSTRACT

Target recognition by DNA-binding ligands, such as drugs, occurs in an aqueous environment, in which water (near unit mole fraction, ~55 M) dominates every solute. A quantitative account of how water molecules are disposed in DNA/ligand binding is indispensable for understanding the driving forces that confer high-affinity and selectivity. We are investigating the DNA sequence selectivity of a model DNA minor groove-binding heterocyclic diamidine, DB1976, which shows therapeutic activity in acute myeloid leukemia, systemic fibroses, and obesity-related liver disorders *in vivo*. The DNA minor groove is richly populated with water molecules. Studies based on explicit-solvent MD simulation have shown distinct DNA dynamics upon drug-DNA complexes. We have cooperated the role of hydration and conformational dynamics in contributing to drug selectivity. Moving forward, our goal is to evaluate the structure-hydration relationships of designed diamidines to site-specific and nonspecific DNA as part of their biophysical characterization as potential therapeutic agents.

INDEX WORDS: DNA dynamics, hydration, target selection, DB1976, volumetric analysis.

INVESTIGATING THE ROLE OF HYDRATION AND DNA DYNAMIC ALTERATIONS IN DNA RECOGNITION BY A HETEROCYCLIC DIAMIDINE

by

VAN HA

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

Master of Science

in the College of Arts and Sciences

Georgia State University

2020

THE ROLE OF HYDRATION AND DNA DYNAMIC ALTERATIONS IN DNA RECOGNITION BY A HETEROCYCLIC DIAMIDINE

by

VAN HA

Committee Chair: Gregory M. K Poon

Committee: W. David Wilson

Jenny J. Yang

Donald Hamelberg

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

December 2020

DEDICATION

This work is dedicated to my parents and sisters.

ACKNOWLEDGEMENTS

I am thankful for everyone who I have met since I started college at Georgia State University. I am grateful for all the moments I have been through the past few years: I have smiled, laughed, frowned, cried, gave up, and stood back up again. Those lessons that I have learned make me who I am today – I stand up for what I love, follow my passion, and I am responsible for my decisions.

Since the day I got accepted to be a lab member of Dr. Gregory Poon's lab, I have learned many things from scientific perspectives to life aspects. I want to say thank you to my advisor Dr. Gregory Poon for giving me opportunities to discover and pursue my interests in science. The support and trust that he has put into me helped me build up experiences and skills for my future career. I will always respect him for everything he has done for me: all the time, advice, and encouragement he's willing to give out to support his student's decisions.

I want to thank all the committee members: Dr. W. David Wilson, Dr. Jenny J. Yang, and Dr. Donald Hamelberg. In a collaborated meeting with Dr. Wilson's group, I have always learned new knowledge from the presentations and all advice, questions, and answers that Dr.Wilson has given us. In Dr. Yang's class, I have always been inspired by her enthusiasm when teaching about the unique techniques and applications of science. I am thankful for Dr. Hamelberg, who gave me a chance to present my thesis to him. Every member has supported and helped me believe in myself again.

I want to thank all the current and past members of Dr. Poon's lab. They have been great lab-mates and friends who always support me.

I want to say thank you to my parents and sisters, who will always be there for me. My mother and father have sacrificed many things to let us live a comfortable life in the United States.

They love me unconditionally. Thank you, Mom, for always waiting for me to get back home. Thank you, Dad, for giving out the best lessons to me. Thank you, my sisters, who always sounded "a little bit" nice when they talked to me, but I know that they always love me and support me through all the arguments, fights, random talks we have had, and all the toys and foods that we have played, learned and ate together. Family is a gift that lasts forever.

Thank you to my loved ones and best friends for always listening to me, supporting me, and encouraging me. Thank you for always protecting me, listening to all my stories, jokes, and complaints. I always end up with happiness when I am around you guys.

Thank you again, everyone, for being a part of my life. From fate, we are family; from strangers, we become friends; from density, we fall in love – together, we grow and succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Acute myeloid leukemia (AML)

Diet-induced obese (DIO)

Non-alcoholic steatohepatitis (NASH)

Transcription factor (TF)

High mobility group box (HMG-box)

Molecular dynamic (MD)

Center of mass (COM)

Median absolute deviation (MAD)

Posterior predictive checks (PPC)

Markov Chain Monte Carlo (MCMC)

Highest posterior density (HPD)

Solvent-accessible surface area (SASA)

Free (F)

Bound (B)

Protein Data Bank (PDB)

Kirkwood-Buff (KB)

Calf thymus DNA (CtDNA)

Triethylene glycol (TEG)

1 INTRODUCTION

1.1 The history of small molecules and current effort in therapeutic application

1.1.1 Brief introduction about small molecules applications

A small molecule drug is an organic compound with a low molecular weight that possesses specific therapeutic activity. Common small molecule drugs include aspirin tablets, and antibiotics such as penicillin (bacterial infection) and atorvastatin (high cholesterol treatment). The development of new small molecule in therapeutics or biological research application is an auspicious field as those molecules cover a wide range of targets from DNA binding, protein binding, or RNA interaction (1, 2). Such as the small molecule – Diminazene, defined as an inhibitor of protein mesotrypsin - a protein associated with various tumor progression (2).

1.1.2 The small molecules-DNA minor groove binding mode

There are multiple binding modes that are classified in DNA-small molecules binding modes, such as intercalators, groove binders, and single-strand binders (the least common type) (3-5). Intercalation behavior was first observed when Lerman studied the interaction between DNA and acridine in 1961. The DNA helix was unwound due to the insertion of planar aromatic substituents into DNA base pairs (4). However, the minor groove binders, proposed by Wartell's study on netropsin, use the concept of fitting to bind to DNA without any damage or minimal perturbation to the natural structure of duplex (5).

The non-covalent sequence-specific interactions with small molecules in the minor groove of B-DNA presents a high level of interest for drug design, especially potential anti-cancer drugs (6). The Hoechst dye families have been well known as blue-fluorescent stains widely used for DNA staining. The Hoechst family belongs to a class of small molecules that preferably bind to

AT-rich regions (7). The Figure 1.1 illustrates the crystal structure of Hoechst 33258 binds to the minor groove of B-DNA and doxorubicin, an intercalation binder.

1.1.3 The classical model of DNA minor groove binders – aromatic amidine derivatives

The studies of small-molecule minor group binders, that belonging to the class of aromatic amidines, have shared similar features of A-T rich binding. For example, the crescent shape that fits into the helical minor groove which is similar to the induce-fit mechanism in enzyme binding. Many studies have found that small molecules bind to A-T rich region of DNA minor groove via non-covalent interaction due to the narrow groove width of in A-T region, allowing for optimization of H-bonding and van der Waals' contacts. Besides complementary shapes, the positive charged ends also allow for electrostatic interaction. A variety of structures also have flexible torsion or called "flexible arms" which help maximize the surface contact between the compound and the DNA to form H-bond (8).

However, in some instances, compounds that have a linear shape rather than the classical curvature shape, violate the classical fitting model. The pentamidine, a synthetic derivative of amidine, has been used clinically as an anti-infective agent against human protozoan infections, treatment for Pneumocystis carinii pneumonia, and African trypanosomiasis (9-11). A crystal structure of pentamidine and dodecane nucleotide d(CGCGAATTCGCG)2 complex establishes the binding to AT-rich regions at the minor groove of the duplex DNA (12). The pentamidine possesses a linear- central shape structure, compared to the shape of Hoechst 33258 or netropsin, but are still be able to access the minor groove curvature orientation, indicating a flexible structure upon complex formation (Figure 1.2). A study on DB921 – a structurally linear diamidine and DB DB911 – a curvature model shape diamidine, has shown distinct characteristics. Based on the

biosensors-SPR experiment, DB921 binds to AATT region stronger compared to DB911 in which the binding constant (K_A) are approximately 2×10^8 M−1 and 2×10^7 M−1, respectively. This result indicates that compounds with nonstandard shapes are still able to bind to the minor groove. The flexibility of the structure such as change in torsion angles may help maximize the binding effect. However, the extended allowance for possible binding is also limited. The compound DB2232, structurally presented in Figure 1.2, is an example of extending linkage diamidine that binds weakly to the minor groove AATT region. One explanation for the weaker selectivity is the limited conformational flexibility in its structure due to the narrow groove width of AATT region (8).

1.2 Why do we choose DB1976 as a role model?

1.2.1 DB1976 is an active inhibitor of transcription factor PU.1 which associates with acute myeloid leukemia (AML)

The DB1976 is a three-steps replacement derivative of DB75 (furamidine), which actively works against antiparasitic activity such as African trypanosomes *in vitro*, and Plasmodium falciparum (13, 14). Despite the therapeutic property of DB75, the derivatives from a clinically tested compound may provide a new discovery in disease treatments.

The ETS-family transcription factor member called PU.1 plays an important role in hematopoiesis. The PU.1 is conserved between humans and mice. The PU.1 is a direct regulator of myeloid and lymphoid differentiation (15, 16). The study has found that deregulation of PU.1 expression leads to multiple hematopoietic abnormalities and associates with acute myeloid leukemia (AML) in mouse models and patient samples (16-18).

A research conducted on the inhibition of low-level PU.1 expression has established a strategy for the potential treatment of AML by novel heterocyclic diamidines, which are minor

groove binders that are structurally derived from clinical compounds. Using a modified mouse model with approximately 20% of normal PU.1 expression level, the treatment of heterocyclic diamidines decrease the growth of PU.1 reduced cells with minimal effect on normal hematopoietic cells. Out of the three compounds selected in the study, namely DB2313, DB2115, and DB1976, the DB1976 has established a weaker PU.1 inhibition activity (19).

1.2.2 DB1976 is an active agent as a PU.1 inhibitor in obesity-related liver disorders

Obesity is a primary health concern that is associated with many risk factors such as the development of insulin resistance, which can also lead to non-alcoholic steatohepatitis (NASH). A recent study on diet-induced obese (DIO) mice model has identified an upregulated level of PU.1/SPI1 transcription factor (TF) in the liver up to approximately 5-fold higher compared to chow diet-fed mice. The knock-down study of PU.1 level in the liver has proven to show an improvement in glucose homeostasis such as improving glucose tolerance, lowering the fasting glucose level. The inhibition of PU.1 emerges as a novel therapeutic strategy for treatment of liver dysfunction, dysregulation of glucose homeostasis caused by obesity. Treatment with DB1976 as an active drug for PU.1 inhibitor has shown remarkably effect on normalizing glucose tolerance in DIO mice. Overall, treatment of DB1976 has a positive effect on hyperglycemia, hepatic inflammation and glucose intolerance by the action of inhibition PU.1 rather than hepatotoxicity (20).

1.2.3 DB1976 shows potential treatment effect on fibrotic fibroblasts, which express high PU.1level.

The fibroblast plays an important role in tissue integrity. It synthesizes extracellular matrix and collagen for tissues and is especially responsible for the wound healing process. In inflammatory fibroblasts, the fibroblasts develop a degradation of extracellular matrix phenotype,

while in fibrotic diseases, the progressive accumulation of extracellular matrix is observed. A recent study has found that the transcription factor PU.1 is highly expressed in fibrotic fibroblasts but is silent in extracellular matrix-degrading inflammatory fibroblasts. They also have indicated that inhibition or inactivation of PU.1 activity disrupts the progressive accumulation of extracellular matrix-producing network, which helps reprogram fibrotic fibroblasts to a resting state. Under the treatment of DB1976, an anti-fibrotic effect in vivo is observed across several organs. Mice models have minimal side effects such as change in body weight, stress, and pain. DB1976 not only inhibits the pro-fibrotic gene-mediated skin fibrosis, but induces regression of pre-establish fibrosis (21).

1.2.4 DB1976 covers a broader range of binding affinity

The ETS family proteins bind to DNA sequences that contain a 5′-GGAA-3′ core consensus sequence. The PU.1 carries a more specific binding specificity which aims for a sequence containing AT-rich tracks flanking both sides of the ETS consensus (22). DB1976 has actively inhibited PU.1 both in biophysical study and in vivo with minimal effects or no binding interaction with PU.1. The SPR data on the binding affinity of heterocyclic diamidines on the λB site, a high-affinity cognate sequence for PU.1, shows a structural dependence variation in binding affinity. The modification steps from the parent compound, DB75, establishes a greater binding affinity going down to DB1976 modification. The DB1976-λB affinity (K_D = 12 nM) is remarkable compared to the parent compound, DB75 which K_D = 0.53 μM (23, 24). To develop a structural relationship for the development of small-molecule inhibitors of PU.1, DB 1976 is a dominant model due to its symmetrical structure, high A-T sequence binding affinity, and an active PU.1 inhibitor.

1.3 The role of hydration in biological studies

1.3.1 What is hydration layer?

The hydration layer is the term used to describe the water cluster layer due to the interaction between substance with water molecules in a solution. The hydration of ions has been well defined; for example, in Na^+ and Cl^- cases. The positive Na^+ ion is attracted to the slightly negative charge of oxygen in the water molecule and the negative Cl^- ion is attracted to the small positive charge of the hydrogen in water. The slight charge on water molecules are created from dipole moments. The ions drive the formation of a condensed water network called hydration layer, in which the water molecular dynamic is different from the bulk properties. While the water from the bulk moves randomly, the water in hydration plays an important role in the solute's activities (25).

1.3.2 Examples of hydration roles in biological applications

The biological system such as the Na^+ and K^+ ion channel is responsible for electrical conduction in the nervous system. The system can distinguish between its specific and impermissible ions. Despite the slight difference in dimension of the ions, the ionic channel's selectivity depends on many factors, such as the hydration shell of the ions, to allow for polar interaction with the selectivity filter in the channel. The hydration shell radii are approximately 0.33 nm and 0.39 nm for Na^+ and K^+, respectively (26). The hydration shell of Na^+ is more stable than K^+ ion which is commonly make a transient associate with water rather than a compact hydration layer as Na^+. The selectivity of each channel follows the hydration characteristic such that highly hydrated ions Li^+ and $HONH3^+$ can pass thought Na+ channel, while Rb^+ and Tl^+ cations can pass through K^+ channel (26).

The hydration is not only limited to single ionic molecules in the system but also contributes to larger extents. The hydration plays a vital role in proteins, nucleic acids,

polysaccharides, phospholipid bilayer and functional organelles. A study on the relationship between hydration and enzyme activity of lysozyme has shown its digestive enzyme activity depends on the amount of water content per protein called the “degree of hydration”. There is a critical level of hydration of h = ~ 0.2 g required for lysozyme to surpass and become functional. In dehydrated states, enzymatic activity is not observed (27).

The role of water is also established in the protein-ligand binding profile. The serine proteases - trypsin contains a specific hydration pattern that mediates contact between the ligands and the protein. In the comparison of the change in the hydration layer in the binding pocket, amidino ligands bind to trypsin with incomplete displacement of water molecules inside the pocket. The small uncharged molecule such as m-chlorobenzyl binds to the subpocket (S1) with complete dehydration of all water molecules. Even though the serine proteases recognize substrates comprising basic residues prior to the peptide bond cleavage, the complete displacement of water from the binding pocket in serine protease enhances a substantial enthalpy-favored binding signature and improves the potency of neutral small molecules serine inhibitors (28, 29).

1.3.3 Distinct hydration properties of DNA minor groove binders through the thermodynamic binding profile

Upon the complex formation, the thermodynamic parameters that include enthalpy, entropy, and Gibbs free energy, provide significant information for the effort to optimize of binding affinity. The value of Gibbs free energy (ΔG) is strictly dependent on the change in enthalpy (ΔH) and entropy (ΔS) following the defined equation:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

Multiple conditions can be established, such as ΔG is minimized (ΔG ~ 0) while enthalpy and entropy compensation occur; or ΔG depends either on ΔH or ΔS or in couples of ΔH and ΔS.

For example, thermodynamic characteristic of melting ice is an endothermic process possess a large positive change in entropy ($\Delta S^{\ominus} \sim +22.1$ J. K^{-1}.mol^{-1}) and small change in enthalpy ($\Delta H^{\ominus} \sim +6.03$ kJ. mol^{-1}) to compensate for the energy supplies to break the hydrogen bond in the crystalline structure. Breaking of H-bond or van der Waals bonding between solute-solute and solute-solvent has a great contribution to the enthalpy and entropy of binding. Especially between the solute and hydration layer as the desorption of tightly bound water can affect the flexibility and mobility of the structural system and lead to a change in entropy. The possibility of major solvent change thermodynamically is considered and observed in some cases as it also contributes to the overall change entropy of the system. For example, a review article has pointed out that upon the binding of DNA to the high mobility group box (HMG-box), HMGD-100 and HMGD-74 - a binding domain found in many transcription factors - has a change in entropy of approximately 80 kJ/mol. One explanation for the increase in entropy of the system is the major change in structural conformation. However, based on the crystal structure and NMR study, the complex formation is completed with minimal distortion. The role of hydration is the major explanation as the change in water compact structure (releasing of water in hydration layer) leads to an increase in entropy, assumed the contribution of enthalpy and entropy between the bond formation between the DNA and the domain is ignorable (30).

Does hydration water always get released as complexes form? Each binding mode in the molecular system possesses its own thermodynamic profile. For two common types of small molecules binding, the intercalation and minor groove binders, the two dominant thermodynamically profiles established. A review by Chaires on drug-DNA binding mode has

pointed out that binding of groove binders dominantly entropical driven while intercalation binding favors enthalpic contributions (Figure 1.3). As minor groove binders bind to DNA with minimal DNA disruption, the well-establish explanation for its dominant entropic contribution is the displacement of hydrated water in the spine of hydration water (31). The presence of the "spine of hydration" in the minor groove has been defined by a variety of methods from X-ray crystallography (32-35), NMR spectroscopy (36, 37), other biophysical methods, and molecular dynamic simulation (38). While intercalation binders dominantly gain net uptake of water in complex formation (39).

However, depending on the ligand structural heterogeneity, the thermodynamics of minor groove binders may have completely different profiles (40). For the bisbenzimide derivatives or the Hoechst family, there is presence of an entropy-enthalpy compensation. The favored entropy profile ($-T\Delta S$ = -21.8 kcal mol^{-1}) is coupling with a penalty in enthalpy of +10.0 (kcal mol^{-1}) in Hoechst 33258. The diphenylfuran derivatives of DBs compounds such as DB 226, DB75, and DB 293 are dominantly entropy driven with minimal distribution or penalty in enthalpy. The entropic contribution to binding of DBs ranges from -8.0 to -6.8 kcal mol^{-1}, which are nearly double to enthalpy values of -3.6 to -0.5 kcal mol^{-1}. The pyrrole-amidine linked group (propamidine, distamycin, and netropsin) establishes an equal distribution of both enthalpy and entropy. For example, distamycin has -5.8 kcal mol^{-1} in ΔH and -4.7 kcal mol^{-1} in $-T\Delta S$ (Figure 1.3).

1.3.4 Water mediates contact in DNA minor groove target

The X-ray structure of DB921 in complex formation with AT -rich minor groove DNA duplex shows DB921 binds to the AATT site with two distinct mediations at the two ends. The benzimidazole end of DB921 binds to the groove by a complementary curvature fit, while there is

a water-mediated contact between the phenylamidine and the groove as an interchange bridge (Figure 1.4). The binding of bases in the groove-water and water-DB's end provided an energetically stable complex. Interestingly, this water also associates with the whole water network surrounding the minor groove (8, 41, 42). The complexes stabilization in the assistance of water of hydration is also absorbed in classical curve shape compounds with minor groove (42). As hydration water indicates an important role in maximizing the binding affinity, a study on hydration properties of minor groove binders are necessary to obtain the full profiles of compounds thermodynamically and structurally.

1.3.5 Probing the hydration change based on volumetric properties

Density is defined as a fraction of mass to volume. Density is a physical property as matters have their own unique density-profile. Measurement of density of water over a range of temperature has established the varies of density–dependence temperature. The kinetic energy of the molecular particle is an explanation for this behavior. Besides temperature, density can also be affected by dissolved material, such as in a mixed system. Interpreting the system by the volume factor of a mixture, such as methanol and water; the molar volume of the system is not a sum of the molar volume of pure water and methanol alone. The partial molar volume of solutes in particular solutions is predictably different as the molecular size and intermolecular interaction need to be taken into consideration. The partial molar volume measurement provides an insight into the thermodynamic properties of the system in which the partial molar volume provides the information between solute-solvent and solute-solute interactions, interpreted by its dependence on the concentration of dissolved material. The partial molar volume ($V°$) is defined following this equation, which it describes as the apparent volume occupied by one mole of solute at an infinite solution (43):

$$V^{\circ} = \frac{M}{\rho_0} - \frac{\rho - \rho_0}{\rho_0 C} \quad (2)$$

Where ρ_0 is the density of the buffer, C is the molar solute concentration, and M is the molecular weight of the solute, ρ is the density of the solution contained solute.

The partial molar volume and the volume change in association of with change in the solutes such as ligand binding or protein transition can be measure using vibrating tube densimetry, picnometry, dilatometry, etc. Other volumetric measurement such as partial molar adiabatic compressibility ($K^{o}{}_{S}$), the pressure derivative of the partial molar volume at constant entropy, are also commonly used. The $K^{o}{}_{S}$ of the solute can be obtained from differential solution of density and sound velocity measurement, defined as:

$$K^{\circ}_{S} = \beta_{S0}\left(2V^{\circ} - 2[U] - \frac{M}{\rho_0}\right) \quad (3)$$

Where β_{S0} is the coefficient of adiabatic compressibility of the solvent and $[U]$ is the molar sound velocity measurement obtained by:

$$[U] = \frac{(U - U_0)}{U_0 C} \quad (4)$$

The U and U_0 are sound velocities of the solution and the buffer (44).

In the review study by Chalikian, the idea of using contributions of individual components (such as amino acid residues or alcohol, sugar, carboxylic acid) to probe the hydration properties of the whole system is reliable using partial molar volume. Based on the chemical structure of functional groups, the partial molar volume of the whole solute can be estimated. A study on using the partial molar volume of amino acid sidechains to calculate partial molar volume of proteins is

successfully carried out which the results are comparable to experimental data. They have used the model of tripeptide gly[X]gly, where X is a variety of choice for amino acids and gly is glycine, to extract the partial molar volume of each amino acid side chain based on the idea that:

$$V^\circ(\mathrm{R}) = V^\circ(\text{gly-X-gly}) - V^\circ(\text{gly-gly-gly}) \quad (5)$$

Where $V^\circ(\mathrm{R})$ is the partial molar volume of side chain amino acid (R) or X in the tripeptide.

In the application of group additivity for the overall partial molar volume calculation of unfolding proteins, they have found that this idea is an effective method. For example, the experimental partial specific value of the native state of ribonuclease A and their predicted values are 0.704 and 0.705 $cm^3\,g^{-1}$, respectively. The partial molar volume of the back bond glycyl group and the ionic end groups of the polypeptide chain are also taken into consideration (45).

However, the application adiabatic compressibility on the same additivity approach results in a larger error. The possible explanation is that the system is more sensitive to subtle intramolecular interactions that can be silent to partial molar volume (46). The adiabatic compressibility of individual solutes may not completely encounter the intramolecular interactions found in the whole solutes. These observations point out the great application of additivity of partial molar volume in the examination of the whole system. However, the partial molar compressibility is more sensitive to solute-solvent interaction.

The volumetric property has been applied to study on hydration properties of nucleic acids, and drug interactions (43, 47, 48). Volumetric parameters are nonspecific probes of global hydration, while other techniques such as x-ray crystallography, which probed highly localized

water or NMR technique that probes immobilize water molecules. The crystallization of flexible proteins are challenging, and larger protein in NMR study can lead to sensitive loss to probe the topology of the molecules (49, 50). The quality of the X-ray structure depends on many factors that can help influencing the crystal quality, such as concentration, pH, ionic strength, and temperature (51). Beside the limitation, the X-ray diffraction and NMR studies have provided interlinked properties (microscopic perspective) on hydration patterns of nucleic acid. For protein and small-molecule complexes with DNA, X-ray diffraction and molecular dynamic simulation have revealed interfacial water molecules at the binding region (52, 53). A macroscopic characterization of hydration is necessary to quantify the hydration changes of the whole system in addition to specific interactions (microscopic events). Volumetric measurements such as volume and compressibility are nonselective, which sample the entire water population that interacts with the solute (54, 55). In protein studies, conformational transitions such as fluctuation of the intramolecular void region, intrinsic packing or protein hydration can be reflected by volumetric measurement (46).

To address the partial molar volume into hydration, the following equation can be applied:

$$V^{\circ} = V_M + \Delta V_h = V_{\mathrm{M}} + \mathrm{n_h}\,(V_{\mathrm{h}} - V_0) \qquad (6)$$

Where V_M represents a solute's intrinsic (molecular) volume that is excluded from water, ΔV_{h} is the volume contraction in the hydration layer relative to bulk solvent, $(V_{\mathrm{h}} - V_0)$ and $\mathrm{n_h}$ is hydration number.

Based on scaled particle theory, the effective void volume arising from mutual vibrational motions of solute and solvent molecules (V_T), need to be added to ΔV_h, given the complete aspect into partial molar volume as the sum of:

$$V^{\circ}=V_M+V_T+V_I+\beta_{TO}RT \quad (7)$$

The interaction volume V_I reflects solvent contraction due to solute–solvent interactions, a similar expression of $n_h\ (V_h - V_0)$, the β_{TO} is the coefficient of isothermal compressibility of the solvent, R is the universal gas constant, and T is the absolute temperature. The $\beta_{TO}RT$ can be neglected in macromolecular studies.

