Biophysical Properties of DNA Minor Groove Binding by Heterocyclic Cations of Varying Structures

Noa Erlitzki
Georgia State University, nerlitzki1@student.gsu.edu

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_theses

Recommended Citation

This Thesis is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
ABSTRACT

Small heterocyclic cations that bind to DNA are interesting systems to study due to their structural diversity, pharmaceutical potential, and characteristic target recognition patterns. Clinically, such compounds offer an attractive therapeutic approach as inhibitors of protein-DNA interactions implicated in disease. Due to their typical intrinsic fluorescence, these compounds also have potential as convenient biotechnological probes for studying DNA. Finally, from a biophysics perspective, an intricate understanding of the factors driving DNA binding by these compounds can extend our understanding of DNA targeting more broadly. In this thesis, the electrostatics and hydration properties of DNA binding by eight of these heterocyclic cations in complex with various DNA sequences are investigated.

INDEX WORDS: DNA minor groove, preferential hydration, polyelectrolyte theory, heterocyclic cations, fluorescence polarization, small compounds
BIOPHYSICAL PROPERTIES OF DNA MINOR GROOVE BINDING BY HETEROCYCLIC CATIONS OF VARYING STRUCTURES

by

NOA ERLITZKI

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

2019
BIOPHYSICAL PROPERTIES OF DNA MINOR GROOVE BINDING BY HETEROCYCLIC
CATIONS OF VARYING STRUCTURES

by

NOA ERLITZKI

Committee Chair: Gregory Poon

Committee: David Wilson

Dabney Dixon

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2019
DEDICATION

To those who told me I could do it,
who were there for me when I was too preoccupied to be there for myself,
and who make my life beautiful and fill it with meaning.

To my grandfathers, Sevek and Dani, who will never know of this
but who would be so proud.

To the one who showed me the kind of person I don’t want to be –
thank you for teaching me to look at the silver lining.

And to me,
because I owe it to myself
to celebrate.

Je suis prest.

— Diana Gabaldon (Outlander)
ACKNOWLEDGEMENTS

A great deal of gratitude goes to my parents who, as with all my other pursuits in life, supported me fully and unconditionally through the highs and lows of this adventure. I am so incredibly lucky. To my brother – the last several years have seen us grow up from bickering siblings to the closest of friends. I cherish our bond and could not have made it this far without your humor and support.

This thesis would certainly not be what it is without the influence of my research advisor, Dr. Gregory Poon. The past four years, starting as an undergraduate and then as a graduate student, have been quite the adventure! My time in your lab has brought many opportunities that helped me grow as a scientist and has seen me through adversities that helped me grow as a person. I thank you for your part in both. I would also like to thank my committee members, Drs. David Wilson and Dabney Dixon, whose time, guidance, and scientific insight were very valuable in the completion of this thesis. In addition, I want to express my thanks toward Drs. David Boykin, Abdelbasset Farahat, Ananya Paul, and Paul Guo for many helpful discussions.

I could not have done this without the support of those I look up to as mentors in both science and life: Samer Gozem, Dabney Dixon, and Donald Hamelberg. I owe a great deal of thanks to the