The partial molar adiabatic compressibility (K^{o}_S) is defined as the adiabatic compressibility of 1 mole of a solute at infinite dilution, following similar expression of partial molar volume:

$$K^{\circ}_S=K_M+\Delta K_{Sh}=K_M+n_h\ (K_{Sh}-K_{S0}) \quad (8)$$

Where K_M is the measurement of intramolecular interaction, ΔK_{Sh} is the change in solvent compressibility induced by hydration change. The term n_h in both application of partial molar volume and partial molar adiabatic compressibility carries the same description about the quantity of water molecules that change in reference to the solute (43).

The combination of volumetric study: density measurement for partial molar volume and sound velocities measurements for the partial molar adiabatic compressibility are used commonly by Chalikian's group. They have studies the hydration properties of small-molecule (Hoechst 33258), netropsin in DNA binding (44, 48), analyzing the urea-dependence volumetric properties

of proteins (56), the hydration characteristic of cyclic AMP with cAMP-binding domain (57), etc. On the study of effect of urea on protein's volumetric properties, the have found that apocytochrome c has similar hydration number (n_h) that calculated from partial molar volume and compressibility which are 660 ± 40 and 590 ± 10, respectively (56). Due to the similarity in hydration number between the two measurements, the change in number of hydration water related to the bulk (n_h) obtained from the change in partial molar volume is used to extrapolate other terms in compressibility such as the a change in the intrinsic compressibility of EPAC1 (57), or lysozyme that occurs by binding-induce change (58).

Other probing methods, such as the idea of "osmotic stress" in application to determine the number of water molecules releases or uptake upon binding as the effect of dispersion of water activity by osmolytes at equilibrium can be express using this equation:

$$-\left(\frac{\partial \log K_{\mathrm{D}}}{\partial Osm}\right) = -\frac{\Delta V_{\mathrm{w}}}{55.5 \ln 10} \tag{9}$$

where 55.5 is the molal volume of pure water and ΔV_{w} is the preferential hydration change between the unbound and bound states, the binding constant is $-\log K_{\mathrm{D}}$ taken as the equilibrium constant, and assuming complete exclusion in binding by osmolyte (59, 60).

In contrast with volume, density, and sound velocity measurement, which have shown water release upon binding of small-molecule to DNA (48), the osmotic stress studies have shown a net uptake of water (61, 62). For examples, the hydration changed upon binding of Hoechst 33258 with the minor groove binding site of the (AATT)2 by osmotic stress has shown a net uptake of 60±13 water molecules (61), while in volumetric study the binding of Hoechst to d(CGCGAATTCGCG)2 involves a net release to the bulk of 55 ± 8 water molecules (44). A study

on water activity of Hoechst 33258 - calf thymus DNA (CtDNA) has detected the net uptake of water is dependent on the osmolytes used. In the presence of triethylene glycol (TEG) as osmolytes, there is an the increased uptake of 74 ± 2 water molecules; while sucrose as osmolytes shows uptake of 30±1water molecules (63). As different hydration properties are observed in different techniques, we have combined a global and domain-specific probe by volumetric measurement (ΔV) and linked osmotic changes with ions (Na^+) not osmolytes, in conjunction with molecular dynamic simulation to study the hydration properties of DB1976.

1.4 The dynamic profile of short DNA fragment upon ligands binding

Oligonucleotides are synthetic short DNA fragments that are widely used in research applications to study drug-binding or protein-DNA binding characteristics. The binding of small molecules or proteins may induce both structural and dynamic changes in double helix. Using oligonucleotides as a tool to represent native double helix binding may underestimate the dynamic profile of binding. Therefore, the dynamic properties of oligonucleotides should be investigated to comprehend the binding profile of ligands.

The stability of DNA double helix is built by the stacking forces (van der Waals, dipole-dipole) establishing between adjacent overlapped bases and H-bond base pairing between complementary strands. However, the oligonucleotide is shortened in those properties, as proved that the thermal stability of oligonucleotide depends on the length, and stacking energy also depends on the sequence order.

The DNA duplex melting has been proposed to be initiate by both ends melting, called “the fraying-peeling mechanism” (64). The end terminal base pairs appear to be less stable and undergo the “zipper unfolding model”, a reversible terminal end opening events (65). The NMR study on the chemical shift of aromatic protons provided a thermodynamic profile of terminal base pairs of

B-DNA duplexes, which are enthalpically unfavorable and entropically stable. Comparing between the two duplexes 5'-d(CGCGATCGCG) and 5'-d(TAGCGCTA), the enthalpy and entropy value of base-pair opening of terminal C-G duplex are + 38. 0 kJ/mol and 0.110 kJ/mol K, respectively. While the enthalpy and entropy value of terminal T-A duplex are approximately +22. 8 kJ/mol and 0.077 kJ/mol K, respectively (66, 67). The thermodynamic cost of opening terminal ends is relatively small to the whole system profile. However, in couple with ligand binding, the end-opening dynamic of the bound DNA should be considered to fully capture the binding profiles of ligands. A recent study on Hoechst 33258 binding effect on base pairs nearby its binding site has concluded that small ligands can manipulate the base-pairs in neighbor with the binding site. Using The DNA dodecamer d (CGCAAATTTGCG)2, they found that up to 66% of G:C pair dissociation found upon complexes formation compared to the 17% possible end fraying in native DNA (68). The molecular dynamic simulation has also established multiple noncanonical structures present at the end of the duplex, in which terminal pairs are unstacked from their neighbor and break the H-bond between strands. (67) The dynamic of the overall DNA-binding system may be affected by these characters. A better understanding of the fraying dynamic based on DNA sequence and ligand-binding induce fraying is imperative.

1.5 Contributions to the project

I acknowledge all the contributed authors of the manuscript presented in Chapter 2 for all their performed date sets and analysis.

My contribution to the manuscript including the following details:

I performed all the experiments related to volumetric technique (density measurements) and analyzed molecular dynamic simulations.

I designed and optimized procedures on using the Densimeter Anton Paar Model DMA-5000. The machine is designed with a precision of 1.5×10^{-6} g/cm^3. My objective is to determine directly the volume change in binding. This is a challenging project because water in the hydration of DNA and proteins is different in (partial molar) volume from bulk water by only about 10%. During the titration experiments, the procedure needed to be carried out in the extreme performance such that no air bubbles were present in the column of the machine to minimize any negative effect on the results. Any mistakes found during the process, will change the density of the solution, and only small changes in the last digits could affect the calculation process and lead to failure. Also, the titration experiment had to be performed in one day (all data points are performed on the same day). For example, the whole titration curve could not be cumulative points of data that performed separately. Due to the sensitivity of the machine to the concentration of solutes, the titrations also had to be performed correctly on volume adding, which were carried out by hand. The optimized condition for the compound (buffer used for the experiments) was also determined by multiple efforts to reach a complete dissolved solution. The compound used in the project was also found to be sticky to plastic; therefore, I had to design a way to avoid direct contact of sticking the syringe plunger into the solution. The machine is extremely sensitive, and all of the preparation steps were carried out by hand works, which led to a potential failure of a large number of trials.

I also distributed to the analyses of molecular dynamic (MD) simulations. I performed the de novo docked of DB1976/DNA complexes (Figure 2.5); the analysis of the off-set position of the docked structures (Figure 2.6); the analysis of ends deformation and their relationships on hydrations (Figure 2.9). I also performed all the calculations and analysis of volume-related MD simulations (Table 4, 5).

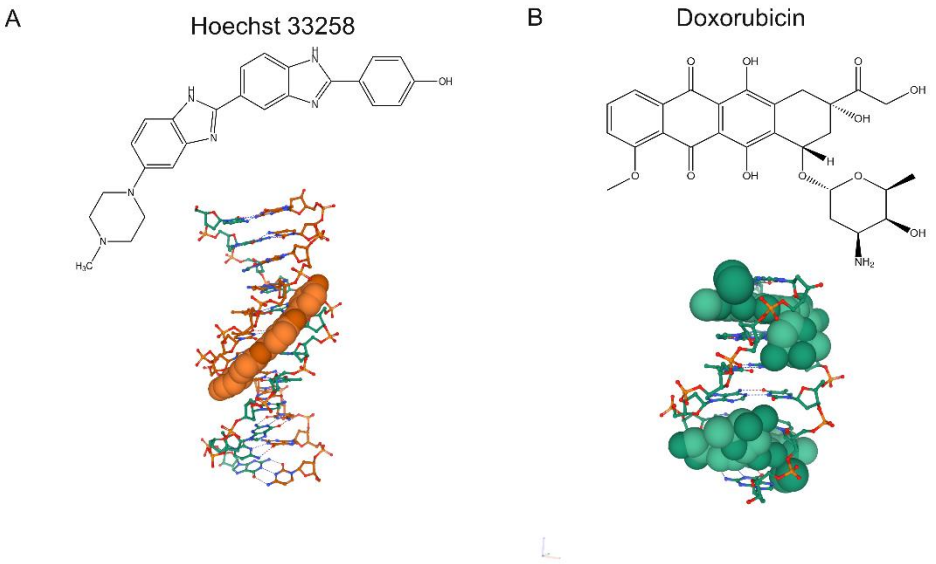


Figure 1-1 The small ligands-DNA binding modes

A. The minor groove binder - Hoechst 33258 to DNA dodecamer CGCGAATTCGCG (PDB:1D44). B. The intercalation binding mode between doxorubicin to D(CGATCG) (PDB: 1P20). The DNA is shown in ball and stick representation. The ligand is shown in space-fill (sphere) representation.

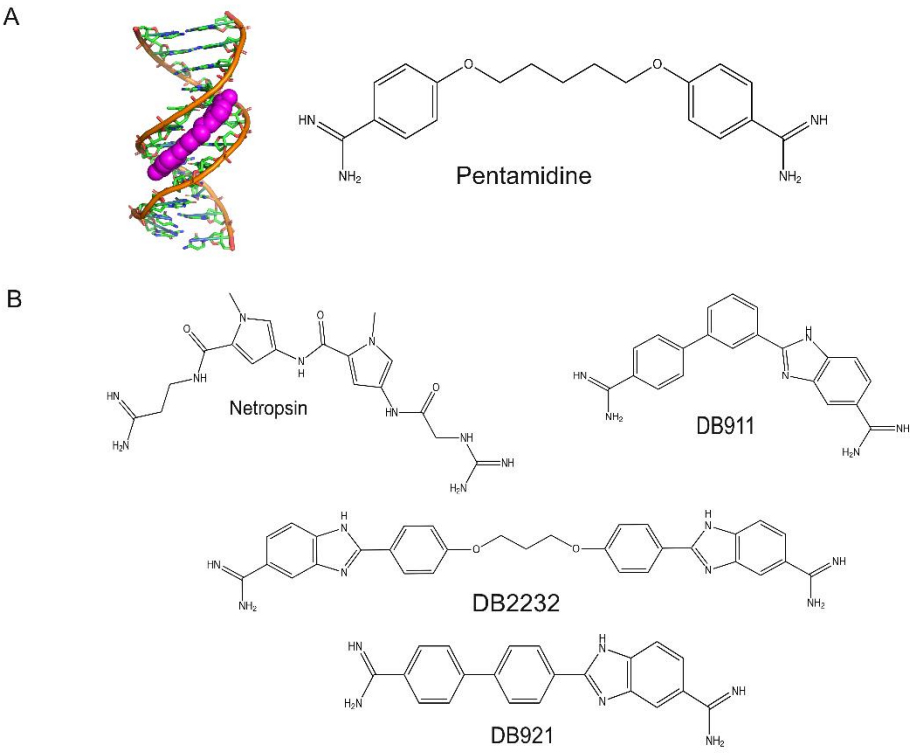


Figure 1-2 Examples of groove binders

A. The crystal structure of pentamidine to AT rich region of minor groove of CGCGAATTCGCG duplex (PBD: 1D64). B. A list of small molecule compounds that are mentioned in section 1.1.3

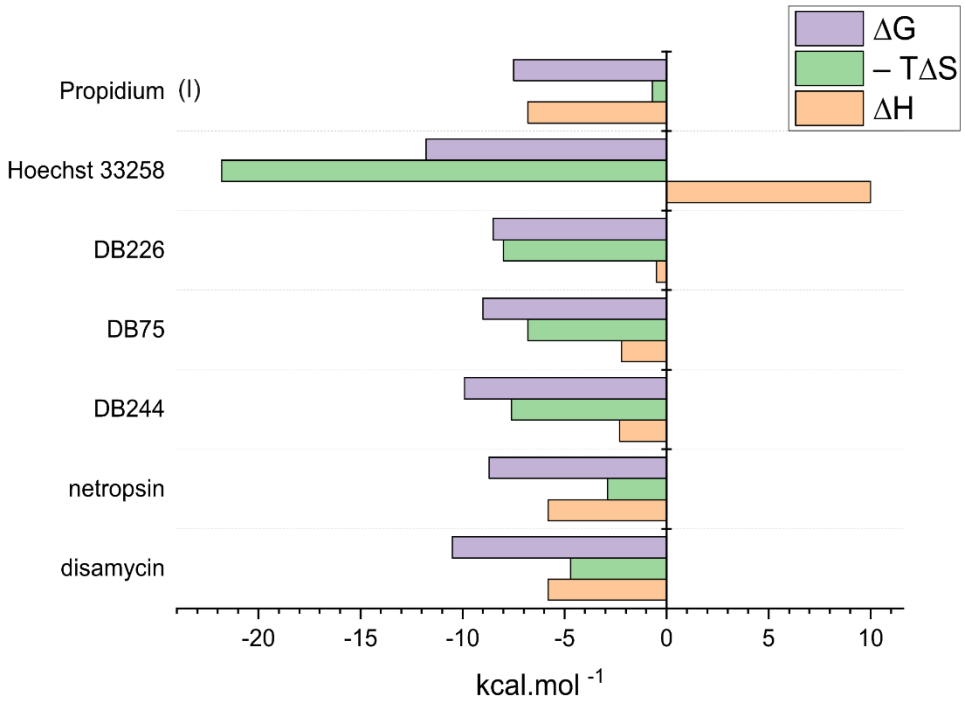


Figure 1-3 The thermodynamic profiles of small molecules-DNA complexes
In which the colors codes for binding free energies **(**ΔG), enthalpy values (ΔH), and entropy contribution (– TΔS) are purple, orange and green, respectively. The (I) indicates the intercalation binding mode, non-labeled molecules are minor groove binders.

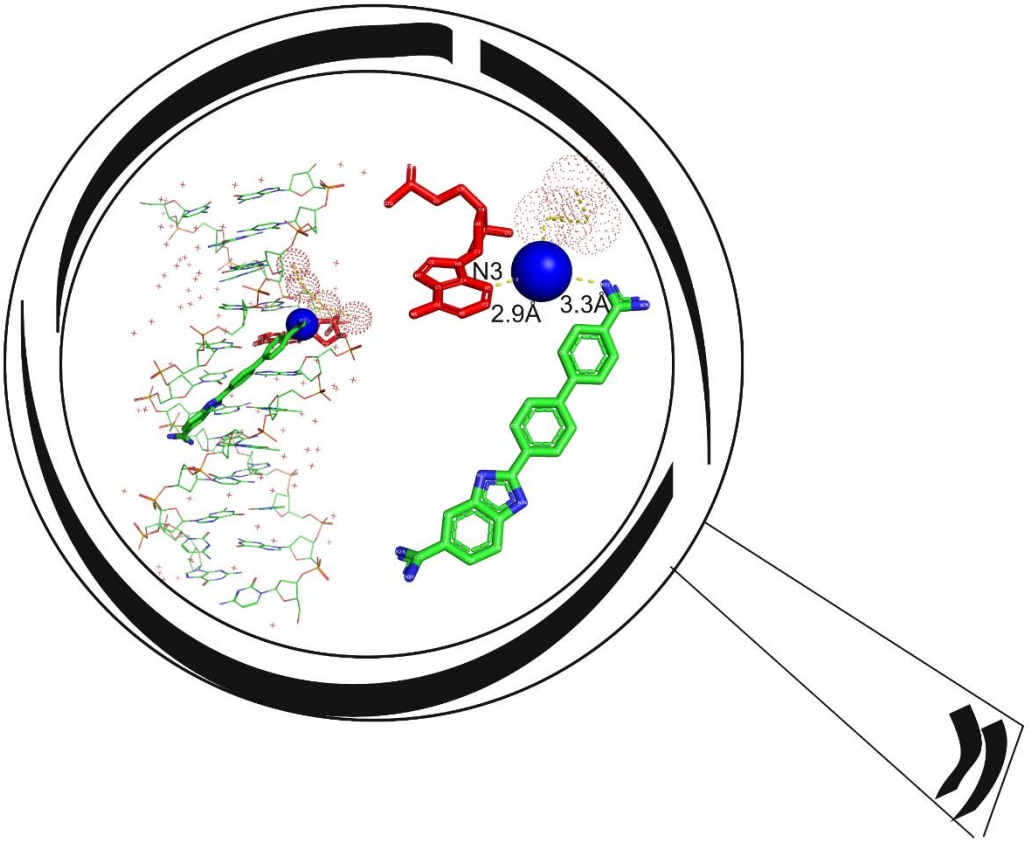


Figure 1-4 The crystal structure of DB921-D(CGCGAATTCGCG)2 complex (PDB:2B0K)

A. The left side represents the binding of DB921 into the minor groove of the duplex.. Where the blue sphere marks the water molecule that shows intermediate interaction between DB921 and A5 of the duplex. The ligand is coded in green. The small dots surround the complex are water molecules. An example of the water networks associate with the imtermediate water are colored with the bigger red-dot sphere. B. The right side is the zoom in portion to clearly show the distance between the water and atom N3 on A5 is 2.9 Å. The distance between the water and the N on the phenylamidine group of ligand is 3.3 Å. The base A5 of the duplex is colored in red.

2 DISSECTING THE DYNAMIC AND HYDRATION CONTRIBUTIONS TO SEQUENCE-DEPENDENT DNA MINOR GROOVE RECOGNITION

 [Biophysical Journal, Volume 119, issue 7, pages 1402-1415] 6 October 2020 DOI: 10.1016/j.bpj.2020.08.013

2.1 Abstract

Sequence selectivity is a critical attribute of DNA-binding ligands and underlines the need for detailed molecular descriptions of binding in representative sequence contexts. We investigated the binding and volumetric properties of DB1976, a model bis(benzimidazole)-selenophene diamidine compound with emerging therapeutic potential in acute myeloid leukemia, debilitating fibroses, and obesity-related liver dysfunction. To sample the scope of cognate DB1976 target sites, we evaluated three dodecameric duplexes spanning >103-fold in binding affinity. The attendant changes in partial molar volumes varied substantially, but not in step with binding affinity, suggesting distinct modes of interactions in these complexes. Specifically, while optimal binding was associated with loss of hydration water, low-affinity binding released more hydration water. Explicit-atom molecular dynamics simulations showed that minor groove binding perturbed the conformational dynamics and hydration at the termini and interior of the DNA in a sequence-dependent manner. The impact of these distinct local dynamics on hydration was experimentally validated by domain-specific interrogation of hydration with salt, which probed the charged axial surfaces of oligomeric DNA preferentially over the uncharged termini. Minor groove recognition by DB1976, therefore, generates dynamically distinct domains that can make favorable contributions to hydration release in both high- and low-affinity binding. Since ligand binding at internal sites of DNA oligomers modulates dynamics at the termini, the results suggest both short-

and long-range dynamic effects along the DNA target that can influence their effectiveness as low-MW competitors of protein binding.

2.1.1 Significance

Blockade of the DNA minor groove is increasingly recognized as a tractable strategy to inhibit transcription factors, by denying them access to the promoters and enhancers of genes. The disposition of hydration water is a signature driving force in minor groove binding, but the underlying hydration dynamics remain poorly understood. DB1976 is an attractive model compound for gaining new insight into target selection in the minor groove, besides its therapeutic relevance, due to widely dispersed affinities for a range of cognate DNA sequences. A combination of volumetric, osmotic stress, and molecular dynamics approaches reveal for the first time a heterogeneous response in DNA dynamics that manifests in the hydration of the resultant complexes.

2.2 Introduction

The DNA minor groove is a major molecular target for many DNA-binding compounds of interest. These include classical natural products such as netropsin and distamycin, as well as myriad synthetic derivatives that are used as markers in fluorescence-based imaging (*e.g.,* DAPI and the Hoechst compounds) and therapeutics (*e.g.,* pentamidine, furamidine, diminazene). In addition to specific contacts with the minor groove, the disposition of associated ions and water molecules impact the thermodynamics of binding. The characteristically favorable entropy change in minor groove binding under ambient conditions is typically ascribed to the release of ions and water of hydration (48, 69-71). The ionic component is classically modeled as the release of condensed counter-ions from DNA by the number of cationic charges on the ligands, and does not depend on the sequence of the DNA target (71-73). In contrast, the displacement of hydration

water is strongly sequence-dependent in favor of AT-rich sequences (47, 48). The details of hydration changes are not well understood, as the thermodynamically measured hydration change in solution is an order of magnitude higher than the "spine of hydration" observed in crystal structures (32). Moreover, recent studies in DNA hydration reveal significant heterogeneity in hydration dynamics in the ps-timescale that are highly sequence-dependent (74-76). The nature and dynamics of sequence-dependent hydration changes in minor groove binding are essential to understanding the energetics of minor groove recognition, as hydration numbers dominate ion release and are expected to strongly impact the free energy change of binding.

Previous experiments showed that sequence effects in nucleic acids hydration can be discerned through their volumetric properties (43, 77). To address the hydration dynamics that attend minor groove binding, we report a joint experimental and molecular dynamics study on the binding of the heterocyclic diamidine DB1976 to three cognate DNA sequences covering a wide range of binding affinities [**Figure 2.1**]. DB1976 exemplifies a generation of synthetic dications aimed at enhancing AT-targeting selectivity by optimizing the iso-helical placement of minor groove-binding modules, such as benzimidazole moieties (23). As a result, DB1976 is an attractive model for gaining new insight into the sequence context of minor groove recognition because its wide dispersal of affinities are experimentally inaccessible to weaker classical ligands (*vide infra*) such as the Hoechst dyes, DAPI, and netropsin (44, 71, 78). DB1976 is also translationally relevant because it exhibits biological activities as specific inhibitors of the transcription factor PU.1 and has shown therapeutic potential in PU.1-related acute myeloid leukemia (19), fibrotic diseases (21), and obesity-related liver dysfunction *in vivo* (20). Our characterizations reveal that oligomeric DNA harboring different AT sequence contexts responds to minor groove binding with dynamics and hydration changes that are out of step with binding affinities. The sequence-

dependent contributions arise from differential perturbations at the termini and interior of the double helix. These results have clear implications for the dissection of the configuration entropy change in binding, to which both hydration release and dynamics changes in the bound state make significant contributions.

2.3 Results

2.3.1 Volumetric characterization of DNA recognition by DB1976

As representative targets for DB1976, we selected the established dodecamer 5'-CGCGAATTCGCG-3' (A_2T_2) and two symmetric variants. 5'-CGCAAATTTGCG-3' (A_3T_3) was expected to capture the enhanced affinity of DB1976 for extended AT-rich sequences (19, 79). A third sequence, 5'-CGCAACGTTGCG-3' (A_2CGT_2) was an isomeric variant of A_2T_2 in which a CG step interrupted the central AT-track. Binding assays by fluorescence anisotropy confirmed that DB1976 bound with 1:1 stoichiometry to all three sequences in the presence of 50 mM NaCl [**Figure 2.2A**]. At this salt concentration, the affinities of DB1976 for A_2T_2 and A_3T_3 were unquantifiably high by fluorescence ($K_D << 10^{-10}$ M). This level of affinity was considerably higher than netropsin or Hoechst 33258 which bind A_2T_2 ~100 times more weakly under similar salt conditions (44, 71, 78). At 0.5 M NaCl, DB1976 bound A_3T_3 with a dissociation constant of 0.12 nM, near the limit of quantitation and ~5-fold higher affinity relative to A_2T_2, while binding to A_2CGT_2 was >600-fold weaker than A_2T_2 [**Figure 2.2B, Table 1**]. To establish the interrupted AT-track in A_2CGT_2 as a cognate binding site, we measured binding at 5 mM NaCl. The data showed identical anisotropy at saturation as other salt conditions and sequences (Figure 2.2B, gray points), demonstrating the absence of nonspecific aggregative binding even under permissive low-salt conditions.

Next, we measured the volume change ΔV of DB1976 binding to the three target DNA directly by high-precision densimetry at 50 mM NaCl. Volumetric titrations were conducted in reverse configuration in which an initial concentration of DNA at 70 µM was titrated with buffer or concentrated DB1976 in the same buffer [**Figure 2.2C**]. Expressed as the ratio r = [DB1976]/[DNA], ΔV decreased linearly with an equivalence point at unit r, in agreement with the 1:1 binding observed at much lower concentrations of compound. As the density-detected titrations were carried out at concentrations orders of magnitude above the equilibrium dissociation constants, we modeled the observed volumetric changes in terms of r as follows:

$$\Delta V = \begin{cases} \Delta V_{\text{obs}} \cdot r & r < 1 \\ \Delta V_{\text{obs}} & r \geq 1 \end{cases} \quad (1)$$

The observed volume change ΔV_{obs} included the volume change of complex formation $\Delta V_{\text{c}} = V_{\text{complex}} - V_{\text{DNA}} - V_{\text{DB1976}^{2+}}$ as well as the hydration volumes (6.9 mL/mol) (80) of two Na^+ counter-ions released from DNA backbone phosphates (79):

$$\Delta V_{\text{obs}} = V_{\text{complex}} - V_{\text{DNA}} - V_{\text{DB1976}^{2+}} = \Delta V_{\text{c}} + 2V_{\text{Na}^+} \quad (2)$$

On this basis, the three DNA targets yielded distinct volume changes upon saturation, A_2T_2 < A_3T_3 < A_2CGT_2 (most to least negative), which were out of order with affinity. The partial molar volumes of the three duplexes and DB1976 alone (Figure 2.3) were then used to compute the corresponding values for the complexes using the experimental ΔV_{obs} in Table 1.

2.3.2 Structural basis of DB1976/DNA binding

Minor groove insertion by linear heteroaromatic compounds is structurally well established (81), and has been confirmed specifically for DB1976 in our previous DNase I footprinting and circular dichroism spectroscopy experiments (23). The experimentally indistinguishable fluorescence anisotropies of the three complexes (Figure 2.2B) strongly support a shared binding pose. On this basis, we carried out explicit-solvent molecular dynamics (MD) simulations using the DNA-optimized forcefield parmbsc1 (82). The DNA and compound in both free and bound states were simulated under conditions as close as possible to the volumetric experiments (50 mM NaCl, 25°C). Construction of the models is described in *Materials and Methods*. We used the TIP4P-ew water model which gives bulk properties in agreement with experimental values (83). Following energy minimization and equilibration, the unrestrained *NPT* ensemble was simulated for 600 ns. Equilibration of the unbound species as well as their complexes were judged by their RMS deviations and the burial of DB1976 in the minor groove [**Figure 2.4**]. As 100 ns was more than sufficient to achieve convergence of local water and ion densities to 0.5 Å resolution (84), the final 500 ns of trajectory was used for analysis.

To establish the physical relevance of the MD results, we first examined the contacts made by DB1976 in the three complexes. Using standard criteria of a 0.35 nm distance and 30° angle cutoffs for hydrogen bonds, we computed the frequency of H-bonds between DB1976 and the DNA. Despite the structural symmetry of DB1976 and palindromic DNA sequences, the ligand was offset on the bound DNA by a half base step. This offset was also present in complexes docked *de novo* to the three sequences (Figure 2.5). To test if this orientation was dependent on the choice of initial coordinates, we ran control simulations in which DB1976 was initially positioned at the geometric midpoint of the DNA [**Figure 2.6**]. The compound shifted to the offset position with

sequence-dependent kinetics and relaxation times within 1 ns. The evolution of DB1976-DNA H-bonds, which tracked the translation of the compound in the minor groove, suggested that the offset position in the complexes optimized drug-DNA contacts in the minor groove. Since H-bonding is directionally sensitive, this behavior suggests a dynamic alignment of H-bond donors and acceptors on the ligand and DNA.

Focusing more closely on the ligand/DNA contacts, the three complexes showed qualitatively similar H-bonding patterns. H-bond contacts were made chiefly by the terminal amidiniums and inward-facing benzimidazole NH groups of DB1976 to contacts at the minor groove floor of both strands. The contacted DNA atoms are N3 of A, O2 of T, O2 of C, and N3 of G, which are canonical H-bond acceptors in the minor groove, as well as the exocyclic N2H of G. **Figure 2.7** quantifies the frequencies of contacts made at 10% or higher occupancy over 100-ps intervals (50 frames). To establish sufficient sampling, we ran triplicate simulations for the A_2T_2 complexes with different random initial velocities. Corresponding H-bond occupancies were reproduced to within 10% among the replicates [**Table 2**]. While the overall frequencies were similar between the complexes with A_2T_2 and A_3T_3, the H-bond frequency for the A_2CGT_2 complex was ~20% lower, and two fewer DNA bases were contacted. The simulations therefore affirmed expectations that the benzimidazole NH groups avidly preferred H-bond acceptors in AT base pairs. Water-mediated contacts (i.e., bridging water), made exclusively by the amidinium moieties, were negligibly transient (defined as appearing in fewer than 10% of frames) in both the A_3T_3 and A_2T_2 complexes, but more significant in the A_2CGT_2 complex. Weaker complexes therefore partially compensated for fewer favorable direct contacts by benzimidazole with water-mediated contacts at the termini.

2.3.3 Dynamics of DB1976-bound DNA

For analytical purposes, we separated dynamics at the termini from those at interior positions of the duplex. Since helix-coil transitions are associated with significant hydration changes (85), altered dynamics at the termini may contribute to the hydration changes in binding. Non-canonical conformations have been reported to not be accurately modeled by previous-generation forcefields for DNA (67), such as parmbsc0 (86). However, their current-generation successors such the parmbsc1, which was used in our simulations, are significantly improved in suppressing artefactual fraying and their agreement with experimental structures of A_2T_2 (82, 87). With this in mind, we evaluated the dynamics of the termini in both free and DB1976-bound duplexes. We first enumerated the Watson-Crick H-bonds in the each of the terminal GC base pairs. We defined end deformation as the absence of all three H-bonds in a terminal GC base pair [**Figure 2.8**]. Triplicate simulations for the three sequences showed that deformation at any one terminus could vary substantially. However, the combined frequencies of the two termini of a sequence were reproducible to within ±5%. Comparison among the complexes showed that DB1976 binding reduced deformation in the A_3T_3 complex with no statistically significant effect on the other two complexes (Figure 2.8C).

To scrutinize the dynamics of the termini in greater detail, we characterized their conformations in terms of separation between the centers of mass (COM) of terminal GC base pairs. These trajectories showed well-defined distance régimes that distinguish intact base pairs from two types of deformation [**Figure 2.9A**]. Separation distances shorter than 5.4 Å corresponded to non-canonical stacking of the terminal bases, while separation over 5.9 Å was associated with fraying of the ends [**Figure 2.9B**]. To assess the hydration of these states, we compared their radial distribution functions (RDF) and H-bonding characteristics with water. The

RDF between non-H atoms in the termini and water oxygen increase progressively to the bulk value of unity due to the bulk of the DNA. All the RDFs showed local maxima near 2.8, 3.8, and 5.0 Å. The first maximum at 2.8 Å, which was taken characteristically as the first hydration layer, was higher for both deformed states relative to the intact base pair. The deformed termini were more hydrated due to solvent penetration into the buckled terminus in the stacked state and exposure of the penultimate base pair in the frayed ends. Over longer distances, between 5 and 10 Å, the frayed state was the most hydrated due to protrusion of the ends into the solvent. These RDFs were mirrored by H-bond densities with water. An assignment of the termini to the intact and deformed states showed notable sequence-specific variations [**Figures 2.9C**]. In agreement with the intra-bp H-bond analysis (Figure 2.8C), only A_3T_3 showed significant gain in intact termini upon binding DB1976. Both stacked and frayed states were lower in the DB1976-bound A_3T_3 complex. In contrast, changes in deformed and intact states were overall mixed for A_2T_2 and minimal for A_2CGT_2. As intact termini were less hydrated than deformed states, we enumerated the number of water molecules within 15 Å (based on the RDFs in Figure 2.9B) of the terminal and penultimate base pairs (the latter exposed in frayed termini) over the last 500 ns. Histograms of the local water showed long tails and larger changes in dispersion upon DB1976 binding for A_3T_3 (as evident from the height of the distribution) relative to the other two sequences [**Figure 2.9D**]. To quantify the excess kurtosis in the distributions, we examined the median absolute deviation $\mathrm{MAD}(\mathbf{x}) \equiv \mathrm{median}\left(\left|x_i - \mathrm{median}(\mathbf{x})\right|\right)$, a robust measure of dispersion [**Figure 2.9E**]. Only A_3T_3 exhibited a significant reduction in MAD arising from the loss of the most highly hydrated conformations in the ensemble. Thus, DB1976 binding altered sequence-dependent dynamics beyond its interior binding site in the minor groove, and the resultant net gain in intact termini represented a source of excess hydration release for the A_3T_3 complex relative to A_2T_2.

Proceeding to the interior dynamics of the interior helix, we examined the helical curvature as an intuitive metric that integrates over perturbations of local helical parameters. For each frame, we summed the absolute geometric curvature κ of a globally fitted the helical axis over each interior base step [**Figure 2.10A**]: $\kappa = \frac{d\theta}{dh}$, where θ is the tangential angle and *h* is the arc length along the axis. Summary expressions of κ or RMSF of atomic positions suggest differences among the sequences and states [**Figure 2.11**], but without considering the time-ordered nature of the trajectories, they do not readily distinguish the underlying dynamics. To tackle the time series explicitly, each curvature trajectory was differenced to generate a stationary time series of signed step sizes $\Delta(t) : \Delta(i) = \kappa(i+1) - \kappa(i)$, where *i* is the frame index. As exemplified in **Figure 2.10B** for free A_2T_2 over a 50 ns period, Δ(*t*) (lower panel) revealed clustered changes whose histograms yielded tall peaks and long tails [**Figure 2.10C and inset**]. Such non-Gaussian (leptokurtic) distributions are characteristic of auto-correlated dynamics. To quantify these dynamics, known as stochastic volatility (88, 89), we modeled Δ(*t*) as a random walk with variable step size *s*(*t*). Stochastic volatility arises when *s*(*t*) is itself a random variable, which we modeled as drawn from a normal distribution with standard deviation σ:

$$\begin{aligned} \Delta(t) &= s(t)W(t) \\ s &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3}$$

where *W*(*t*) is a standard Brownian (Wiener) process. σ is the target parameter that specifies the volatility dynamics. A limiting value of $\sigma \to 0$ (constant *s*) represents simple Brownian random walks. Large values of $\sigma > 0$ correspond to a wider dispersion in the distribution of Δ(*t*), and more volatile curvature fluctuations. To estimate σ, we performed Bayesian inference via Markov Chain

Monte Carlo (MCMC) simulations. **Figure 2.10D** shows an overlay of 50 MCMC traces in green for $s(t)$ out of 5,000 total for the 50-ns period shown in Figure 2.10B, to intuit the model's ability to track the time-dependent volatility in the curvature fluctuations. To assess the goodness of fit of the model, we performed posterior predictive checks (PPC) on the statistics of the posterior distributions. More precisely, we compared the medians and median absolute deviations from the MCMC traces with the corresponding values from the MD trajectories (Figure 2.11B). As illustrated for A_2T_2 in **Figure 2.10E**, the PPC statistics were within 5% of those exhibited by the MD trajectories. To further test the specificity of the model, we challenged it with a simulated Brownian random walk without volatility *i.e.,* $\sigma = 0$ and constant s [**Figure 2.12**]. The model was able to return the correct estimate for s and a negligibly small estimate for σ.

For all sequences and states, Bayesian estimates for σ were drawn from posterior distributions as means and uncertainties given by the 95% highest posterior density (HPD, also known as the credible interval; **Figure 2.10F** and parametric values in **Table 3**, σ showed distinctive differences among the three DNA duplexes. DB1976-bound A_3T_3 and A_2T_2 showed significant reduction in volatility relative to unbound DNA. In stark contrast, the curvature fluctuations in bound A_2CGT_2 were more volatile than the unbound DNA. Thus, volatility modeling distinguished the autoregressive fluctuations in helical curvature between high-affinity A_3T_3 and A_2T_2 from A_2CGT_2. Taken together with the end deformation results, MD simulations showed that the interior and terminal portions of the DNA duplex behaved as distinct dynamic domains that in response to minor groove binding in a sequence-dependent manner.

2.3.4 Analysis of hydration changes from volumetric data

In addition to providing a high-resolution window into the drug/DNA complexes, the MD simulations provide structure-based parameters needed for a microscopic analysis of the experimental volume changes. A standard molecular description of volumetric changes is to dissect it into three additive contributions (77):

$$\Delta V_{\mathrm{c}} = \Delta V_{\mathrm{M}} + \Delta V_{\mathrm{T}} + \Delta V_{\mathrm{I}} \tag{4}$$

V_{M} represents a solute's intrinsic (molecular) volume that is excluded from water defined as a probe of radius 1.4 Å. V_{T} is an effective void volume arising from mutual vibrational motions of solute and solvent molecules. The interaction volume V_{I} reflects solvent contraction due to solute–solvent interactions. In Eq. (4), we neglect the compressibility contribution, which is on the order of 1 cm^3/mol for aqueous solutions.

The final 500 ns of each MD trajectory was used to compute the time-averaged changes in solvent-excluded volume ΔV_{M} and solvent-accessible surface areas (ΔSASA). ΔV_{M} was computed from frames extracted every ns and averaged in five equal blocks. Thermal volume was computed by a surface-area approach whereby ΔSASA (similarly block-averaged ± S.D.) is split equally between the constituents and scaled by the effective thicknesses of their thermal layers δ (77):

$$\Delta V_{\mathrm{T}} = \delta \cdot \Delta\mathrm{SASA} \tag{5}$$

where δ was taken as 0.55 Å as suggested by Chalikian and Macgregor for non-protein solutes (77, 90). With $\Delta V_{\mathrm{I}} = \Delta V_{\mathrm{c}} - \Delta V_{\mathrm{M}} - \Delta V_{\mathrm{T}}$, the hydration change of binding (Δn_h) was given by

$$\Delta n_h = \frac{\Delta V_{\mathrm{I}}}{V_h^0 - V_0^0} \tag{6}$$

where the volume contraction in the hydration layer relative to bulk solvent, $V_h^0 - V_0^0$, is taken to be -1.8 cm^3/mol.

As **Table 4** shows, molecular volume contributions ΔV_{M} were small in magnitude relative to ΔV_{T}. The negative thermal volume contributions ΔV_{T}, which were 10-fold larger in magnitude relative to ΔV_{M}, showed little differences among the three substrates (within 1%). As a result, the changes in interaction volume and magnitude of global hydration release ($|\Delta n_h|$) [**Table 5**] gave the rank order: $A_2CGT_2 > A_3T_3 > A_2T_2$.

2.3.5 Domain-specific analysis of hydration changes

The differences in global hydration change among the three complexes were remarkable in that the time-averaged ΔSASA values were essentially identical for binding to all three DNA. Using a nominal cross section of 9 Å² per water, ΔSASA alone corresponds to a release of ~86 water molecules, an amount bracketed by the range of observed Δn_h. Clearly, this analysis does not consider dynamic contributions to hydration. Hydration release due to a net gain of intact termini in the A_3T_3 complex is consistent with its Δn_h relative to A_2T_2. What would account for the even greater hydration release for the A_2CGT_2 complex? The MD results show that the duplex termini and interior behave as dynamically distinct domains in response to minor groove binding.

If the dynamics of the duplex interior also contribute to the hydration change of binding, the distinct dynamics of the three duplexes could account for their observed hydration ranking.

To test this notion, we sought a hydration probe that could differentiate contributions arising from ends and interior (axial) positions of the DNA helix. Synthetic oligonucleotides used in experiments bear uncharged 5' and 3' hydroxyl termini [**Figure 2.13A**] with near-zero surface electrostatic potential [**Figure 2.13B**]. Given this feature, the linked osmotic changes with ions would selectively probe the hydration population at strongly anionic axial positions. Hydration changes at the termini could be taken as silent to Na^+ (91, 92). As previously described (79, 93, 94), the hydration change in binding can be estimated from the curvilinear dependence of the equilibrium dissociation constant K_D on the mean ionic activity, $a_{\pm}$:

$$-\left(\frac{d\log K_D}{d\log a_{\pm}}\right) = \Delta\nu_{\pm} - \frac{2m_{\pm}}{55.5}\Delta\nu_w \qquad (7)$$

where $\Delta\nu_{\pm}$ and $\Delta\nu_w$ represent the linked disposition of ions and water upon binding (negative values denote net release), and $m_{\pm}$ is the mean ionic molality. The first term on the right side of Eq. (7) is strictly the limiting slope for the log-log dependence of the binding constant on the mean ionic activity. Increasing ionic activity modifies the slope depending on whether the linked hydration change is release (convex) or uptake (concave), respectively.

The applicability of the linkage of ion and water to DNA binding as embodied by Eq. (7) requires a valid formulation of the equilibrium constant (*i.e.,* 1:1 binding with the release of ions), which is well justified by both the fluorescence polarization and densimetric titrations. In the general case, linkage parameters $\Delta\nu$ are not guaranteed to represent physical stoichiometries (60, 94). However, in the specific case of DNA and monovalent salts, condensation (95, 96) and

Poisson-Boltzmann theories (97, 98) support the identification of $\Delta\nu_{\pm} = \Delta n_{\pm}$, the physical number of counter-ions transacted upon binding by ionic ligands. This correspondence is widely used to analyze electrostatic contributions in ligand-DNA interactions in the literature. Using a counter-ion condensation formulation, the ion number may be decomposed as follows:

$$\Delta n_{\pm} = -\psi z + \Delta n_{-} \tag{8}$$

The first contribution represents counter-ion release from the ionic atmosphere around the DNA: z is the number of phosphate neutralized, and ψ is a coefficient that describes contributions from condensation and screening effects (95, 96). The second term Δn_{-} accounts for any residual anion release from the cationic ligand (DB1976·2H^{+}). The hydration change, now $\Delta\nu_{w} = \Delta n_{w}$, refers to the attendant change in number of water molecules from the DB1976/DNA complex consequent to ion release (99). Since cations preferentially sample the anionic duplex interior, a sub-domain of the total hydrated volume, we expect $|\Delta n_{w}| < |\Delta n_{h}|$ where Δn_{h} is the global hydration change as given volumetrically in Eq. (6).

Experimentally, we fitted the dependence of the dissociation constant on mean ionic activity $a_{\pm}$ to the integrated form of Eq. (7), substituting Δn for $\Delta\nu$:

$$-\log K_{D} = -\log K_{0} + \Delta n_{\pm} \log a_{\pm} - \frac{2\Delta n_{w}}{55.5 \ln 10} m_{\pm} \tag{9}$$

where $-\log K_{0}$ is the constant of integration and $m_{\pm}$ depends on $a_{\pm}$ according to the interpolating function:

$$m_{\pm} = \varphi_1 a_{\pm} + \varphi_2 a_{\pm}^2 \quad (10)$$

The coefficients φ depend on salt identity. For NaCl, φ_1 = 1.64 and φ_2 = -0.113 (100). With molar NaCl concentrations converted to $a_{\pm} / m_{\pm}$ as independent variable, the fitted parameters are $\Delta n_{\pm}$, Δn_w, and $-\log K_0$.

Figure 2.13C shows that binding to the three sequences could be perturbed with NaCl over a 10^4-fold difference in affinity, equivalent to ~30 kJ/mol in $\Delta\Delta G°$ at 25°C. Given the structural similarity in the three DNA and complexes, the data was globally fitted with Eq. (9) with a shared $\Delta n_{\pm}$. The model rendered a good fit to the data and gave an estimate of $\Delta n_{\pm}$ = -2.4 ± 0.1. To decompose $\Delta n_{\pm}$ using Eq. (8), we took *z* = 2 for DB1976, as previous studies have established a 1-1 correspondence between the number of positive charges on the minor groove ligand and *z* (71-73). For the 12-bp synthetic oligomers used here (11 × 2 = 22 phosphates), ψ = 0.77 (101). These considerations yielded the release of $|-z\psi|$ = 2 × 0.77 = 1.5 condensed Na^+ from two neutralized DNA phosphates. To check for non-colligative contributions to cation release, which would indicate deviation from condensation theory or salt-induced changes in DNA structure, we tested the effects of substituting NaCl with KCl. Replacement of the cation produced the same $\Delta n_{\pm}$ values within experimental uncertainty across the tested salt range [**Figure 2.14**]. We therefore concluded that ionic component of DB1976 binding consisted of the release of condensed counter-ions and residual Cl^- from the DB1976 dication.

For the hydration component, Δn_w values for A_2T_2 and A_3T_3 were equal within experimental error, at Δn_w ~ -36 ± 3, and significantly smaller in magnitude than the value for

A_2CGT_2 at Δn_w = -50 ± 4 [**Table 6**]. All Δn_w values were ~50% smaller in magnitude than their respective values of Δn_h (*c.f.* Table 5), as expected given the domain-specific nature of Na^+. Since Na^+ probed only the axial (interior) surface, the similar Δn_w values for A_3T_3 and A_2T_2 indicated that the excess *global* hydration release Δn_h for A_3T_3 over to A_2T_2 was due to contributions from the DNA termini. The MD results have shown that deformed termini were over-hydrated and DB1976 binding uniquely stabilized termini of A_3T_3. We could therefore now attribute end stabilization as the source of excess hydration release from the A_3T_3 complex relative to A_2T_2.

In contrast, the A_2CGT_2 complex released significantly more hydration water than the other two sequences when probed by Na^+ data, indicating that the interior portion of the DNA was responsible for the excess hydration release. Stochastic volatility modeling has revealed that interior helical dynamics are uniquely induced in this complex. To better understand the consequent effect on hydration, we analyzed the trajectories of water hydrating the interior of free DNA and DB1976-bound complexes from MD simulations. Excluding the DNA termini, we defined the first hydration layer as water O atoms within 5 Å from a non-H atom of the solute. This criterion was derived in analogy with contact definitions by Halle (102) and Hamelberg (103) for proteins. Non-overlapping instances of the hydration ensemble were followed in time, yielding decay curves over 3 ns such as shown in **Figure 2.13D**. The time profiles $N(t)$ represented the net efflux of water from the first hydration layer as a result of water-water and water-solute interactions within the layer as well as translational diffusion into and out of the layer (104). The profiles revealed significant sequence-dependent differences in their free (F) and bound (B) states. Quantitatively, the non-linear appearance of the decay on semi-logarithmic scales indicated distinct kinetics on distinct timescales. Fitting the phenomenological profiles with exponential functions required three phases to fully capture for the initial hydration population [**Figure 2.15**]:

$$N(t) = N_1 e^{-\frac{t}{\tau_1}} + N_2 e^{-\frac{t}{\tau_2}} + N_3 e^{-\frac{t}{\tau_3}} \quad (11)$$

For convenience, we ordered $\tau_1 < \tau_2 < \tau_3$ (most rapid to slowest). We fitted 5 non-overlapping trajectories instantiated 200 ps apart during the final 100 ns of each simulation. The parameters are presented in **Figure 2.13E** and **Table 7** as means ± S.E.

The coefficients $N(0) = \sum_{i=1}^{3} N_i$ represent the population of hydration water. For all constructs, the majority of $N(0)$ were captured by phases i = 1 and 2 with lifetimes of τ_1 ~ 10 ps (~30%) and τ_2 ~ 80 ps (~55%). The slowest phase described the remainder (i = 3, ~10%) in the 10^2-ps timescale. The DB1976 complexes with A_2T_2 and A_3T_3 showed reduced $N(0)$ relative to the unbound DNA: $\Delta N(0)$ = -8 ± 4. This reduction arose mostly from the fastest-phase sub-population N_1. In comparison, A_2CGT_2 released significantly more water upon binding: $\Delta N(0)$ ~ -20 ± 5, due primarily to a loss of ~50% of the water from N_3 of the slowest phase. The differences in $\Delta N(0)$ between the complexes and respective free DNA accounted for the experimental differences in Δn_w among for the three sequences. Parenthetically, we did not include hydration of the *free* compound. This did not detract from the analysis because such a contribution would be constant for all three complexes.

In terms of the decay lifetimes, the A_2T_2 and A_3T_3 complexes showed accelerated hydration decay in all three phases relative to unbound DNA (Figure 2.13E). Their reduction in lifetimes were similar and was greatest for the slowest phase. In stark contrast, the A_2CGT_2 complex exhibited similar decay lifetimes as the free DNA. To probe these kinetics in greater detail, we scrutinized the trajectories of the water molecules individually for the three complexes. We noted

that many water molecules exited and re-entered the first hydration layer repeatedly until the final member exited the layer for the last time [**Figure 2.16**]. Analysis of the re-entrant hydration showed a significant shift to higher frequencies of re-entry for the A_2CGT_2 complex over the other two complexes [**Figure 2.13F**]. Thus, the interior portion A_2CGT_2, which was the most dynamic among the three complexes, was the most recurrently hydrated by local water molecules. Since the conformational dynamics of the solute were in relative motion with the translation of water molecules, the results suggested that increased conformational dynamics could entrap local hydration water and transiently re-populate the hydration layer. Nevertheless, since the water was ultimately lost to the bulk, re-entrant transport did not alter the release characteristics as defined by initial populations N_i. In summary, the hydration decay profile of the A_2CGT_2 complex suggested complex dynamics whereby increased conformational dynamics resulted in a net hydration loss even as the turnover of water in the hydration layer was slowed by re-entrant statistics.

2.4 Discussion

A combination of global and domain-specific interrogation of DNA hydration, the former by direct volumetric measurements and the latter using Na^+ as probe revealed distinct hydration contributions to minor groove binding by the DB1976 dication. The use of co-solutes to probe the osmotic properties of biomolecules depends on the preferential interactions or exclusion between the co-solutes and biomolecules (60, 94). More precisely, co-solutes detect the hydration changes associated with the surfaces to which a co-solute preferentially interacts or from which it is excluded (105, 106). This feature is often regarded as a weakness of probing hydration with co-solutes. In the present case, we have taken advantage of this thermodynamic linkage, Eq. (7), to dissect the hydration contributions along the axial surfaces of the DNA, with which Na^+

preferentially interacts, and contributions arising from the uncharged DNA termini, which can be taken as non-reporting sites for Na^+. Volumetric parameters such as volume and compressibility are nonspecific probes of global hydration (43, 77). With appropriate controls, monovalent cations represent accessible domain-specific hydration probes that experimentally inform the global hydration analysis of DNA and other poly-anionic solutes.

On the computational side, extensive MD simulations of DNA hydration dynamics provided two important outcomes. First, the results show that the termini and interior of the duplex are dynamically distinct domains. Each domain is linked microscopically in divergent ways to hydration properties in the ps-ns timescale. Second, perturbation of these dynamics by minor groove binding alters hydration in a manner that relates directly to experimental hydration detected volumetrically and osmotically. As the three DNA sequences demonstrate, the impact on hydration is non-continuum with respect to binding affinity. That the lowest-affinity sequence (A_2CGT_2) is associated with the largest hydration release implies that its enthalpic contribution to binding should be significantly poorer than the non-interrupted AT tracks. The relative deficits in intermolecular H-bonds in the A_2CGT_2 complex (Figure 2.7) are consistent with, albeit do not prove, this notion. In the case of A_3T_3, improved base pairing in intact termini may be expected to be a favorable enthalpic contributor that is not present for the other two sequences.

The identical Na^+-detected hydration release for both A_3T_3 and A_2T_2 indicates that the excess hydration release upon binding A_3T_3 is contributed by reduced deformation of the ends, which are silent to Na^+. Chalikian and coworkers have reported the hydration change accompanying the helix-coil transition of CCGTAATGCC, a decamer with similar composition as A_2T_2 (85). The helical state is associated with an average hydration deficit of 18 ± 1 water molecules per base pair relative to the coiled state. If we consider, as a first approximation,

deformation of the two terminal base pairs as dissociation, and DB1976 binding reduces deformation in A_3T_3 by ~20% over binding to A_2T_2 (as suggested by H-bonding and distance criteria in Figure 2.9C), one estimates an excess release of 18 × 2 × 20% = 7 water molecules from end healing. Relative to the observed $\Delta\Delta n_h$ ~ -26 for A_3T_3 over A_2T_2, this suggests involvement of hydration beyond the first hydration layer. The excess hydration of deformed termini beyond the first hydration layer (>5 Å in Figure 2.9B) relative to intact termini supports this interpretation.

Kinetically, the decay profiles for the first hydration layer of the three sequences and their complexes exhibit multiple lifetimes, ranging from 10 to ~500 ps. These phenomenological kinetics reflect both intrinsic H-bonding dynamics as well as water flux into and out of the hydration layer (104). Water flux is strongly influenced by re-entrant behavior, particularly at longer timescales, and is favored by conformational dynamics as the A_2CGT_2 complex shows. Although re-entrant transport strongly impacts the kinetics of hydration, equilibrium parameters such as volume and hydration numbers depend on the net changes in the hydration population. The loss of water from the slowest-decaying population of the A_2CGT_2 complex is consistent with the mobilization of slow hydration water, which has previously been suggested (107), as a dynamic contribution to hydration changes. However, the three phenomenological classes of water modeled from decay profiles should not necessarily be interpreted as non-interconverting sub-populations. In the literature, the Hynes and Laage groups have described the microscopic hydration of A_2T_2 by MD simulations (75). Their results show that hydration water exhibits broad heterogeneity in reorientation lifetimes spanning almost 50-fold, depending on the physicochemical character and topography of the exposed surface. Changes to both properties due to ligand binding may therefore alter the distribution of water among the hydration sub-populations.

While the MD simulations provided microscopic explanations for the sequence-dependent differences in hydration properties of minor groove recognition, absolute equivalence between experimental and hydration numbers from MD simulations will in general be limited at some level. This is due to assumptions in the decomposition of volumetric observables for the macromolecules on the one hand (90), and the approximation of simulational solution conditions on the other hand *e.g.,* choice of cutoff for the hydration layer(s), omission of the buffer species. Given a specification of the solution composition, one may in principle quantify hydration from simulations exactly via Kirkwood-Buff (KB) theory (108, 109) and its inversion (110, 111). However, simulations remain dependent on the specific choice of forcefield, water model, and other details. Thus, absolute estimations of hydration from experiments and simulations involve some assumptions but agree sufficiently for the present characterizations given sufficient differences among sequences and states.

In conclusion, we have combined thermodynamic experiments and simulations to gain insight into minor groove binding by DB1976. Hydration release contributes positively to DB1976 binding, regardless of sequence. However, more hydration water is released upon binding to A_3T_3 and A_2CGT_2, which bind significantly more and less strongly than A_2T_2. Experimental characterization of weaker cognate binding, such as A_2CGT_2, is enabled by the high intrinsic affinities of extended benzimidazole ligands such as DB1976 that are not attained by older minor groove ligands. As a compound with actual therapeutic potential (19-21), interactions with a wide range of cognate sequences which occur frequently in the genome represent foundational knowledge for pharmacologic targeting of DNA in the minor groove (112). In the case of DB1976, the ability of A_3T_3 to perturb the terminal dynamics of oligomeric DNA implies propagated effects that are also expected to operate in polymeric DNA. A_2CGT_2 binding suggests hydration release

from induced dynamics in the bound state as a potential compensatory contribution in low-affinity binding. Both aspects represent novel features of minor groove recognition that deserve further investigation.

2.5 Methods and materials

2.5.1 Compound and DNA

The synthesis of DB1976 has been previously reported (23). Elemental analysis showed that the compound purified as a tetrahydrochloride salt. Deoxyoligonucleotides were purchased from IDT (Midland, IA) as lyophilized solid. The DNA was dissolved in at ~2 mM (strand) and annealed by slow cooling from 95°C. Once reconstituted, the DNA targets were exhaustively co-dialyzed against three changes of 10 mM sodium cacodylate buffer, pH 6.8, adjusted to 0.05 M Na^+ with NaCl. The final dialysate was reserved, filtered at 0.45 µm, and used in all subsequent dilution procedures and to dissolve DB1976. Compound concentrations were calculated based on analytically weighed solid to a precision of 10^{-5} g. DNA concentrations were determined by UV absorption at 260 nm using the nearest-neighbor extinction coefficients (113), in cm^{-1} $(M\ duplex)^{-1}$: 186,075 for 5'-CGCAAATTTGCG-3' (termed "A_3T_3"), 191,511 for 5'-CGCGAATTCGCG-3' ("A_2T_2"), and 190,127 for 5'-CGCAACGTTGCG-3' ("A_2CGT_2").

2.5.2 Volumetric measurements

Solution densities ρ were measured in 10 mM sodium cacodylate, pH 6.8 at 25°C, containing 50 mM NaCl using an Anton Paar Model DMA-5000 vibrating tube densimeter with a precision of 1.5×10^{-6} g/cm^3. At concentrations of DNA and compound used in this study (up to 10^{-5} M), the apparent density is strictly linear in molar concentration, indicating equivalence with measurement at infinite dilution *i.e.,* partial molar values [**Figure 2.3**]. The volume change accompanying DNA binding by DB1976 was determined by a reported titrimetric method (48), in

which the densities of matched concentrations of DNA were measured upon addition to DB1976 or buffer. The volume change per mol of DNA is given by (48):

$$\Delta V_{obs} = \frac{(\rho - \rho_0) - (\rho' - \rho'_0)\left(1 + \frac{V'}{V'_0}\right)}{\rho_0 c} \quad (12)$$

In Eq. (12), ρ is the density of the DNA in buffer, without DB1976. ρ_0 is the density of plain buffer. ρ' is the density of the DNA in buffer containing DB1976. ρ'_0 is the density of a matched solution of DB1976 alone. V' is the cumulative volume of the DB1976 stock solution added to the sample. V'_0 is the starting volume and *c* is the initial DNA concentration in the experiment. Eq. (12) includes a dilution correction.

2.5.3 Fluorescence polarization titrations

Equilibrium titrations of DB1976 by DNA was performed as previously described (79, 114) in 10 mM sodium cacodylate buffer, pH 6.8, containing various concentrations of NaCl. Dissociation constants were estimated by fitting a 1:1 binding model to the data with total DNA oligomers taken as independent variable (24). Molar salt concentrations were converted to mean ionic activities using published activity coefficients in water (100).

2.5.4 Setup of molecular dynamics simulations

Explicit-solvent simulations of DB1976, DNA, and the DB1976/DNA complex were performed using the Amber14sb/parmbsc1 forcefield (82) in version 2019.3 of the GROMACS environment. DB1976 was parameterized as previously described (79). In brief, geometry optimization and single point energy calculations for DB1976 were performed using the Hartree-

Fock/6-31G* basis set in GaussView/Gaussian 09. Partial charges were then derived following the RESP method in Antechamber from AmberTools16. The GAFF2 forcefield was used to assign the atom types. For the selenophene Se atom, which was not contained in GAFF2, we temporarily assigned the S atom type, since the chemical nature of selenophene closely resembles its congener ring thiophene (115-117). Mass, geometry (115), and vDW parameters (118) were hand-edited to revert S back to Se. This approach have been used successfully in other studies of seleno compounds using the Amber forcefields (e.g., 119). Duplex DNA encoding the experimental sequences were generated in canonical B-form using 3DNA (120). Complexes with DB1976 were generated by homology modeling of a published DB1976/DNA complex originally docked using AutoDock Vina (79). Independently, we also performed *de novo* docking with the three DNA sequences in this study using the same parameters. The internal portion of the duplex was defined as the search box for 100 iterations. The top-ranked poses correspond in position with each other and the homology models [**Figure 2.5**]. In all cases, the poses are analogous to co-crystallographic complexes of other iso-helical benzimidazoles such as DAPI (PDB: 1D30, 432D) and Hoechst 33258 (PDB: 8BNA, 296D, 1DNH) bound to similar DNA sequences.

Following topology generation, each system was set up in dodecahedral boxes at least 1.0 nm wider than the longest dimension of the solute, solvated with TIP4P-ew water (83), and neutralized with Na^+ and Cl^- to 0.05 M. Electrostatic interactions were handled by particle-mesh Ewald summation with a 1 nm distance cutoff. All simulations were carried out at an *in silico* temperature and pressure of 298 K (modified Berendsen thermostat) (121) and 1 bar (Parrinello-Rahman ensemble). A timestep of 2 fs was used and H-bonds were constrained using LINCS. After the structures were energy-minimized by steepest descent, the *NVT* ensemble was equilibrated at 298 K for 1 ns to thermalize the system, followed by another 1 ns of equilibration

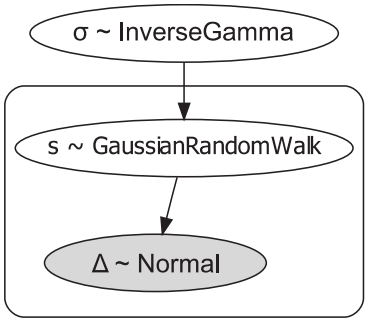
of the *NPT* ensemble at 1 bar and 298 K. The final *NPT* ensemble was simulated without restraints for 600 ns, recording coordinates every 2 ps. Convergence of the trajectories were checked by RMSD from the energy-minimized structures, after adjustments for periodic boundary effects.

2.5.5 Analysis of MD trajectories

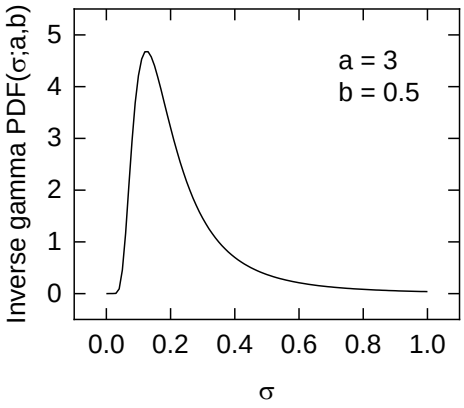
Unless stated otherwise, post-processing was performed with tools provided in GROMACS. DNA helical parameters were extracted using do_x3DNA (120, 122). Solvent-excluded volumes and solvent-accessible surface areas were computed using ^{3}V (123) and VMD, respectively. Ensemble average volumes and surface areas were reported as means ± standard deviations of 100 ns blocks.

2.5.6 Bayesian inference of stochastic volatility

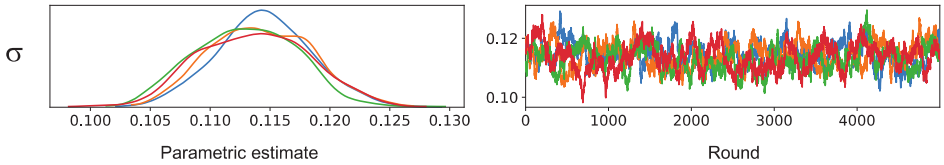
For each system, the trajectory of summed curvature over the 9 interior base steps was differenced frame-by-frame to generate a stationary time series $\Delta(t)$. To model stochastic volatility, we treated $\Delta(t)$ as a random walk whose step size (innovation) $s > 0$ is a random variable drawn from a normal distribution standard deviation $\sigma > 0$. Schematically, the model specification is:



To initialize the Bayesian analysis, a prior distribution for the target volatility-specifying parameter σ is empirically assigned as an inverse gamma distribution to capture its anticipated estimates between 0.1 to 0.2 (*c.f.* Figure 2.10F).



The high dimensionality and a dependency-structure in the random walk distribution renders analytical approaches to the parameters intractable in practice. We therefore utilize a numerical approach by Markov Chain Monte Carlo (MCMC) simulations, using the Hamiltonian NUTS sampler as implemented in the PyMC3 package (version 3.8) (124). Four independent chains of simulations were sampled, each running 5,000 iterations after discarding the initial 500 as burn-in. Other parameters were internally selected by the software. Representative trace plots of the four chains sampling σ for A_2T_2 are shown as follows:



The left panel is a histogram (kernel density plot) for σ. In the right panel, the four chains show convergent mixing during the 5,000 MCMC rounds, reflecting sampling of the underlying

MD-derived distribution. Chain mixing is formally checked with the Gelman-Rubin diagnostic ($\hat{R}$), which is routinely near its theoretical value of unity for complete mixing. To assess the goodness of fit of the model, posterior predictive checks (PPC) were performed to compare the medians and median absolute deviations (MAD) of the traces against those of the MD trajectories. Illustrative results for A_2T_2 are presented in Figure 2.10E in the main text, showing that the model captures the statistics of the MD results. Bayesian estimates of σ for each construct are reported in Figure 2.10F and Table 3 as the mean and 95% highest posterior density (HPD, also known as credible interval) of the posterior distribution. HPD is the Bayesian analog of the 95% confidence interval.

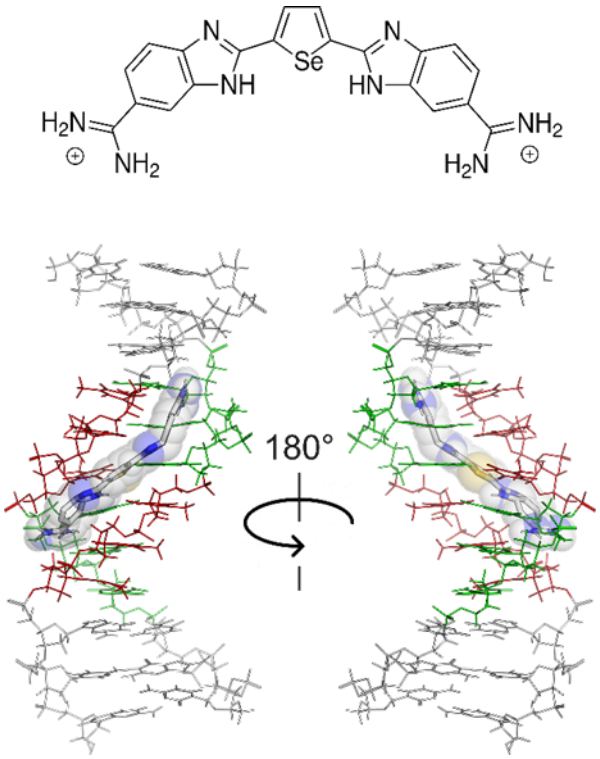


Figure 2-1 Structure of DB1976

Top: the diamidinium cation, the expected ionization state at neutral pH. Bottom: docked model of DB1976 in complex with 5'-d(CGCA3T3GCG)-3' in which dA is colored in green and dT in red.

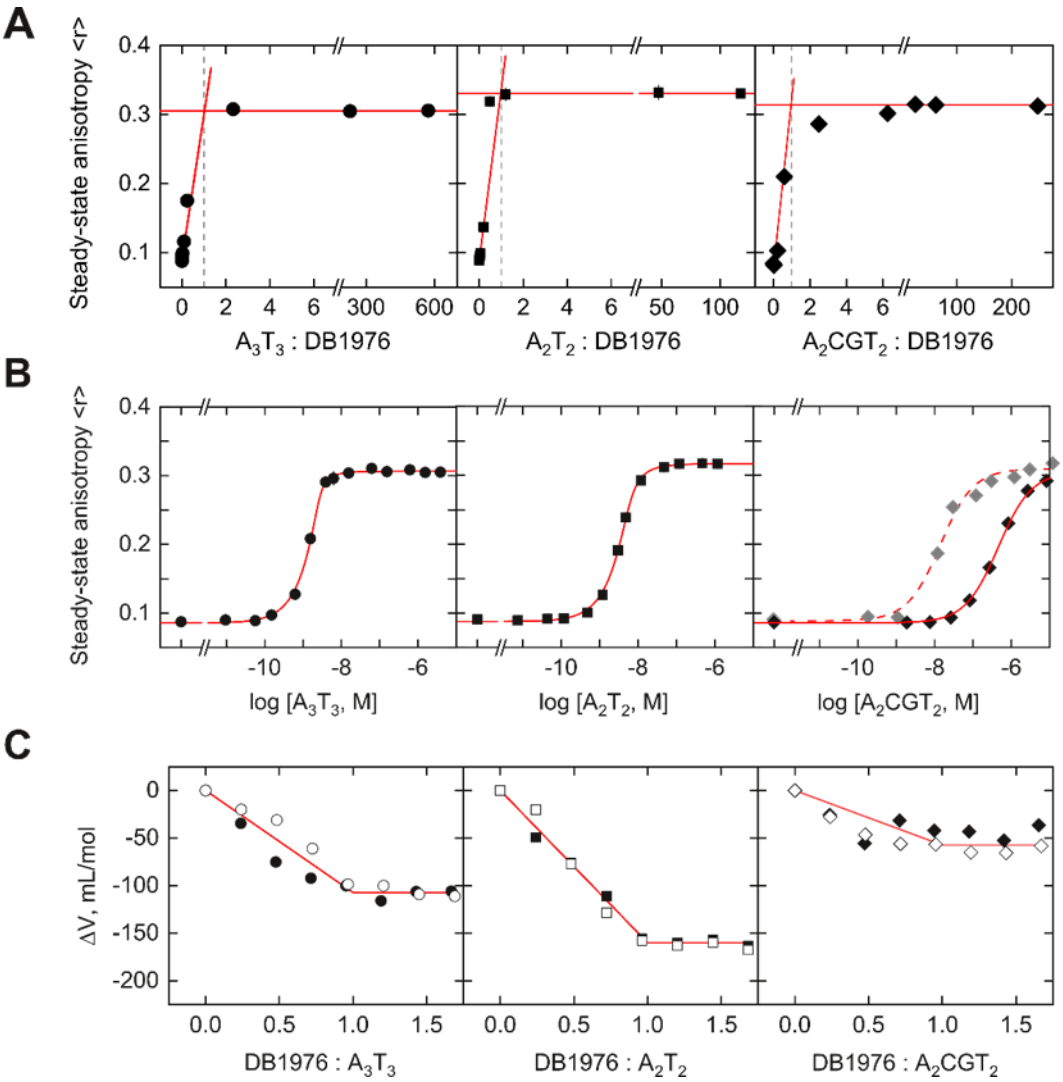


Figure 2-2 Fluorescence polarization and densimetric titrations of DB1976 to three AT sequence contexts

The solution conditions were 10 mM sodium cacodylate, pH 6.8, plus NaCl at various concentrations as described. Error bars representing the precision of the anisotropy measurements are on the order in size as the symbols. **A**, Titrations of 0.4 µM DB1976 with A_3T_3, A_2T_2, and A_2CGT_2 DNA for stoichiometric determination in 50 mM NaCl. Dashed lines represent unit DNA:DB1976 ratio. **B**, Titrations of 4 nM DB1976 with the same DNA, but in 0.5 M NaCl, the lowest salt concentration at which binding affinities to all three sequences could be quantified. The gray points for A_2CGT_2 represent data acquired in 5 mM NaCl to establish the lack of nonspecific aggregation at very low ionic strength. Curves represent fits of 1:1 binding with total DNA concentration taken as independent variable. **C**, Density-detected volumetric titrations of DNA with DB1976. Two independent experiments are shown. Lines represent least-square fits of the data to Eq. (12). Parametric values are given in Table 1.

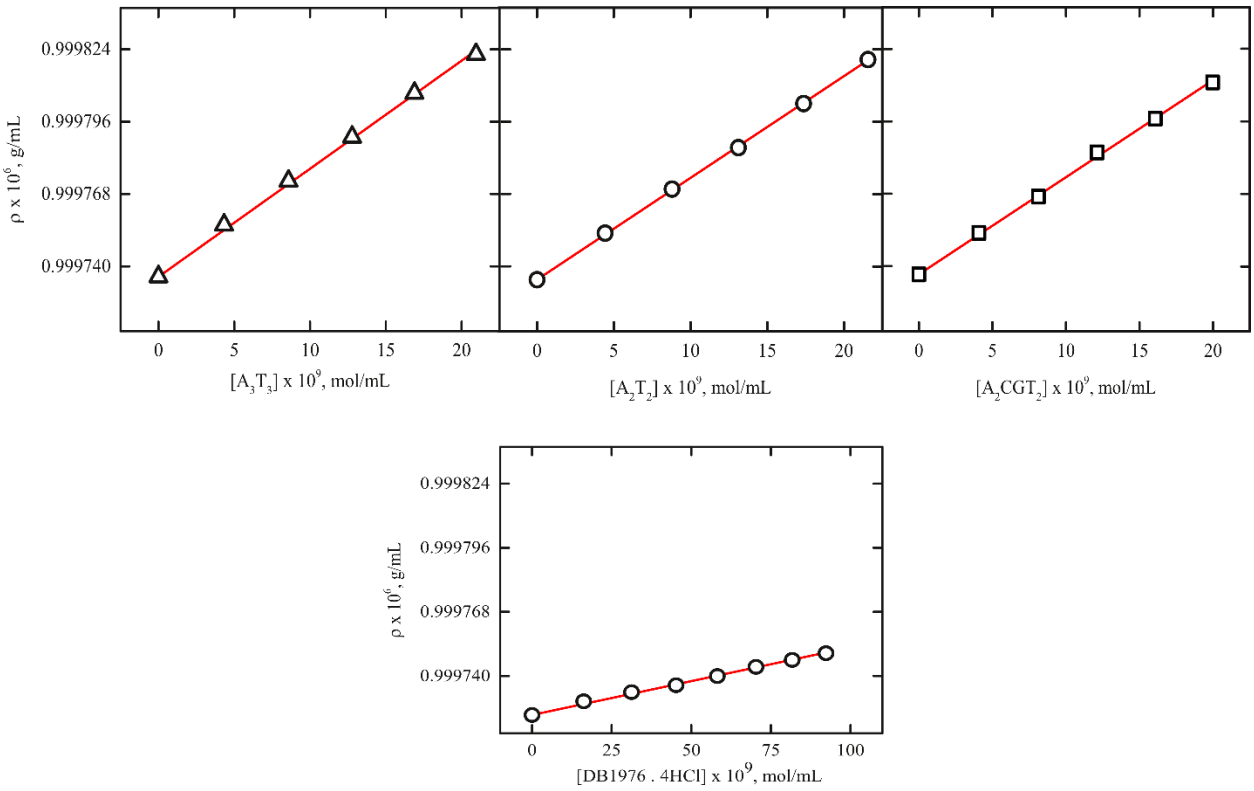


Figure 2-3. Partial molar volumes of unbound DNA and DB1976.
The solution conditions were 10 mM sodium cacodylate, pH 6.8, with 50 mM NaCl. The apparent molar volumes *V*° of the compound and free DNA were determined from the following relationship:

$$\rho = \rho_0 + (M - V^\circ \rho_0)c$$

where ρ_0 is the density of the buffer, *c* is the molar solute concentration, and *M* is the molecular weight of the solute (125). *Lines* represent linear fits to the data. For the DNA, the partial molar volumes of A_3T_3, A_2T_2, and A_2CGT_2 are 3,113 ± 40 cm^3/mol, 3,365 ± 14 cm^3/mol, 3,551 ± 30 cm^3/mol, respectively. The partial molar volume for DB1976 (297 ± 4 cm^3/mol) as measured includes the neutralization of 2 equivalents of H^+ by the buffer and hydration of 4 equivalents of Cl^- to generate the DB1976 diamidinium cation, the "active" species in DNA binding (231 ± 4 cm^3/mol). For this calculation, ΔV^0_{ion} = 13 mL/mol for sodium cacodylate (44), and V^0_{Cl} = 23 mL/mol (80).

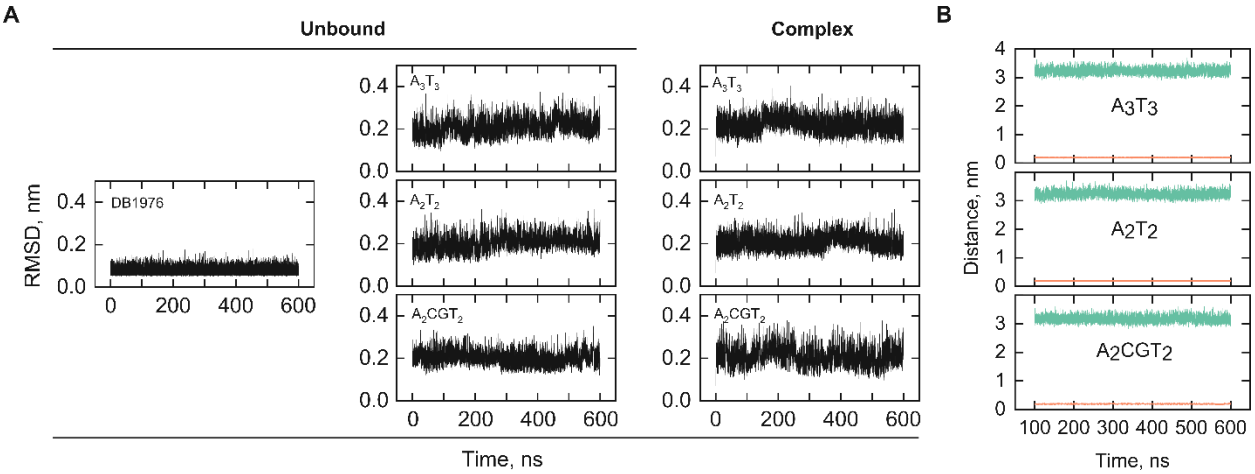


Figure 2-4. Global behavior of the of DB1976, unbound DNA and their complexes during the simulations.

A, All-atom root-mean-square deviations (RMSD). The reference coordinates were the energy-minimized structures for each species. **B**, Separation between DB1976 and DNA in the complexes. The maximum (green) and minimum (orange) distance between the compound and DNA are plotted. The initial 100 ns is discarded as burn-in.

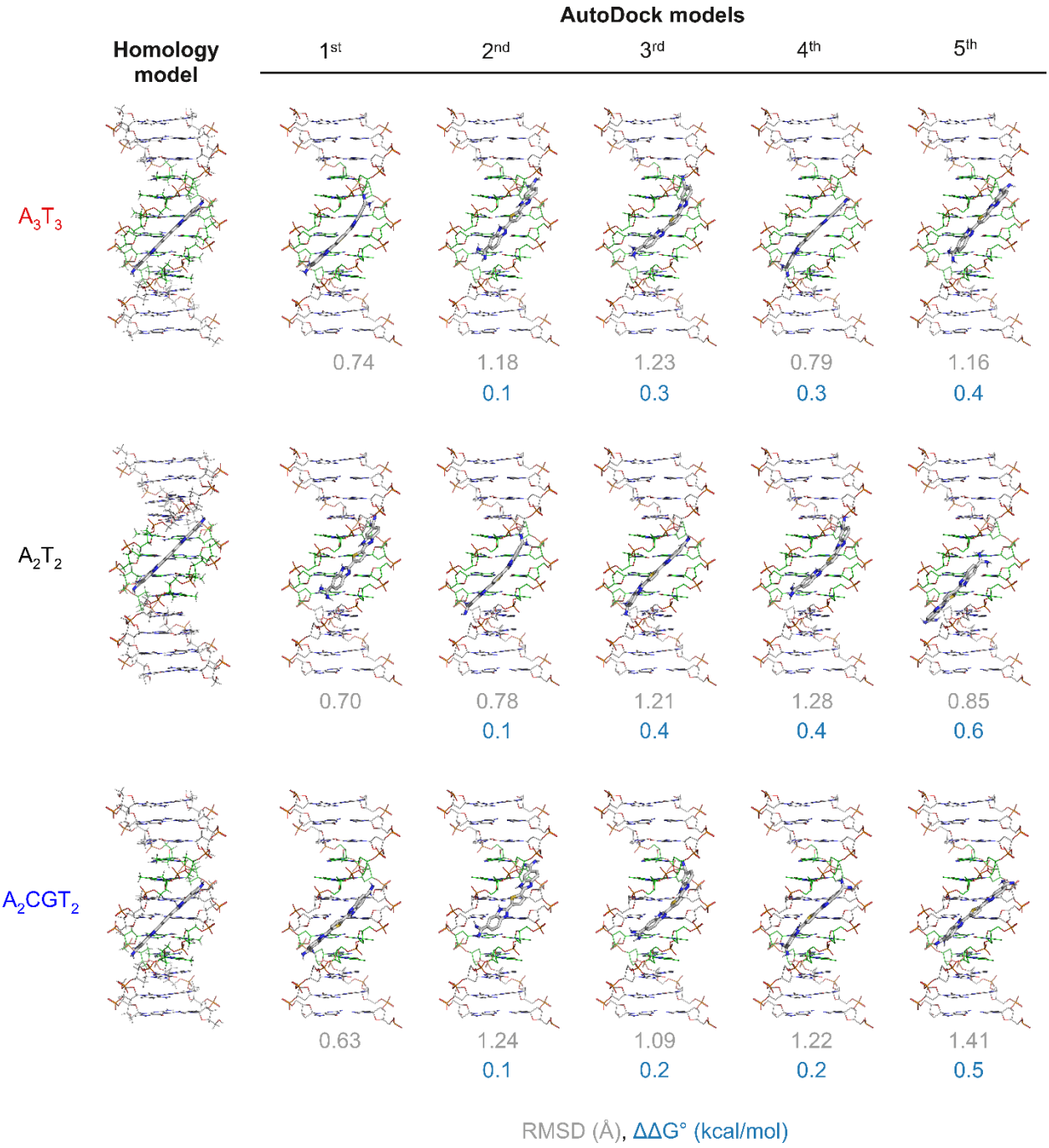


Figure 2-5. De novo docked DB1976/DNA complexes.

Docking to B-DNA encoding the three dodecameric sequences in this study was performed with Autodock Vina as described in Materials and Methods. The internal portion of the helix was used as the search box. dA and dT bases are colored in green. The five top-ranked poses are shown with the homology model used in the simulations. RMS deviations are relative to the homology model and the nominal ΔΔG° (as computed by AutoDock) is relative to to the first-ranked pose. Note the offset positions of the ligand in the homology model and first-ranked pose.

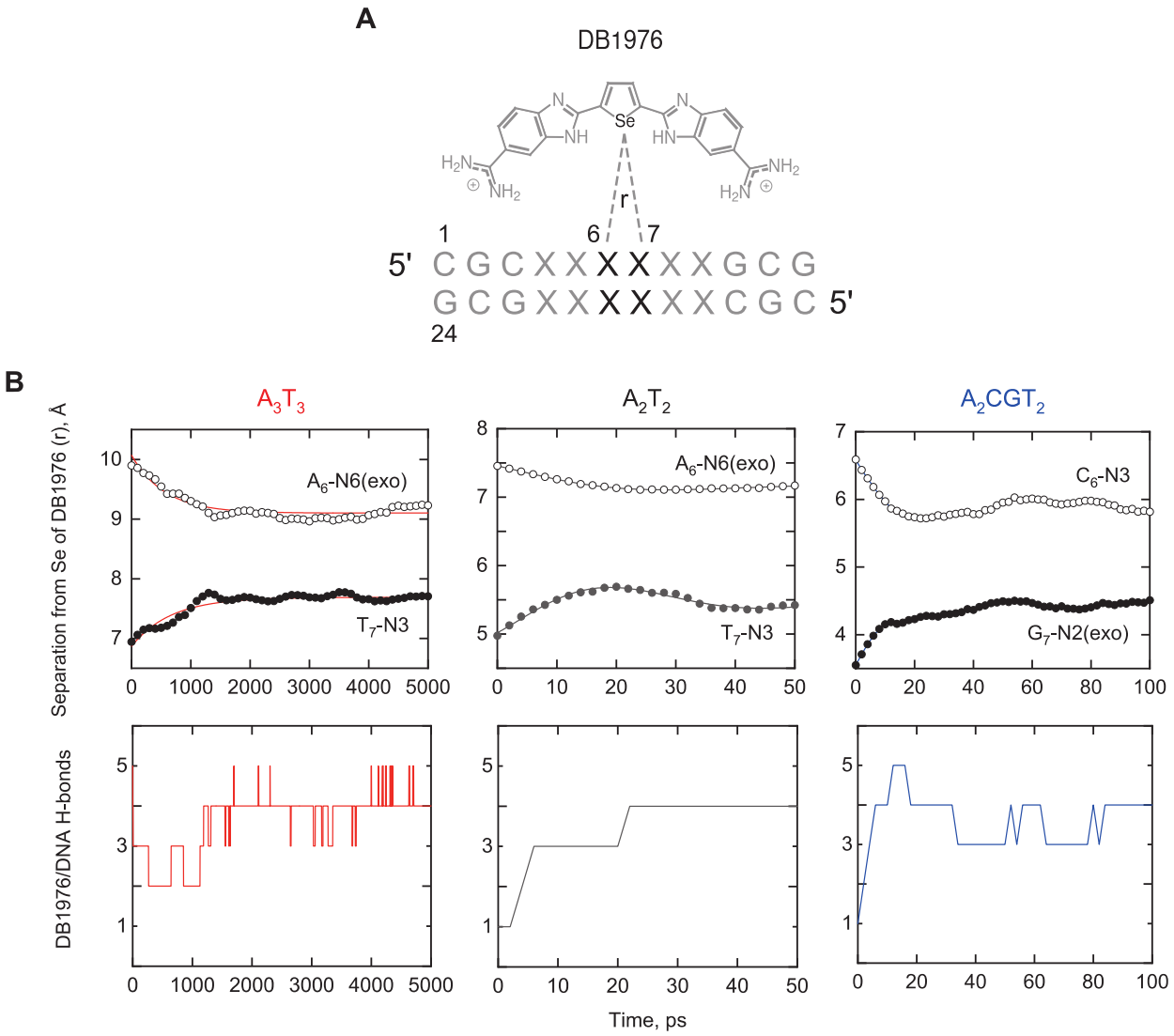


Figure 2-6. Relaxation dynamics of non-optimally docked DB1976. A. To probe the preference for DB1976 for an offset position in the complexes, the ligand was initially positioned at the geometric center of each oligomer. Following solvent equilibration, the separation between the Se atom in DB1976 and characteristic atoms in the central base step (A_6 and T_7 for A_2T_2 and A_3T_3, C_6 and G_7 for A_2CGT_2) were followed during unconstrained simulations. These atoms are the exocyclic -NH_2 groups in the purines R = A and G, and N3 of the pyrimidines Y = T and C. **B**, Trajectories of the separation distances show translation of the Se atom of DB1976 towards one position and away from the other. Also shown are the H-bond trajectories between DB1976 and DNA. These trajectories reveal different relaxation kinetics, which are ~10^2 slower for A_3T_3 than A_2T_2 and A_2CGT_2. The points were adjacent-averaged with 10-point moving window to clarify trends. Every 20th point in the A_3T_3 trajectory is shown. At short timescales, damped harmonic motion is also apparent.

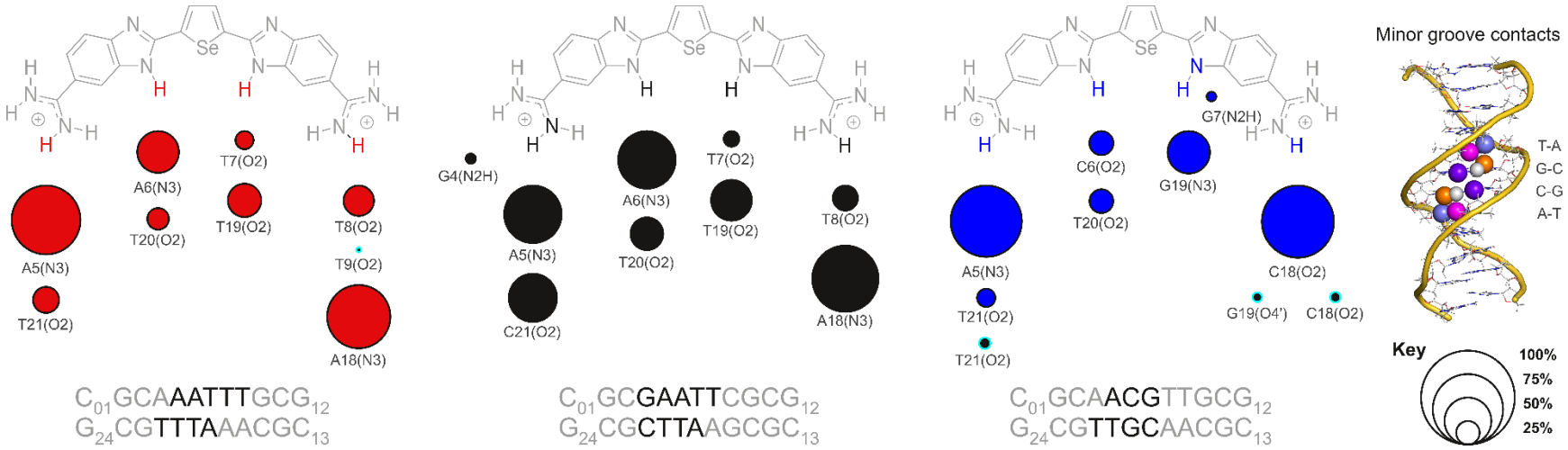


Figure 2-7. Drug-DNA contacts in simulational complexes with A_3T_3, A_2T_2, and A_2CGT_2.
The DB1976 dication is schematically rendered in gray with H-bonding atoms with DNA colored in red (A_3T_3), black (A_2T_2), and blue (A_2CGT_2). DNA contacts are bolded. The size of each blob corresponds to the fractional occupancy of the indicated H-bond in the simulations (see key). Water-mediated contacts are marked with cyan outlines. Some contacts e.g., T21(O2) in A_2CGT_2 are alternately made directly or via a bridging water. Contacts made at lower than 10% occupancy are not shown. The minor groove contacts are rendered as spheres: A(N3), T(O2), C(O2), G(N3), and G(N2H).

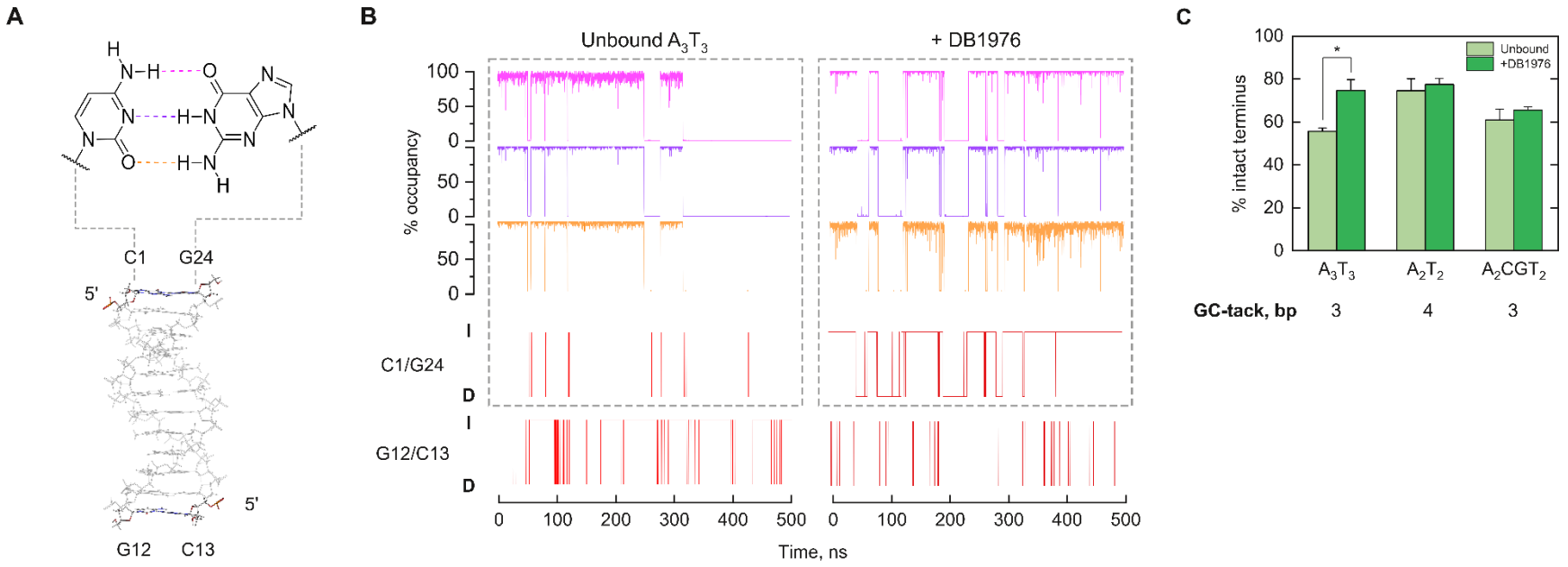


Figure 2-8. H-bond analysis of end deformation.

A, Terminal GC base pairs in the DNA oligomers showing the three Watson-Crick H-bonds. **B**, Representative trajectories of the two termini as intact (I) or deformed (D) base pairs in unbound and complexed the A_3T_3 oligomer. The status as I or D is determined based on the occupancy of the three terminal H-bonds over a 10-ps moving window. **C**, Intact termini are defined as the retention of at least one Watson-Crick H-bond. For the free duplexes, A_2T_2 has more stable ends than A_3T_3 and A_2CGT_2, consistent with the former's longer GC-tacked termini (4 bp) than the latter two duplexes (3 bp). H-bond occupancy was computed using the python script readHBmap by Ricardo Soares (University of São Paulo, Brazil), obtained at http://www.gromacs.org/Downloads/User_contributions/Other_software.

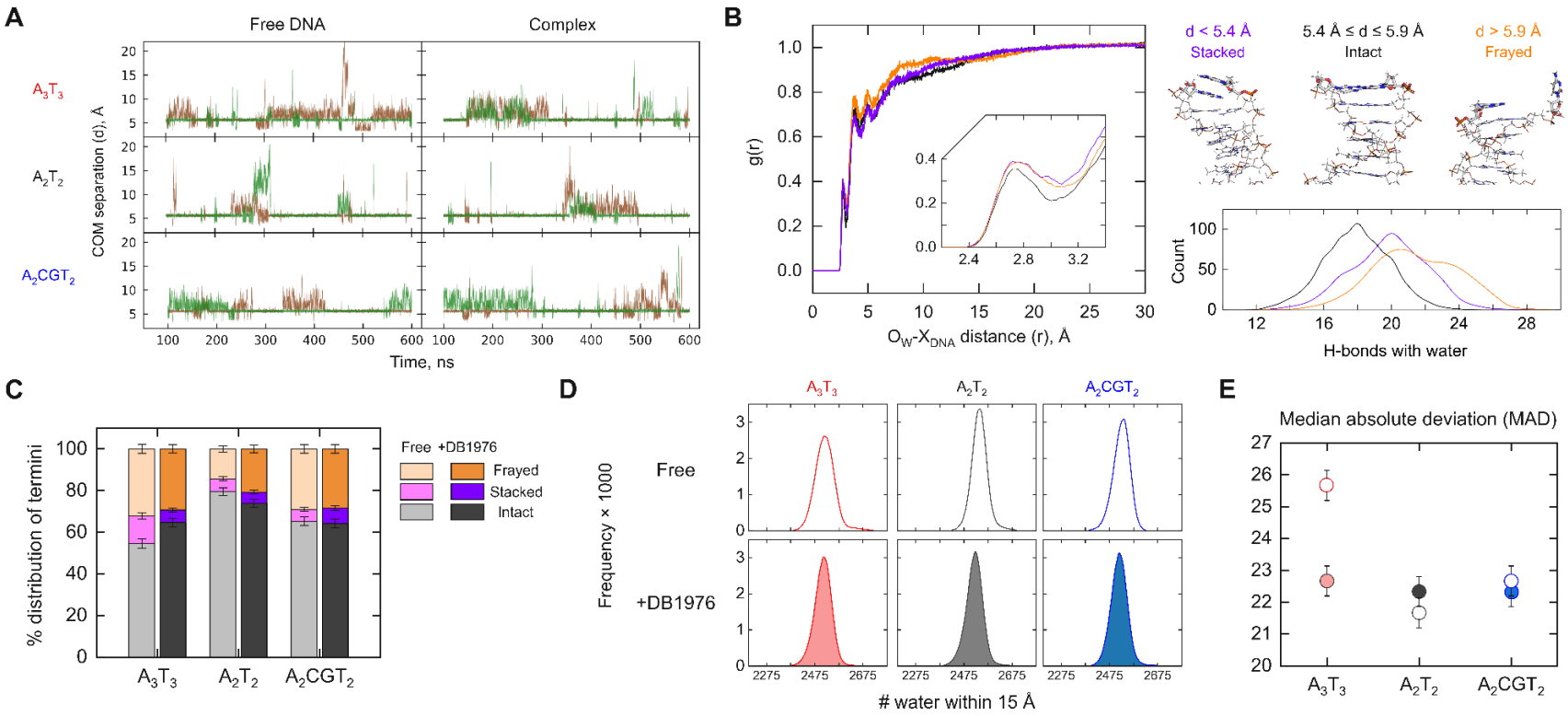


Figure 2-9 Analysis of end deformation and hydration in simulational DNA and DB1976-bound complexes

A, Trajectories of the distances separating the centers of mass (COM) of the two pairs of terminal GC bases over the final 500 ns. Intact ends are bounded tightly between 5.4 and 5.9 Å. **B**, Radial distribution functions $g(r)$ for non-H terminal DNA atoms and water oxygen over representative 1-ns segments associated with states defined by distance criteria. Distances longer than 5.9 Å showed progressive end fraying of terminal base pairs, while distance shorter than 5.4 Å showed non-canonical stacking by the buckled base pair. Intermediate separation corresponds to intact base pairs. Also shown are histograms of H-bonds of the termini with water. **C**, Fraction of the terminal base pairs as intact or one of the two deformed states in the free and bound DNA over the 500 ns trajectories. **D**, Representative histograms of water within 15 Å of the terminal and penultimate base pairs of free and DB1076-bound DNA. In each case, the total frequency over 500 ns was 2.5×10^5 frames. **E**, Median absolute deviations (MAD) of histograms from three replicas, following the labeling scheme of Panel D.

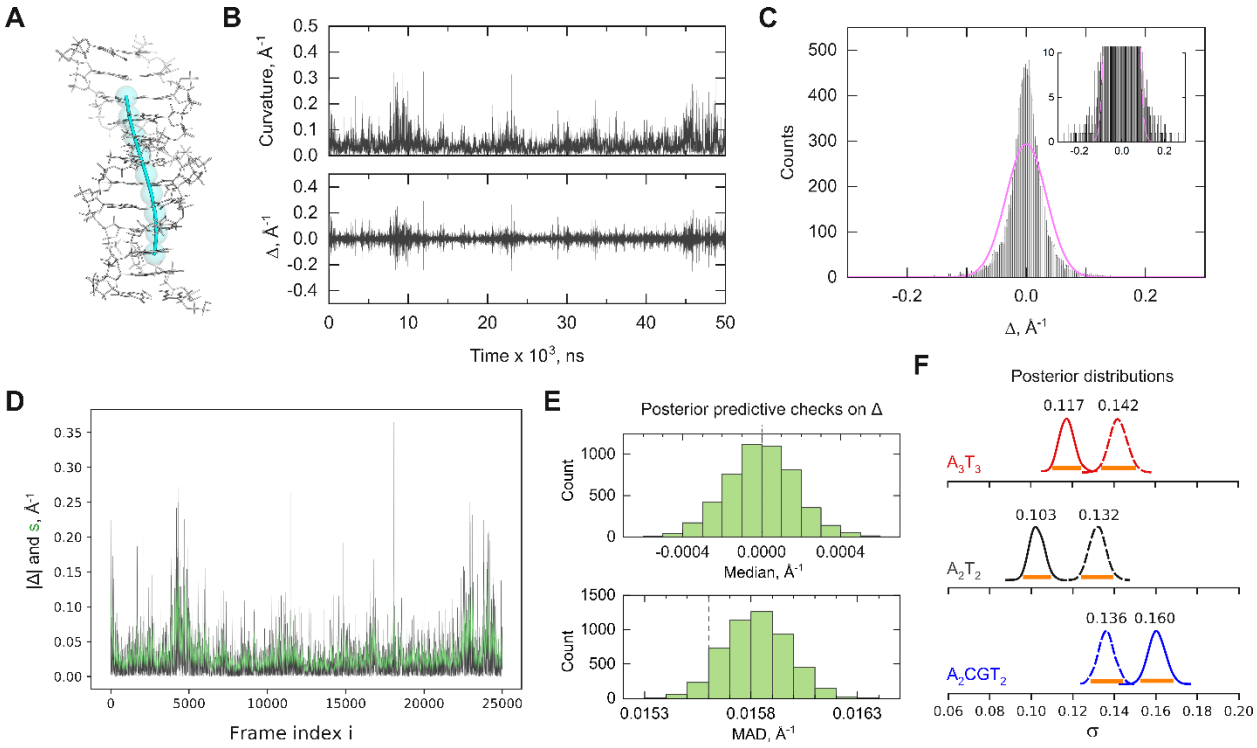


Figure 2-10 Stochastic volatility modeling reveals sequence-dependent dynamic perturbation by DB1976

A, Curvature at each internal base step was computed as the absolute geometric curvature of a globally fitted helical axis. The values over the 9 interior base steps were summed. **B**, As an illustrative example, the curvature trajectory of free A_2T_2 and its differenced stationary time series (termed Δ). A 50-ns window (2.5×10^4 frames) is shown to visualize the details. **C**, Histogram of Δ showing leptokurtosis relative to a Gaussian distribution (magenta). **D**, Overlay of 50 (green, out of 5,000 total) traces from the MCMC simulations of the stochastic volatility model for the same window as Panel B. **E**, Posterior predictive checks (PPC) aimed at probing how closely the posterior distributions reproduce the statistical properties of the MD-derived distribution. Shown are histograms of the medians and median absolute deviations (MAD) of the 5,000 traces for A_2T_2. Dashes represent the corresponding values for the MD-derived distribution as shown in Panel C and Figure 2.11. **F**, Summary of the posterior distributions of σ, the target volatility-specifying parameter, for the three sequences (dashes) and their DB1976-bound complexes (solid). The Bayesian estimates of σ are given as means with uncertainty given by the 95% highest posterior density (HPD, or credible interval; orange bars).

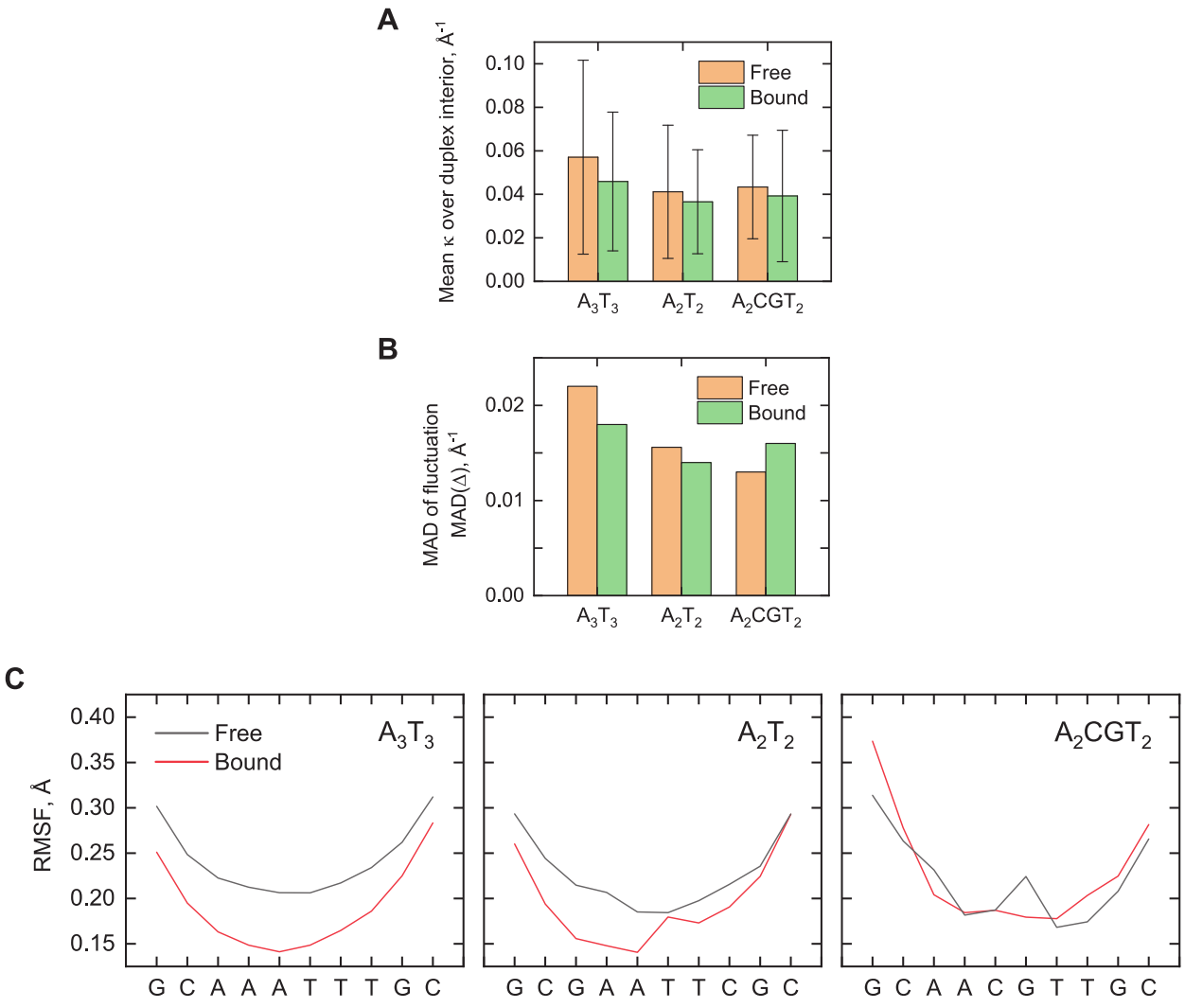


Figure 2-11. Summary statistics of the curvature of the interior helix in free and bound DNA.

A, Curvature κ values as means ± S.D. **B**, Mean absolute deviations (MAD) in the frame-by-frame differences in κ, denoted as Δ. $\mathrm{MAD}(\mathbf{x}) \equiv \mathrm{median}\left(\left|x_i - \mathrm{median}(\mathbf{x})\right|\right)$. The median values of Δ are uniformly 0. **C**, RMS fluctuations (RMSF) of the all-atomic positions, expressed per interior base pair.

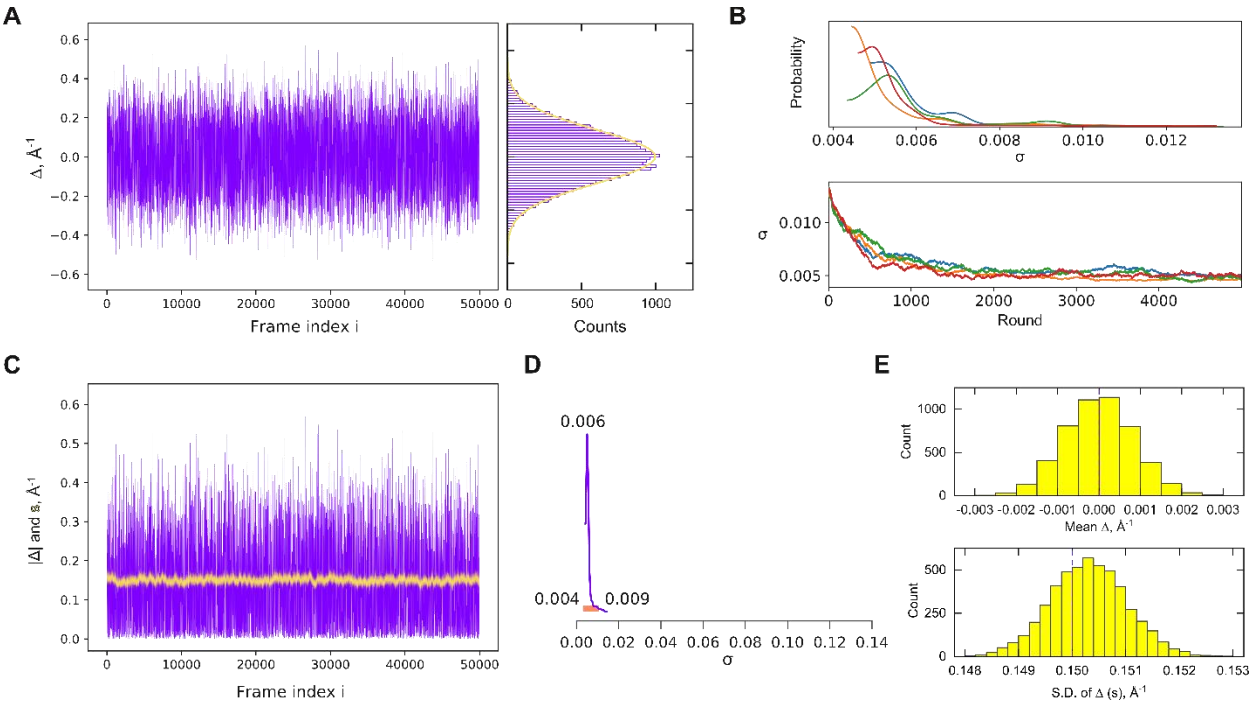


Figure 2-12 Control stochastic volatility (SV) analysis.

We tested the specificity the SV model by subjecting it to a simulated random walk in Δ with zero mean and a constant step size of s = 0.15 i.e., no volatility. In the context of the model, this corresponds to σ = 0. This is a challenging use case for the model because standard deviations take on strictly positive values (σ > 0). **A**, Under these conditions, a histogram of Δ is Gaussian-distributed (yellow) with standard deviation s. **B**, 5,-000 rounds of MCMC sampling with four chains. Distributions of σ (top) and its evolution during the simulation is shown. **C**, Overlay of 50 of the 5,000 sampling traces of s(t) is shown (yellow), hovering tightly around the simulated constant value of 0.15. Contrast this behavior with the time-varying s(t) inferred from the MD-derived curvature fluctuations in Figure 2.10D. **D**, Bayesian estimate of σ as given by the mean bracketed by the 95% highest posterior density (HPD) drawn from the posterior distribution. The distribution is presented with the same scale spacing as Figure 2.10F to emphasize the uncertainty estimate afforded by the model. **E**, Posterior predictive check on Δ in terms of the distribution of means and standard deviations, as Δ is normally distributed in this case. Purple dashed lines represent the challenge values for the mean (zero) and standard deviation (s = 0.15).

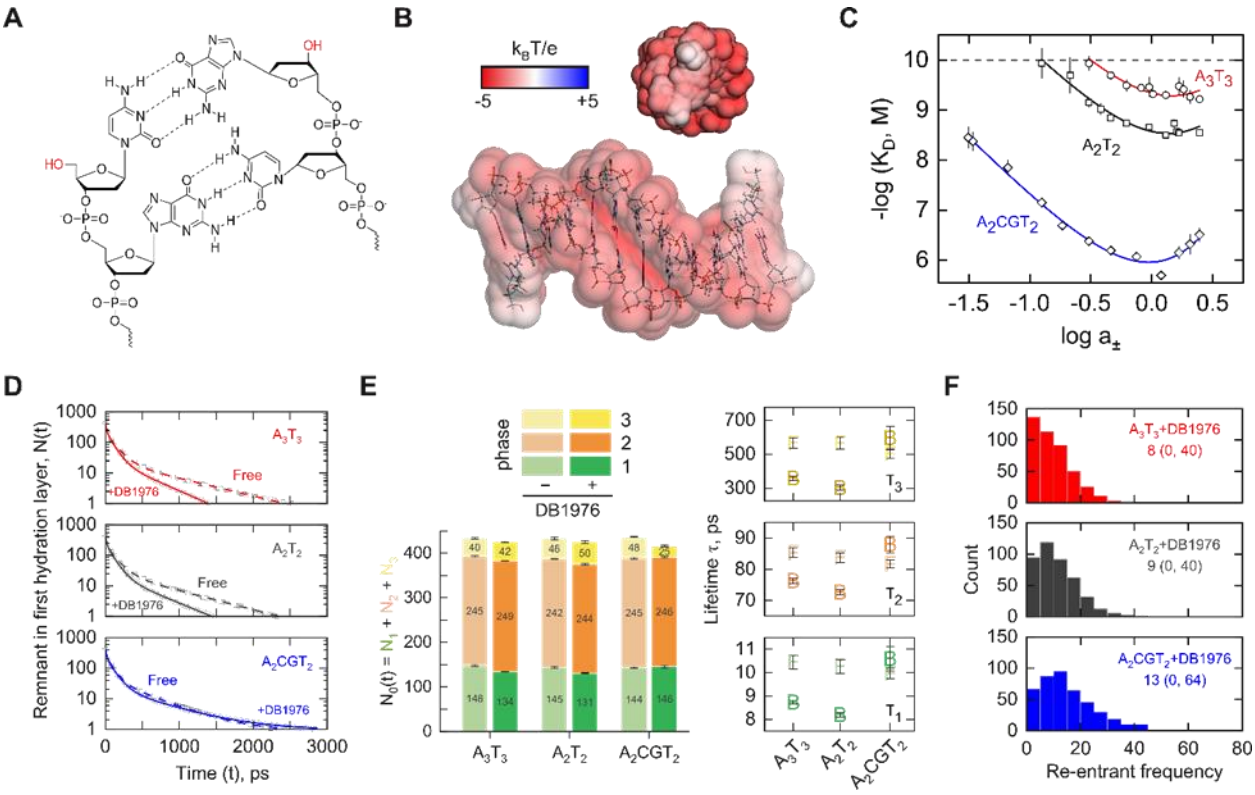
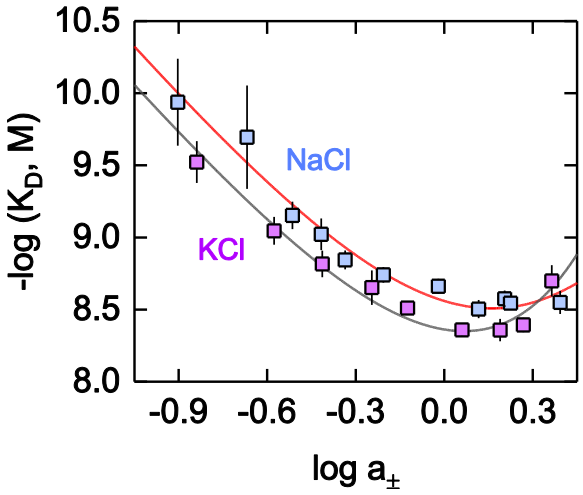


Figure 2-13 Domain-specific hydration properties of DB1976/DNA complexes

A, Structure of the termini of solid-phase synthetic oligonucleotides, which are uncharged hydroxyls (red). B, Surface electrostatic potential of the average NMR structure of A2T2 (PDB: 2DAU) in implicit water containing 0.15 M NaCl using APBS. Note the near-zero potential at the termini. C, Ionic dependence of equilibrium DB1976 binding to A3T3, A2T2, and A2CGT2 was determined with NaCl at the indicated mean ionic activities. Titration profiles at 0.5 M NaCl (~ 0.5) are shown in Figure 2.2B. Curves represent global fits of the data by the linkage relation, Eq. (9). Parametric values are given in Table 6. Dashed line represents the upper limit of quantitation for the equilibrium constant given the concentration of DB1976 (5 nM) used in the titrations. Unshaded symbols indicate a subset of the data reported preliminarily in a thesis (N.E., Department of Chemistry, Georgia State University, 2019). D, Representative phenomenological decay curves of the first hydration layer of free and DB1976-bound DNA, defined as within 5 Å between the water oxygen atom and a non-H atom on the solute. Every 20th point is shown. Curves represent fits to Eq. (11). The initial 5 ps of decay is shown in Figure 2.15. E, Fitted parameters of Eq. (11) to the decay curves, expressed as mean ± S.E. of five non-overlapping instantiations of the hydration layer. Free (F) and bound (B) complexes are presented in light and dark shades as indicated. F, Representative histograms of re-entry into the first hydration layer of the DB1976-bound complexes. Labeled values represent median (minimum, maximum) re-entrant frequencies. See Figure 2.16 for trajectories of representative water molecules.



	$\Delta n_{\pm}$	Δn_w	$\log K_0$
NaCl	-2.4 ± 0.2	-35 ± 2	-7.73 ± 0.04
KCl	-2.4 ± 0.2	-33 ± 2	-7.46 ± 0.04

Figure 2-14 Ion-water linkage in DB1976/A_2T_2 binding in NaCl and KCl

To probe the potential release of Na^+ from a non-condensation régime and associated perturbation of DNA structure across the salt range employed, we compared the effects of substituting NaCl with KCl. In counter-ion condensation theory, released cations arise from the diffuse ionic atmosphere surrounding the polyanion (DNA) in a manner that depends on the geometry of the charges on the DNA but independently of cation identity (126). If locally held cations or significant changes in DNA conformation were involved, the two different cations are expected to produce distinct perturbations on DB1976 binding. Titrations data was globally fitted with the linkage relation, Eq. (9), and the attendant conversion between $m_{\pm}$ and $a_{\pm}$, Eq. (10), from the main text. The salt-specific coefficients φ that enter Eq. (10) for NaCl are $\varphi_1 = 1.64$ and $\varphi_2 = -0.113$ as indicated in the text. For KCl, $\varphi_1 = 1.74$ and $\varphi_2 = -0.0164$ (100). A switch was written into the fitting routine to assign the appropriate coefficients to each data set. Note that similarity in the ion and hydration numbers does not translate to parallel curves between different salts due to the identity-specific relationship between $m_{\pm}$ and $a_{\pm}$. Binding in the presence of KCl was weaker than corresponding activities of NaCl, in agreement with the affinity of condensed cations for DNA (127), with identical $\Delta n_{\pm}$ and Δn_w within experimental error. The data therefore confirms that, across the ionic activities tested, the disposition of cations is accounted for as the release from the diffuse ionic atmosphere around poly-anionic DNA.

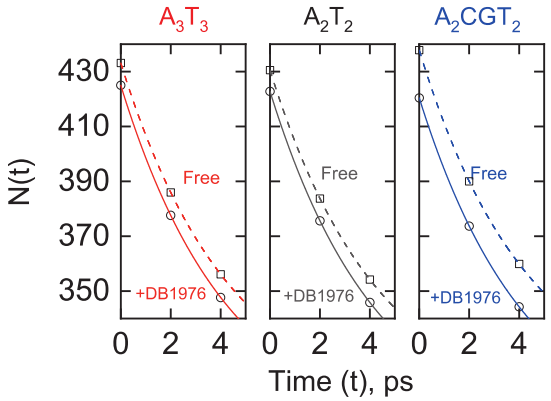


Figure 2-15Initial decay profiles of free DNA and DB1976-bound complexes
Initial 5 ps of the decay profiles shown in Figure 2.13D in the main text. *Curves* represent fits to Eq. (11).

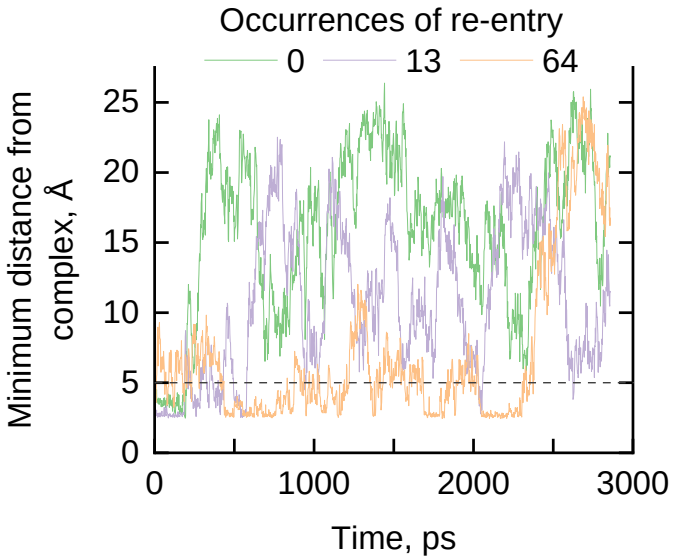


Figure 2-16 Illustrative re-entrant trajectories of first-layer hydration water in the A_2CGT_2/DB1976 complex

Three water molecules were selected from the ensemble hydrating the A_2CGT_2 complex shown in Figure 2.13F. They represented the minimum (zero), median, and maximum occurrence of re-entry into the first hydration layer, defined as 5 Å (dashed line) between the water oxygen atom and a non-H atom of the solute. A re-entry (RE) is defined as an instance in which the water was farther than 5 Å from the solute and subsequently returned to within 5 Å during the decay trajectory. RE is computed as multiples of even numbers of 5 Å-crossings ($X_{5Å} \in \mathbb{N}$) as follows:

$$\mathrm{RE} = \frac{X_{5Å} - (X_{5Å} \bmod 2)}{2} .$$

Note that, by definition, the trajectory begins at a separation less than 5 Å.

Table 1 Experimental parameters of DNA recognition by DB1976

The solution conditions were 10 mM sodium cacodylate, pH 6.8 at 25°C. Binding and free energy data corresponded to the presence of 0.5 M NaCl while the volumetric data was acquired in 50 mM NaCl.

	A_3T_3	A_2T_2	A_2CGT_2
$K_D \times 10^{-9}$ M	0.12 ± 0.04	0.70 ± 0.15	452 ± 25
$\Delta G°$, kJ/mol at 25°C	-56.60 ± 0.83	-52.23 ± 0.53	-36.20 ± 0.14
$V^0_{unbound}$, cm^3/mol	3,113 ± 40	3,365 ± 14	3,551 ± 30
ΔV_{obs}, cm^3/mol	-107 ± 2	-162 ± 2	-57 ± 2
$V^0_{complex}$, cm^3/mol	3,303 ± 40	3,499 ± 15	3,790 ± 30

Table 2 H-bonding between DB1976 and DNA in the simulational complex with A_2T_2
Occupancies were computed with standard 3.5 Å and ±30° cutoffs over the final 500 ns of three independent simulations. Minor contacts with lower than 10% average occupancy were omitted. Coefficients of variation (CV = S.D./average) for contacts in replicate simulations with other sequences span similar ranges.

	H-bond pair		Occupancy				
	Donor	Acceptor				Average	CV, %
1	DB1976(H34)	A18(N3)	70.6	72.6	71.9	71.7	1.4
2	DB1976(H34)	T8(O2)	26.7	25.6	26.1	26.1	2.1
3	DB1976(H25)	T19(O2)	43.4	43.3	43.4	43.4	0.1
4	DB1976(H25)	T7(O2)	16.6	16.4	16.3	16.4	0.9
5	DB1976(H13)	C21(O2)	51.4	51.2	50.7	51.1	0.7
6	DB1976(H13)	A5(N3)	61.3	58.6	61.7	60.5	2.8
7	DB1976(H4)	T20(O2)	34.8	34.9	34.4	34.7	0.8
8	DB1976(H4)	A6(N3)	28.5	29.3	29.5	29.1	1.8
9	G4(H21)	DB1976(N12)	10.8	10.5	9.7	10.3	5.5

Table 3 Stochastic volatility parameters of interior helical dynamics
The Bayesian estimates on the volatility-specifying parameter σ are given with uncertainties as the 95% highest probability density (HPD, also known as the 95% credible interval).

Sequence	Free DNA	DB1976-bound
A_3T_3	0.142 (0.134, 0.151)	0.117 (0.110, 0.124)
A_2T_2	0.132 (0.124, 0.140)	0.103 (0.096, 0.109)
A_2CGT_2	0.136 (0.128, 0.144)	0.160 (0.152, 0.168)

Table 4 Structural parameters derived from MD simulations of free and bound DB1976 and DNA

Solvent-excluded volumes and SASA were determined based on a probe radius of 1.4 Å. Errors were estimated by averaging five equally sized blocks over the trajectory. 1 Å3 ~ 0.602 cm3/mol. 1 $Å^2$ = 10-16 cm^2.

	V_M, $Å^3$ or (cm^3/mol)			SASA, $Å^2 = 10^{-16}$ cm^2		
	A_3T_3	**A_2T_2**	**A_2CGT_2**	**A_3T_3**	**A_2T_2**	**A_2CGT_2**
DB1976		413 ± 1 248 ± 1			625 ± 2	
DNA	6,651 ± 2 (4,005 ± 2)	6,625 ± 3 (3,989 ± 2)	6,623 ± 2 (3,988 ± 2)	4,792 ± 13	4,759 ± 5	4,739 ± 10
Complex	7,119 ± 3 (4,287 ± 2)	7,084 ± 2 (4,266 ± 1)	7,090 ± 4 (4,269 ± 2)	4,636 ± 7	4,600 ± 5	4,596 ± 10
Change	56 ± 4 (34 ± 3)	46 ± 4 (28 ± 2)	54 ± 5 (33 ± 3)	-781 ± 15	-784 ± 7	-768 ± 14

Table 5 Estimates of global changes in hydration from volumetric measurements
SASA were converted to thermal volumes based on a thermal layer of δ = 0.55 Å. Interaction volumes and hydration changes were computed using Eqs. (5) and (7).

	A_3T_3	**A_2T_2**	**A_2CGT_2**
ΔV_c, cm^3/mol	-93 ± 2	-149 ± 4	-43 ± 2
ΔV_M, cm^3/mol	34 ± 2	28 ± 2	33 ± 3
ΔV_T, cm^3/mol	-259 ± 5	-260 ± 2	-254 ± 4
ΔV_I, cm^3/mol	132 ± 6	84 ± 5	178 ± 6
Δn_h	-73 ± 3	-47 ± 3	-99 ± 3

Table 6 Ion-water linkage parameters of DB1976/DNA complexes

Titrations of DB1976 with A3T3, A2T2, and A2CGT2 duplexes in the presence of up to 3.5 M NaCl were fitted globally with Eq. (9) with as a shared parameter. is shown for completeness but should not be physically interpreted beyond a constant of integration.

DNA	$\Delta n_{\pm}$	Δn_w	$\log K_0$
A_3T_3		-35 ± 3	-8.51 ± 0.08
A_2T_2	-2.4 ± 0.1	-36 ± 3	-7.70 ± 0.09
A_2CGT_2		-50 ± 4	-4.76 ± 0.11

Table 7 Hydration dynamics of DB1976/DNA complexes

Decay curves such as shown in Figure 2.13D are fitted to Eq. (11) in the main text and presented as mean ± S.E. of five non-overlapping instantiations. F, free; B, bound complex.

Population, N_i

DNA	N_1	N_2	N_3	$N(0)=\sum_{i=1}^{3} N_i$
A_3T_3	148 ± 1 (F) 134 ± 1 (B)	245 ± 2 249 ± 1	40 ± 2 42 ± 1	433 ± 3 425 ± 2
A_2T_2	145 ± 2 131 ± 1	242 ± 1 244 ± 2	46 ± 2 50 ± 3	433 ± 3 425 ± 4
A_2CGT_2	144 ± 1 146 ± 3	245 ± 1 246 ± 2	48 ± 1 25 ± 2	436 ± 2 416 ± 4

Lifetimes, τ_i

DNA	τ_1	τ_2	τ_3
A_3T_3	10.4 ± 0.3 (F) 8.7 ± 0.1 (B)	10.3 ± 0.3 8.2 ± 0.1	10.0 ± 0.2 10.6 ± 0.5
A_2T_2	85 ± 2 76 ± 1	84 ± 2 73 ± 1	82 ± 1 88 ± 3
A_2CGT_2	565 ± 31 358 ± 14	564 ± 36 302 ± 15	502 ± 26 597 ± 66

3 CONCLUSION

The previous studies on A-T specific minor groove binding compounds establish a strong binding with at least four adjacent A-T base pairs. DB1976 is able to cover a wide dispersal of affinities, including non-specific binding that is limited in the binding profile of DAPI, Hoechst dyes, and netropsin (44, 71, 78). We have shown that insertion of a GC base pair into a central A-T sequence causes a huge decrease in the binding constant, which K_D is approximately 0.70×10^{-9} M and 4.52×10^{-7} M in A_2T_2 and A_2CGT_2, respectively, at 0.5M NaCl solution. Besides A-T rich contiguous region, a genome that contains mixed GC and AT regions is also observed. For example, the mitochondrial DNAs of minicircle of kinetoplastids contain multiple three consecutive A-T bases separated by a single GC base pair (128). Understanding the driving force and factors that drive a non-specific binding target is necessary to minimize non-specific binding in order to maximize specific binding; or to maximize non-specific bindings to an extent that achieves the cooperative system in an adjacent mixed AT and GC region, and target a more consensus binding sites of other diseases for drug development.

3.1 MD simulation in application of ligand binding's effects on DNA

The ability of DB1976 to bind to a wide range of binding affinity with a distinct mode of terminal dynamic profiles deserves further attention. As indicated in Chapter 2, binding of DB1976 to A_3T_3 perturbs the terminal fraying of oligomeric DNA while A_2CGT_2 binding has minimal effect on terminal deforming. The question is how the stability of base pairs adjacent to the binding site is affected upon specific and non-specific binding.

3.1.1 A quick recap about role of hydration water in the event of opened terminal base pairs

In general, the intact termini are less hydrated than deformed states, as indicated in Figure 2.9B in chapter 2. However, we are missing the information about what factors can prolong the terminal-end opening states.

A simulation with various temperatures, for example, 25^0, 50^0, and 70 ^{0}C should be tested. The fraying (or deformed states) is observed as a reversible process at room temperature. The temperature difference is used to weaken the duplex stability at the termini and prolong the opened-end states. If we probe the separation distance (SD) between terminal ends, a plateau in SD, where the open ends are prolonged, can be coupled with water and ions counting at that specific time frame. This study can help probe the role of water molecules and ions in the separation of terminal base pairs. The question is whether the number of water molecules in the first hydration shell increases or decreases during the plateau phase? The increase in the water of hydration around the terminal ends may indicate that water plays a role in maintaining the fraying states.

3.1.2 A study on minor-groove width to understand the stability generated between the ligand-DNA interaction

The minor-groove width is also an important factor to be studied in temperature range study. Upon formation, the minor-groove width can be affected to a minimal extend by DB1976. If we compare the minor-groove width at the binding site of unbound and bound states as temperatures increase, the information about the level of stability between DNA and ligand can be observed. This test provides the effort to observe the stability of ligand-DNA formation between specific and non-specific binding. The two possibilities include: the groove width of both specific

and non-specific has a slight difference with rising temperature, or the width at the bound state possesses a significant difference.

3.1.3 What will happened if the DNA length is expanded?

The dynamic of terminal ends should be investigated in a longer DNA sequence. A 2-4 base pairs extension steps can provide further information about the effect of DB1976 binding on base pairs adjacent to the binding site. This also helps clarify if the opened-termini event is the result of short oligo's nature or the binding of DB1976 modifies the stability of base pairs adjacent to the binding site.

3.1.4 Is interaction between DB1976 and single strand DNA (ssDNA) possible?

The well-defined binding model of the minor-groove binder is discussed in Chapter 1. DB1976 binding occurs between opposite strands of double-stranded DNA (*interstrand*); however, we did not clarify the "binding action steps". Overall, the binding occurs by fitting into the groove. Does DB1976 fit the entire structure (or its body) into the groove in a one-way step? Or does DB1976 interact with one-side of the complementary strand first and adjust the position along the groove to search for the possibility of bonding? A simulation of ssDNA for both specific and non-specific binding with DB1976 can be used to validate any possible interaction between DB1976 and ssDNA. This test may not provide enough information the address the question; however, it can provide an insight into the binding intermediate steps that require further experimental studies.

3.2 The hydration and dynamic properties of diamidines family clarify by the difference in structure and binding mode.

3.2.1 The hydration and dynamic properties of diamidines in different binding profiles

Based on the properties of binding affinity, hydration, and the dynamic effect on the DNA of DB1976, the following compounds should also be investigated: DB 293 and DB 1242 (Figure 3.1). DB293 is previously discovered that binds to the single GC inserted sequence (5′- ATGA-3′) as a 2:1 complex while binding to A-T rich sequence in a 1:1 manner (129). The ability of both DB1976 and DB293 to bind to specific and non-specific bindings with two distinct binding characteristics should be investigated on both hydration and binding dynamic aspects. The DB1242 has an amidine-phenyl-pyrimidine and phenyl-amidine moieties with a distinct linear shape (8). DB1242 binds to AT-containing DNA sequence weakly; however, it binds strongly to the GC rich sequence, 5′-GCTCG-3′ as a dimer. A G-C sequence containing a single or double A-T insertion should be studied between DB1976 and DB1242.

3.2.2 The structure-dependent profiles on hydration and dynamic properties

DB1977 is different from DB1976 at the terminal dication and shows a strong binding site to the AT-rich region. By the SPR study, DB1977 establishes a familiar binding affinity to the λB site, compared to DB1976. The λB site is a sequence that contains an ETS consensus binding site with an extended AT-rich region (23). The difference at the structurally terminal ends between the two compounds can provide more information about the diamidine's binding profile on hydration and dynamic variation (Figure 3.1B).

DB270 is the bisbenzimidazole-furan analog of DB293, as mentioned in Chapter 2. DB1976 binds to each sequence with an offset by a half base step. DB270 contains oxygen instead

of selenium. The binding positions and their effect on the terminal-end dynamic may be observed by the comparison between DB270 and DB1976 (Figure 3.1B).

DB1281, a monobenzimidazole-selenophene, contains a non-symmetrical shape. If the dynamic on each side of terminal base pairs is different, the structural effects of diamines on the adjacent bases pair to the binding site can be interpreted (Figure 3.1B).

Those compounds should be investigated further to be able to draw a complete profile of the role of the dynamic domain at the terminus on non-specific binding optimization.

3.3 An experimental design to probe end fraying of the unbound and bound oligonucleotides

The nature of end fraying in oligonucleotides is discussed in Chapter 1. The reduction of structural constraints at the termini is the major explanation for its open state. Even the open ends may not structurally affect the nearby helix; the environmental conditions surrounding the interior of the duplex may have changed when it goes through the opening ends. When the termini open, it provides greater water exposure to the interior area. The overall water of the hydration network may process a different orientation. Also, the overall counter ions network condensation should have shown a different packing. The overall behavior at the terminus of native DNA and ligand-DNA is important to address the question of how the specific binding of DB1976 gains the stability of terminal ends. An experimental design to directly probe the fraying events are necessary to compromise with MD simulation data.

3.4 Does the binding affinity and hydration release of DB1976 depend on the order of binding site sequence?

We have compared the binding affinity and hydration properties of A_3T_3 and A_2T_2. With an addition of a single AT base pairs, the binding affinity increases in occupation of higher number

of water release. The order of base pairs in DNA is important as the base-stacking forces follow an order-dependent manner. To fully optimize the binding affinity of the compounds, a mixed order of the binding site is necessary.

3.5 DB1976 binding behavior in the presence of both specific and non-specific binding

DB1976 has established a distinct profile on binding to both specific and non-specific binding. In the presence of both specific and non-specific binding sites, such that DNA contains both AAATTT and AACGTT regions, the question is to identify which region DB1976 will bind to with a favorable thermodynamic profile. As mixed regions of specific and nonspecific binding are common in the genomic sequence, this study helps evaluate the selectivity of diamidines on DNA binding.

One remarkable binding ability of DB1976 is that it can expand to binding of duplexes spanning $>10^3$-fold in binding affinity. We may apply this property in targeting base pairs mismatch in duplexes such as C-C or G-G region. DB1976 and the DB compound library possess a unique fluorescent characteristic where the intensity is usually affected by the structure of the compounds. A study on the binding affinity of DB1976 to the mismatch region in duplexes should provide a new approach in using DB compounds as a mix-match DNA recognition probe.

3.6 Overall picture for future direction

The study has contained some weaknesses, such as the interpretation of volumetric data is completed based on molecular dynamic simulation to grant access to averaged changes in solvent-excluded volume (ΔV_M) and change in a solute's intrinsic (molecular) volume. We have implied those values by molecular dynamic simulation as no high-resolution structural data is available for the DB1976-DNA complex.

The goal of the future project is to establish a structure-hydration relationship for future molecular design in therapeutic application. The structural elements that contribute to the minor groove recognition can be used as a key factor to address drug-DNA targeting design. From the discussion above, additional experiments should be conducted on other DB compounds. Those studies will help provide future directions to design rules for manipulating hydration changes and dynamic induced to the DNA based on structural modifications to maximize the desired binding affinity of the drugs.

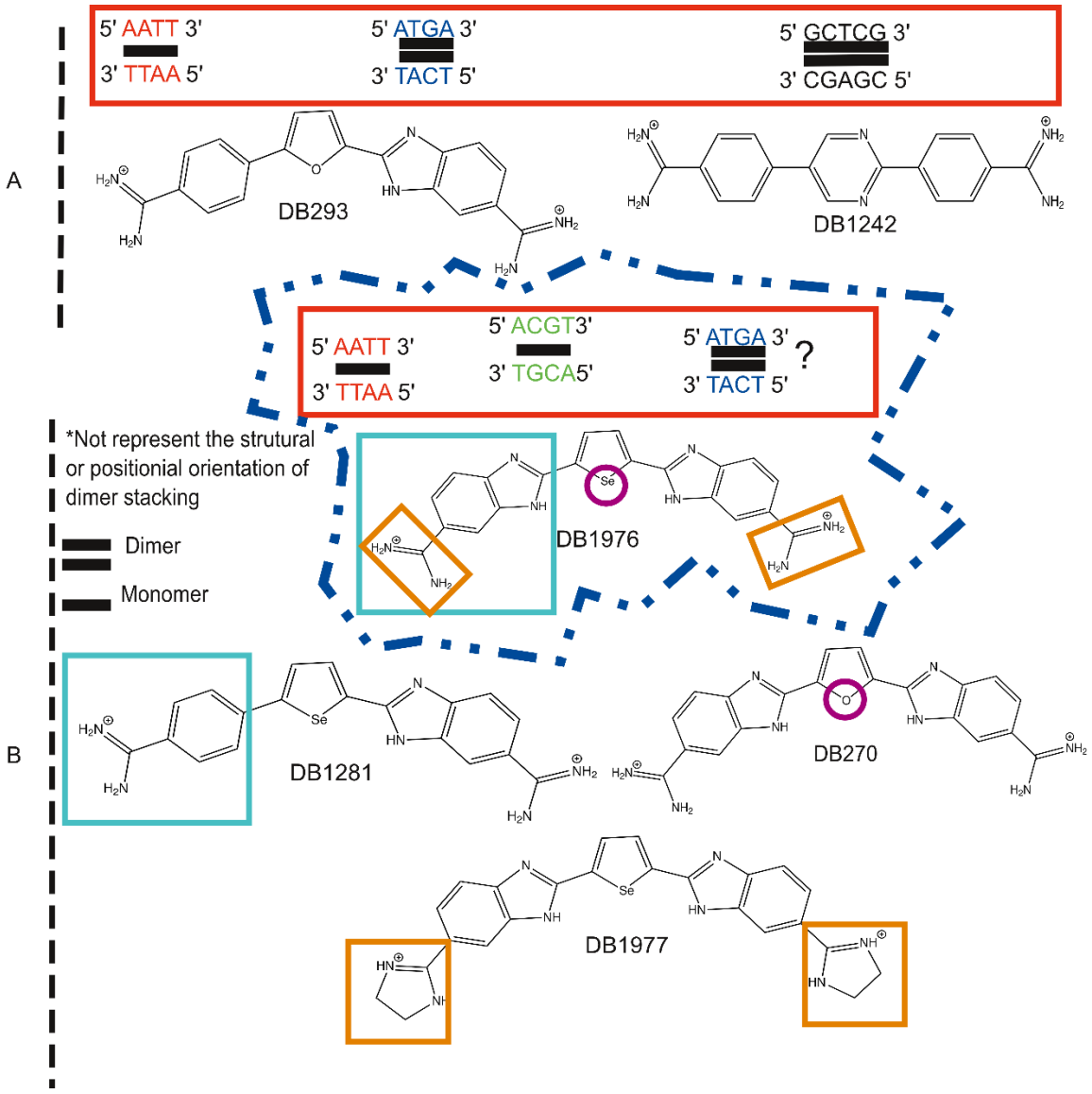


Figure 3-1 A comparison chart between DB1976 and other diamidines compounds to visualize the direction of the future study

A. The comparison between the binding modes as pointed out in section 3.2.1. B. The comparation in structural difference between each compound (section 3.2.2). Note, the color of each square or circle guides the functional groups that are compared to the reference compound, DB1976. The red rectangles are used to group the binding mode category. Only the central part of the DNA groove is listed. The dimer and monomer positions are just illustration, not a particular position of binding or positional stacking of dimers.

REFERENCES

1. Costales MG, Childs-Disney JL, Haniff HS, Disney MD. How We Think about Targeting RNA with Small Molecules. J Med Chem. 2020;63(17):8880-900. Epub 2020/03/28. doi: 10.1021/acs.jmedchem.9b01927. PubMed PMID: 32212706; PMCID: PMC7486258.
2. Kayode O, Huang Z, Soares AS, Caulfield TR, Dong Z, Bode AM, Radisky ES. Small molecule inhibitors of mesotrypsin from a structure-based docking screen. PLoS One. 2017;12(5):e0176694. Epub 2017/05/04. doi: 10.1371/journal.pone.0176694. PubMed PMID: 28463992; PMCID: PMC5413004.
3. Gurova K. New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. Future Oncol. 2009;5(10):1685-704. Epub 2009/12/17. doi: 10.2217/fon.09.127. PubMed PMID: 20001804; PMCID: PMC2821823.
4. Waring M. Variation of the supercoils in closed circular DNA by binding of antibiotics and drugs: evidence for molecular models involving intercalation. J Mol Biol. 1970;54(2):247-79. Epub 1970/12/14. doi: 10.1016/0022-2836(70)90429-8. PubMed PMID: 5500444.
5. Wartell RM, Larson JE, Wells RD. Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. J Biol Chem. 1974;249(21):6719-31. Epub 1974/11/10. PubMed PMID: 4371420.
6. Baraldi PG, Bovero A, Fruttarolo F, Preti D, Tabrizi MA, Pavani MG, Romagnoli R. DNA minor groove binders as potential antitumor and antimicrobial agents. Med Res Rev. 2004;24(4):475-528. Epub 2004/06/02. doi: 10.1002/med.20000. PubMed PMID: 15170593.

7. Pjura PE, Grzeskowiak K, Dickerson RE. Binding of Hoechst 33258 to the minor groove of B-DNA. J Mol Biol. 1987;197(2):257-71. Epub 1987/09/20. doi: 10.1016/0022-2836(87)90123-9. PubMed PMID: 2445998.

8. Nanjunda R, Wilson WD. Binding to the DNA minor groove by heterocyclic dications: from AT-specific monomers to GC recognition with dimers. Curr Protoc Nucleic Acid Chem. 2012;Chapter 8:Unit8 Epub 2012/12/21. doi: 10.1002/0471142700.nc0808s51. PubMed PMID: 23255206; PMCID: PMC3539175.

9. Pearson RD, Hewlett EL. Pentamidine for the treatment of Pneumocystis carinii pneumonia and other protozoal diseases. Ann Intern Med. 1985;103(5):782-6. Epub 1985/11/01. doi: 10.7326/0003-4819-103-5-782. PubMed PMID: 3901852.

10. Sands M, Kron MA, Brown RB. Pentamidine: a review. Rev Infect Dis. 1985;7(5):625-34. Epub 1985/09/01. doi: 10.1093/clinids/7.5.625. PubMed PMID: 3903942.

11. Babokhov P, Sanyaolu AO, Oyibo WA, Fagbenro-Beyioku AF, Iriemenam NC. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. Pathog Glob Health. 2013;107(5):242-52. Epub 2013/08/07. doi: 10.1179/2047773213Y.0000000105. PubMed PMID: 23916333; PMCID: PMC4001453.

12. Edwards KJ, Jenkins TC, Neidle S. Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties. Biochemistry. 1992;31(31):7104-9. Epub 1992/08/11. doi: 10.1021/bi00146a011. PubMed PMID: 1643044.

13. Thuita JK, Karanja SM, Wenzler T, Mdachi RE, Ngotho JM, Kagira JM, Tidwell R, Brun R. Efficacy of the diamidine DB75 and its prodrug DB289, against murine models of human African trypanosomiasis. Acta Trop. 2008;108(1):6-10. Epub 2008/08/30. doi: 10.1016/j.actatropica.2008.07.006. PubMed PMID: 18722336.

14. Purfield AE, Tidwell RR, Meshnick SR. The diamidine DB75 targets the nucleus of Plasmodium falciparum. Malar J. 2009;8:104. Epub 2009/05/16. doi: 10.1186/1475-2875-8-104. PubMed PMID: 19442305; PMCID: PMC2689252.

15. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science. 1994;265(5178):1573-7. Epub 1994/09/09. doi: 10.1126/science.8079170. PubMed PMID: 8079170.

16. McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, Klemsz M, Feeney AJ, Wu GE, Paige CJ, Maki RA. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J. 1996;15(20):5647-58. Epub 1996/10/15. PubMed PMID: 8896458; PMCID: PMC452309.

17. Steidl U, Rosenbauer F, Verhaak RG, Gu X, Ebralidze A, Otu HH, Klippel S, Steidl C, Bruns I, Costa DB, Wagner K, Aivado M, Kobbe G, Valk PJ, Passegue E, Libermann TA, Delwel R, Tenen DG. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. Nat Genet. 2006;38(11):1269-77. Epub 2006/10/17. doi: 10.1038/ng1898. PubMed PMID: 17041602.

18. Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, Akashi K, Fiering S, Tenen DG. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. Nat Genet. 2004;36(6):624-30. Epub 2004/05/18. doi: 10.1038/ng1361. PubMed PMID: 15146183.

19. Antony-Debre I, Paul A, Leite J, Mitchell K, Kim HM, Carvajal LA, Todorova TI, Huang K, Kumar A, Farahat AA, Bartholdy B, Narayanagari SR, Chen J, Ambesi-Impiombato A, Ferrando AA, Mantzaris I, Gavathiotis E, Verma A, Will B, Boykin DW, Wilson WD, Poon GM, Steidl U. Pharmacological inhibition of the transcription factor

PU.1 in leukemia. J Clin Invest. 2017;127(12):4297-313. Epub 2017/10/31. doi: 10.1172/JCI92504. PubMed PMID: 29083320; PMCID: PMC5707147.

20. Liu Q, Yu J, Wang L, Tang Y, Zhou Q, Ji S, Wang Y, Santos L, Haeusler RA, Que J, Rajbhandari P, Lei X, Valenti L, Pajvani UB, Qin J, Qiang L. Inhibition of PU.1 ameliorates metabolic dysfunction and non-alcoholic steatohepatitis. J Hepatol. 2020;73(2):361-70. Epub 2020/03/07. doi: 10.1016/j.jhep.2020.02.025. PubMed PMID: 32135178.

21. Wohlfahrt T, Rauber S, Uebe S, Luber M, Soare A, Ekici A, Weber S, Matei AE, Chen CW, Maier C, Karouzakis E, Kiener HP, Pachera E, Dees C, Beyer C, Daniel C, Gelse K, Kremer AE, Naschberger E, Sturzl M, Butter F, Sticherling M, Finotto S, Kreuter A, Kaplan MH, Jungel A, Gay S, Nutt SL, Boykin DW, Poon GMK, Distler O, Schett G, Distler JHW, Ramming A. PU.1 controls fibroblast polarization and tissue fibrosis. Nature. 2019;566(7744):344-9. Epub 2019/02/01. doi: 10.1038/s41586-019-0896-x. PubMed PMID: 30700907; PMCID: PMC6526281.

22. Kodandapani R, Pio F, Ni CZ, Piccialli G, Klemsz M, McKercher S, Maki RA, Ely KR. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature. 1996;380(6573):456-60. Epub 1996/04/04. doi: 10.1038/380456a0. PubMed PMID: 8602247.

23. Munde M, Wang S, Kumar A, Stephens CE, Farahat AA, Boykin DW, Wilson WD, Poon GM. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. Nucleic Acids Res. 2014;42(2):1379-90. Epub 2013/10/26. doi: 10.1093/nar/gkt955. PubMed PMID: 24157839; PMCID: PMC3902942.

24. Stephens DC, Kim HM, Kumar A, Farahat AA, Boykin DW, Poon GM. Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. Nucleic Acids Res. 2016;44(9):4005-13. Epub 2016/04/16. doi: 10.1093/nar/gkw229. PubMed PMID: 27079976; PMCID: PMC4872103.

25. Chen X, Weber I, Harrison RW. Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. J Phys Chem B. 2008;112(38):12073-80. Epub 2008/08/30. doi: 10.1021/jp802795a. PubMed PMID: 18754631; PMCID: PMC2768875.

26. Galib M, Baer MD, Skinner LB, Mundy CJ, Huthwelker T, Schenter GK, Benmore CJ, Govind N, Fulton JL. Revisiting the hydration structure of aqueous Na(). J Chem Phys. 2017;146(8):084504. Epub 2017/03/03. doi: 10.1063/1.4975608. PubMed PMID: 28249415.

27. Careri G, Gratton E, Yang PH, Rupley JA. Correlation of IR spectroscopic, heat capacity, diamagnetic susceptibility and enzymatic measurements on lysozyme powder. Nature. 1980;284(5756):572-3. Epub 1980/04/10. doi: 10.1038/284572a0. PubMed PMID: 7366728.

28. Schiebel J, Gaspari R, Wulsdorf T, Ngo K, Sohn C, Schrader TE, Cavalli A, Ostermann A, Heine A, Klebe G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. Nat Commun. 2018;9(1):3559. Epub 2018/09/05. doi: 10.1038/s41467-018-05769-2. PubMed PMID: 30177695; PMCID: PMC6120877.

29. Abel R, Salam NK, Shelley J, Farid R, Friesner RA, Sherman W. Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation

factor serine proteases. ChemMedChem. 2011;6(6):1049-66. Epub 2011/04/21. doi: 10.1002/cmdc.201000533. PubMed PMID: 21506273.

30. Dragan AI, Read CM, Crane-Robinson C. Enthalpy-entropy compensation: the role of solvation. Eur Biophys J. 2017;46(4):301-8. Epub 2016/11/01. doi: 10.1007/s00249-016-1182-6. PubMed PMID: 27796417; PMCID: PMC5384952.

31. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE. The molecular origin of DNA-drug specificity in netropsin and distamycin. Proc Natl Acad Sci U S A. 1985;82(5):1376-80. Epub 1985/03/01. doi: 10.1073/pnas.82.5.1376. PubMed PMID: 2983343; PMCID: PMC397264.

32. Drew HR, Dickerson RE. Structure of a B-DNA dodecamer. III. Geometry of hydration. J Mol Biol. 1981;151(3):535-56. Epub 1981/09/25. doi: 10.1016/0022-2836(81)90009-7. PubMed PMID: 7338904.

33. Kopka ML, Fratini AV, Drew HR, Dickerson RE. Ordered water structure around a B-DNA dodecamer. A quantitative study. J Mol Biol. 1983;163(1):129-46. Epub 1983/01/05. doi: 10.1016/0022-2836(83)90033-5. PubMed PMID: 6834428.

34. Schneider B, Cohen D, Berman HM. Hydration of DNA bases: analysis of crystallographic data. Biopolymers. 1992;32(7):725-50. Epub 1992/07/01. doi: 10.1002/bip.360320703. PubMed PMID: 1391627.

35. Shui X, McFail-Isom L, Hu GG, Williams LD. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. Biochemistry. 1998;37(23):8341-55. Epub 1998/06/19. doi: 10.1021/bi973073c. PubMed PMID: 9622486.

36. Denisov VP, Carlstrom G, Venu K, Halle B. Kinetics of DNA hydration. J Mol Biol. 1997;268(1):118-36. Epub 1997/04/25. doi: 10.1006/jmbi.1996.0862. PubMed PMID: 9149146.

37. Liepinsh E, Otting G, Wuthrich K. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. Nucleic Acids Res. 1992;20(24):6549-53. Epub 1992/12/25. doi: 10.1093/nar/20.24.6549. PubMed PMID: 1480475; PMCID: PMC334570.

38. Chuprina VP, Heinemann U, Nurislamov AA, Zielenkiewicz P, Dickerson RE, Saenger W. Molecular dynamics simulation of the hydration shell of a B-DNA decamer reveals two main types of minor-groove hydration depending on groove width. Proc Natl Acad Sci U S A. 1991;88(2):593-7. Epub 1991/01/15. doi: 10.1073/pnas.88.2.593. PubMed PMID: 1988954; PMCID: PMC50858.

39. Qu X, Chaires JB. Hydration changes for DNA intercalation reactions. J Am Chem Soc. 2001;123(1):1-7. Epub 2001/03/29. doi: 10.1021/ja002793v. PubMed PMID: 11273594.

40. Alniss HY. Thermodynamics of DNA Minor Groove Binders. J Med Chem. 2019;62(2):385-402. Epub 2018/07/31. doi: 10.1021/acs.jmedchem.8b00233. PubMed PMID: 30059627.

41. Liu Y, Kumar A, Depauw S, Nhili R, David-Cordonnier MH, Lee MP, Ismail MA, Farahat AA, Say M, Chackal-Catoen S, Batista-Parra A, Neidle S, Boykin DW, Wilson WD. Water-mediated binding of agents that target the DNA minor groove. J Am Chem Soc. 2011;133(26):10171-83. Epub 2011/06/02. doi: 10.1021/ja202006u. PubMed PMID: 21627167; PMCID: PMC3165004.

42. Nguyen B, Neidle S, Wilson WD. A role for water molecules in DNA-ligand minor groove recognition. Acc Chem Res. 2009;42(1):11-21. Epub 2008/09/19. doi: 10.1021/ar800016q. PubMed PMID: 18798655; PMCID: PMC2668807.

43. Chalikian TV, Breslauer KJ. Volumetric properties of nucleic acids. Biopolymers. 1998;48(4):264-80. Epub 2000/03/04. doi: 10.1002/(sici)1097-0282(1998)48:4<264::aid-bip6>3.3.co;2-#. PubMed PMID: 10699844.

44. Han F, Taulier N, Chalikian TV. Association of the minor groove binding drug Hoechst 33258 with d(CGCGAATTCGCG)2: volumetric, calorimetric, and spectroscopic characterizations. Biochemistry. 2005;44(28):9785-94. Epub 2005/07/13. doi: 10.1021/bi047374f. PubMed PMID: 16008363.

45. Hackel M, Hinz HJ, Hedwig GR. Partial molar volumes of proteins: amino acid side-chain contributions derived from the partial molar volumes of some tripeptides over the temperature range 10-90 degrees C. Biophys Chem. 1999;82(1):35-50. Epub 2006/10/13. doi: 10.1016/s0301-4622(99)00104-0. PubMed PMID: 17030339.

46. Chalikian TV. Volumetric properties of proteins. Annu Rev Biophys Biomol Struct. 2003;32:207-35. Epub 2003/01/25. doi: 10.1146/annurev.biophys.32.110601.141709. PubMed PMID: 12543706.

47. Chalikian TV, Volker J, Srinivasan AR, Olson WK, Breslauer KJ. The hydration of nucleic acid duplexes as assessed by a combination of volumetric and structural techniques. Biopolymers. 1999;50(5):459-71. Epub 1999/09/10. doi: 10.1002/(SICI)1097-0282(19991015)50:5<459::AID-BIP1>3.0.CO;2-B. PubMed PMID: 10479730.

48. Chalikian TV, Plum GE, Sarvazyan AP, Breslauer KJ. Influence of drug binding on DNA hydration: acoustic and densimetric characterizations of netropsin binding to the poly(dAdT).poly(dAdT) and poly(dA).poly(dT) duplexes and the poly(dT).poly(dA).poly(dT) triplex at 25 degrees C. Biochemistry. 1994;33(29):8629-40. Epub 1994/07/26. doi: 10.1021/bi00195a003. PubMed PMID: 8038152.

49. Harding SE. The hydration problem in solution biophysics: an introduction. Biophys Chem. 2001;93(2-3):87-91. Epub 2002/01/24. doi: 10.1016/s0301-4622(01)00213-7. PubMed PMID: 11804718.

50. Frueh DP, Goodrich AC, Mishra SH, Nichols SR. NMR methods for structural studies of large monomeric and multimeric proteins. Curr Opin Struct Biol. 2013;23(5):734-9. Epub 2013/07/16. doi: 10.1016/j.sbi.2013.06.016. PubMed PMID: 23850141; PMCID: PMC3805735.

51. McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. Acta Crystallogr F Struct Biol Commun. 2014;70(Pt 11):1445-67. Epub 2014/11/06. doi: 10.1107/S2053230X14019670. PubMed PMID: 25372810; PMCID: PMC4231845.

52. Neidle S. DNA minor-groove recognition by small molecules. Nat Prod Rep. 2001;18(3):291-309. Epub 2001/07/31. doi: 10.1039/a705982e. PubMed PMID: 11476483.

53. Reddy CK, Das A, Jayaram B. Do water molecules mediate protein-DNA recognition? J Mol Biol. 2001;314(3):619-32. Epub 2002/02/16. doi: 10.1006/jmbi.2001.5154. PubMed PMID: 11846571.

54. Chalikian TV, Sarvazyan AP, Breslauer KJ. Hydration and partial compressibility of biological compounds. Biophys Chem. 1994;51(2-3):89-107; discussion -9. Epub 1994/08/01. doi: 10.1016/0301-4622(94)85007-0. PubMed PMID: 7919045.

55. Chalikian TV, Breslauer KJ. Thermodynamic analysis of biomolecules: a volumetric approach. Curr Opin Struct Biol. 1998;8(5):657-64. Epub 1998/11/18. doi: 10.1016/s0959-440x(98)80159-0. PubMed PMID: 9818272.

56. Son I, Shek YL, Tikhomirova A, Baltasar EH, Chalikian TV. Interactions of urea with native and unfolded proteins: a volumetric study. J Phys Chem B. 2014;118(47):13554-63. Epub 2014/11/05. doi: 10.1021/jp509356k. PubMed PMID: 25365737.

57. Son I, Selvaratnam R, Dubins DN, Melacini G, Chalikian TV. Ultrasonic and densimetric characterization of the association of cyclic AMP with the cAMP-binding domain of the exchange protein EPAC1. J Phys Chem B. 2013;117(37):10779-84. Epub 2013/08/24. doi: 10.1021/jp406451p. PubMed PMID: 23968295.

58. Son I, Shek YL, Dubins DN, Chalikian TV. Volumetric characterization of tri-N-acetylglucosamine binding to lysozyme. Biochemistry. 2012;51(29):5784-90. Epub 2012/06/27. doi: 10.1021/bi3006994. PubMed PMID: 22732010.

59. Timasheff SN. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. Proc Natl Acad Sci U S A. 2002;99(15):9721-6. Epub 2002/07/05. doi: 10.1073/pnas.122225399. PubMed PMID: 12097640; PMCID: PMC124992.

60. Timasheff SN. In disperse solution, "osmotic stress" is a restricted case of preferential interactions. Proc Natl Acad Sci U S A. 1998;95(13):7363-7. Epub 1998/06/24. doi: 10.1073/pnas.95.13.7363. PubMed PMID: 9636154; PMCID: PMC22618.

61. Kiser JR, Monk RW, Smalls RL, Petty JT. Hydration changes in the association of Hoechst 33258 with DNA. Biochemistry. 2005;44(51):16988-97. Epub 2005/12/21. doi: 10.1021/bi051769x. PubMed PMID: 16363812; PMCID: PMC6158785.

62. Degtyareva NN, Wallace BD, Bryant AR, Loo KM, Petty JT. Hydration changes accompanying the binding of minor groove ligands with DNA. Biophys J. 2007;92(3):959-65. Epub 2006/11/23. doi: 10.1529/biophysj.106.097451. PubMed PMID: 17114230; PMCID: PMC1779984.

63. Anuradha, Alam MS, Chaudhury NK. Osmolyte changes the binding affinity and mode of interaction of minor groove binder hoechst 33258 with calf thymus DNA. Chem Pharm Bull (Tokyo). 2010;58(11):1447-54. Epub 2010/11/05. doi: 10.1248/cpb.58.1447. PubMed PMID: 21048335.

64. Qamhieh K, Wong KY, Lynch GC, Pettitt BM. The Melting Mechanism of DNA Tethered to a Surface. Int J Numer Anal Model. 2009;6(3):474-88. Epub 2009/10/06. PubMed PMID: 19802357; PMCID: PMC2755589.

65. Perez A, Orozco M. Real-time atomistic description of DNA unfolding. Angew Chem Int Ed Engl. 2010;49(28):4805-8. Epub 2010/05/19. doi: 10.1002/anie.201000593. PubMed PMID: 20480472.

66. Nonin S, Leroy JL, Gueron M. Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. Biochemistry. 1995;34(33):10652-9. Epub 1995/08/22. doi: 10.1021/bi00033a041. PubMed PMID: 7654719.

67. Zgarbova M, Otyepka M, Sponer J, Lankas F, Jurecka P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. J Chem Theory Comput. 2014;10(8):3177-89. Epub 2014/08/12. doi: 10.1021/ct500120v. PubMed PMID: 26588288.

68. Zhang XX, Brantley SL, Corcelli SA, Tokmakoff A. DNA minor-groove binder Hoechst 33258 destabilizes base-pairing adjacent to its binding site. Commun Biol. 2020;3(1):525. Epub 2020/09/24. doi: 10.1038/s42003-020-01241-4. PubMed PMID: 32963293; PMCID: PMC7508854.

69. Chaires JB. A thermodynamic signature for drug-DNA binding mode. Arch Biochem Biophys. 2006;453(1):26-31. doi: 10.1016/j.abb.2006.03.027. PubMed PMID: 16730635.

70. Wang S, Kumar A, Aston K, Nguyen B, Bashkin JK, Boykin DW, Wilson WD. Different thermodynamic signatures for DNA minor groove binding with changes in salt concentration and temperature. Chem Commun (Camb). 2013;49(76):8543-5. Epub 2013/08/16. doi: 10.1039/c3cc44569k. PubMed PMID: 23945614; PMCID: PMC3791883.

71. Marky LA, Breslauer KJ. Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data. Proc Natl Acad Sci U S A. 1987;84(13):4359-63. Epub 1987/07/01. PubMed PMID: 3037518; PMCID: PMC305088.

72. Haq I, Ladbury JE, Chowdhry BZ, Jenkins TC, Chaires JB. Specific binding of hoechst 33258 to the d(CGCAAATTTGCG)2 duplex: calorimetric and spectroscopic studies11Edited by I. Tinoco. Journal of Molecular Biology. 1997;271(2):244-57. doi: https://doi.org/10.1006/jmbi.1997.1170.

73. Loontiens FG, Regenfuss P, Zechel A, Dumortier L, Clegg RM. Binding characteristics of Hoechst 33258 with calf thymus DNA, poly[d(A-T)], and d(CCGGAATTCCGG): multiple stoichiometries and determination of tight binding with a wide spectrum of site

affinities. Biochemistry. 1990;29(38):9029-39. Epub 1990/09/25. doi: 10.1021/bi00490a021. PubMed PMID: 1702995.

74. Pal SK, Zhao L, Zewail AH. Water at DNA surfaces: ultrafast dynamics in minor groove recognition. Proc Natl Acad Sci U S A. 2003;100(14):8113-8. Epub 2003/06/20. doi: 10.1073/pnas.1433066100. PubMed PMID: 12815094; PMCID: PMC166191.

75. Duboue-Dijon E, Fogarty AC, Hynes JT, Laage D. Dynamical Disorder in the DNA Hydration Shell. J Am Chem Soc. 2016;138(24):7610-20. Epub 2016/05/31. doi: 10.1021/jacs.6b02715. PubMed PMID: 27240107.

76. Laage D, Elsaesser T, Hynes JT. Water Dynamics in the Hydration Shells of Biomolecules. Chem Rev. 2017;117(16):10694-725. Epub 2017/03/02. doi: 10.1021/acs.chemrev.6b00765. PubMed PMID: 28248491; PMCID: PMC5571470.

77. Chalikian TV, Macgregor RB. Nucleic acid hydration: a volumetric perspective. Physics of Life Reviews. 2007;4(2):91-115. doi: 10.1016/j.plrev.2006.11.001. PubMed PMID: WOS:000248605100001.

78. Breusegem SY, Clegg RM, Loontiens FG. Base-sequence specificity of Hoechst 33258 and DAPI binding to five (A/T)4 DNA sites with kinetic evidence for more than one high-affinity Hoechst 33258-AATT complex. J Mol Biol. 2002;315(5):1049-61. Epub 2002/02/06. doi: 10.1006/jmbi.2001.5301. PubMed PMID: 11827475.

79. Erlitzki N, Huang K, Xhani S, Farahat AA, Kumar A, Boykin DW, Poon GMK. Investigation of the electrostatic and hydration properties of DNA minor groove-binding by a heterocyclic diamidine by osmotic pressure. Biophys Chem. 2017;231:95-104. Epub 2017/04/02. doi: 10.1016/j.bpc.2017.02.008. PubMed PMID: 28363467; PMCID: PMC5607081.

80. Marcus Y. The standard partial molar volumes of ions in solution. Part 4. Ionic volumes in water at 0-100 degrees C. J Phys Chem B. 2009;113(30):10285-91. Epub 2009/07/10. doi: 10.1021/jp9027244. PubMed PMID: 19585994.

81. Liu Y, Collar CJ, Kumar A, Stephens CE, Boykin DW, Wilson WD. Heterocyclic diamidine interactions at AT base pairs in the DNA minor groove: effects of heterocycle differences, DNA AT sequence and length. J Phys Chem B. 2008;112(37):11809-18. Epub 2008/08/23. doi: 10.1021/jp804048c. PubMed PMID: 18717551; PMCID: PMC2556899.

82. Ivani I, Dans PD, Noy A, Perez A, Faustino I, Hospital A, Walther J, Andrio P, Goni R, Balaceanu A, Portella G, Battistini F, Gelpi JL, Gonzalez C, Vendruscolo M, Laughton CA, Harris SA, Case DA, Orozco M. Parmbsc1: a refined force field for DNA simulations. Nat Methods. 2016;13(1):55-8. Epub 2015/11/17. doi: 10.1038/nmeth.3658. PubMed PMID: 26569599; PMCID: PMC4700514.

83. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys. 2004;120(20):9665-78. Epub 2004/07/23. doi: 10.1063/1.1683075. PubMed PMID: 15267980.

84. Stumpe MC, Blinov N, Wishart D, Kovalenko A, Pande VS. Calculation of local water densities in biological systems: a comparison of molecular dynamics simulations and the 3D-RISM-KH molecular theory of solvation. J Phys Chem B. 2011;115(2):319-28. Epub 2010/12/23. doi: 10.1021/jp102587q. PubMed PMID: 21174421; PMCID: PMC3407544.

85. Son I, Shek YL, Dubins DN, Chalikian TV. Hydration changes accompanying helix-to-coil DNA transitions. J Am Chem Soc. 2014;136(10):4040-7. Epub 2014/02/20. doi: 10.1021/ja5004137. PubMed PMID: 24548168.

86. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, 3rd, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys J. 2007;92(11):3817-29. doi: 10.1529/biophysj.106.097782. PubMed PMID: 17351000; PMCID: PMC1868997.

87. Galindo-Murillo R, Robertson JC, Zgarbova M, Sponer J, Otyepka M, Jurecka P, Cheatham TE, 3rd. Assessing the Current State of Amber Force Field Modifications for DNA. J Chem Theory Comput. 2016;12(8):4114-27. Epub 2016/06/15. doi: 10.1021/acs.jctc.6b00186. PubMed PMID: 27300587; PMCID: PMC4980684.

88. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15:1593-623. PubMed PMID: WOS:000338420000013.

89. Jacquier E, Polson NG, Rossi PE. Bayesian Analysis of Stochastic Volatility Models. J Bus Econ Stat. 2002;20(1):69-87. doi: 10.1198/073500102753410408. PubMed PMID: WOS:000172708800007.

90. Chalikian TV, Macgregor RB, Jr. On empirical decomposition of volumetric data. Biophys Chem. 2019;246:8-15. Epub 2019/01/01. doi: 10.1016/j.bpc.2018.12.005. PubMed PMID: 30597448.

91. Manning GS. Approximate Solutions to Some Problems in Polyelectrolyte Theory Involving Nonuniform Charge Distributions. Macromolecules. 2008;41(16):6217-27. doi: 10.1021/ma800628v. PubMed PMID: WOS:000258580700041.

92. Allison SA. End Effects in Electrostatic Potentials of Cylinders: Models for DNA Fragments. The Journal of Physical Chemistry. 1994;98(46):12091-6. doi: 10.1021/j100097a040.

93. Record MT, Jr., Zhang W, Anderson CF. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. Adv Protein Chem. 1998;51:281-353. Epub 1998/06/06. PubMed PMID: 9615173.

94. Tanford C. Extension of the theory of linked functions to incorporate the effects of protein hydration. J Mol Biol. 1969;39(3):539-44. Epub 1969/02/14. doi: 10.1016/0022-2836(69)90143-0. PubMed PMID: 5357211.

95. Record MT, Jr., Lohman ML, De Haseth P. Ion effects on ligand-nucleic acid interactions. J Mol Biol. 1976;107(2):145-58. Epub 1976/11/04. doi: 10.1016/s0022-2836(76)80023-x. PubMed PMID: 1003464.

96. Record MT, Jr., Anderson CF, Lohman TM. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q Rev Biophys. 1978;11(2):103-78. Epub 1978/05/01. PubMed PMID: 353875.

97. Stigter D, Dill KA. Binding of ionic ligands to polyelectrolytes. Biophys J. 1996;71(4):2064-74. Epub 1996/10/01. doi: 10.1016/S0006-3495(96)79405-5. PubMed PMID: 8889181; PMCID: PMC1233673.

98. Wilson RW, Rau DC, Bloomfield VA. Comparison of polyelectrolyte theories of the binding of cations to DNA. Biophys J. 1980;30(2):317-25. Epub 1980/05/01. doi: 10.1016/S0006-3495(80)85097-1. PubMed PMID: 7260278; PMCID: PMC1328737.

99. Shimizu S, Matubayasi N. Preferential solvation: dividing surface vs excess numbers. J Phys Chem B. 2014;118(14):3922-30. doi: 10.1021/jp410567c. PubMed PMID: 24689966.

100. Haynes WM. CRC handbook of chemistry and physics. 94th editi ed. Boca Raton, Florida: CRC Press; 2013. 2672 pages p.

101. Record MT, Lohman TM. A semiempirical extension of polyelectrolyte theory to the treatment of oligoelectrolytes: Application to oligonucleotide helix-coil transitions. Biopolymers. 1978;17(1):159-66. doi: 10.1002/bip.1978.360170112.

102. Persson F, Soderhjelm P, Halle B. The spatial range of protein hydration. J Chem Phys. 2018;148(21):215104. Epub 2018/06/10. doi: 10.1063/1.5031005. PubMed PMID: 29884061.

103. Yao XQ, Momin M, Hamelberg D. Establishing a Framework of Using Residue-Residue Interactions in Protein Difference Network Analysis. J Chem Inf Model. 2019;59(7):3222-8. Epub 2019/07/04. doi: 10.1021/acs.jcim.9b00320. PubMed PMID: 31268315.

104. Luzar A, Chandler D. Hydrogen-bond kinetics in liquid water. Nature. 1996;379(6560):55-7. doi: 10.1038/379055a0.

105. Courtenay ES, Capp MW, Anderson CF, Record MT. Vapor Pressure Osmometry Studies of Osmolyte−Protein Interactions: Implications for the Action of

Osmoprotectants in Vivo and for the Interpretation of “Osmotic Stress” Experiments in Vitro. Biochemistry. 2000;39(15):4455-71. doi: 10.1021/bi992887l.

106. Shimizu S. Estimation of excess solvation numbers of water and cosolvents from preferential interaction and volumetric experiments. J Chem Phys. 2004;120(10):4989-90. Epub 2004/07/23. doi: 10.1063/1.1646373. PubMed PMID: 15267361.

107. Bano M, Marek J. How thick is the layer of thermal volume surrounding the protein? Biophys Chem. 2006;120(1):44-54. Epub 2005/10/26. doi: 10.1016/j.bpc.2005.09.024. PubMed PMID: 16242836.

108. Kirkwood JG, Buff FP. The Statistical Mechanical Theory of Solutions. I. The Journal of Chemical Physics. 1951;19(6):774-7. doi: 10.1063/1.1748352.

109. Shimizu S. Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. Proc Natl Acad Sci U S A. 2004;101(5):1195-9. Epub 2004/01/21. doi: 10.1073/pnas.0305836101. PubMed PMID: 14732698; PMCID: PMC337029.

110. Ben-Naim A. Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system. The Journal of Chemical Physics. 1977;67(11):4884-90. doi: 10.1063/1.434669.

111. Smith PE. On the Kirkwood-Buff inversion procedure. J Chem Phys. 2008;129(12):124509. Epub 2008/12/03. doi: 10.1063/1.2982171. PubMed PMID: 19045038; PMCID: PMC2671658.

112. Bouhlel MA, Lambert M, David-Cordonnier MH. Targeting transcription factor binding to DNA by competition using DNA binders as an approach for controlling gene expression. Curr Top Med Chem. 2015. PubMed PMID: 25866275.

113. Tataurov AV, You Y, Owczarzy R. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. Biophys Chem. 2008;133(1-3):66-70. doi: 10.1016/j.bpc.2007.12.004.

114. Erlitzki N, Farahat AA, Kumar A, Boykin DW, Poon GMK. DNA recognition by linear indole-biphenyl DNA minor groove ligands. Biophys Chem. 2019;245:6-16. Epub 2018/12/05. doi: 10.1016/j.bpc.2018.11.004. PubMed PMID: 30513446.

115. Fringuelli F, Marino G, Taticchi A. Tellurophene and Related Compounds. In: Katritzky AR, Boulton AJ, editors. Advances in Heterocyclic Chemistry Volume 21: Academic Press; 1977. p. 119-73.

116. Brown RD, Burden FR, Godfrey PD. The microwave spectrum of selenophene. Journal of Molecular Spectroscopy. 1968;25(4):415-21. doi: 10.1016/s0022-2852(68)80012-8.

117. Bak B, Christensen D, Hansen-Nygaard L, Rastrup-Andersen J. The structure of thiophene. Journal of Molecular Spectroscopy. 1961;7(1-6):58-63. doi: 10.1016/0022-2852(61)90341-1.

118. Torsello M, Pimenta AC, Wolters LP, Moreira IS, Orian L, Polimeno A. General AMBER Force Field Parameters for Diphenyl Diselenides and Diphenyl Ditellurides. J Phys Chem A. 2016;120(25):4389-400. Epub 2016/06/09. doi: 10.1021/acs.jpca.6b02250. PubMed PMID: 27267296.

119. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, SantaLucia J. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. J Chem Theory Comput. 2007;3(4):1464-75. Epub 2007/07/01. doi: 10.1021/ct600329w. PubMed PMID: 26633217.

120. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic acids research. 2003;31(17):5108-21. PubMed PMID: 12930962; PMCID: PMC212791.

121. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126(1):014101. Epub 2007/01/11. doi: 10.1063/1.2408420. PubMed PMID: 17212484.

122. Kumar R, Grubmuller H. do_x3dna: a tool to analyze structural fluctuations of dsDNA or dsRNA from molecular dynamics simulations. Bioinformatics. 2015;31(15):2583-5. Epub 2015/04/04. doi: 10.1093/bioinformatics/btv190. PubMed PMID: 25838463.

123. Voss NR, Gerstein M. 3V: cavity, channel and cleft volume calculator and extractor. Nucleic acids research. 2010;38(suppl_2):W555-W62. doi: 10.1093/nar/gkq395.

124. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science. 2016;2:e55. doi: 10.7717/peerj-cs.55. PubMed PMID: WOS:000437456500001.

125. Millero FJ. The apparent and partial molal volume of aqueous sodium chloride solutions at various temperatures. J Phys Chem. 1970;74(2):356-62. doi: 10.1021/j100697a022.

126. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q Rev Biophys. 1978;11(2):179-246. Epub 1978/05/01. doi: 10.1017/s0033583500002031. PubMed PMID: 353876.

127. Bleam ML, Anderson CF, Record MT. Relative binding affinities of monovalent cations for double-stranded DNA. Proc Natl Acad Sci U S A. 1980;77(6):3085-9. Epub 1980/06/01. PubMed PMID: 16592827; PMCID: PMC349556.

128. Thomas S, Martinez LL, Westenberger SJ, Sturm NR. A population study of the minicircles in Trypanosoma cruzi: predicting guide RNAs in the absence of empirical RNA editing. BMC Genomics. 2007;8:133. Epub 2007/05/26. doi: 10.1186/1471-2164-8-133. PubMed PMID: 17524149; PMCID: PMC1892023.

129. Bailly C, Tardy C, Wang L, Armitage B, Hopkins K, Kumar A, Schuster GB, Boykin DW, Wilson WD. Recognition of ATGA sequences by the unfused aromatic dication DB293 forming stacked dimers in the DNA minor groove. Biochemistry. 2001;40(33):9770-9. Epub 2001/08/15. doi: 10.1021/bi0108453. PubMed PMID: 11502170.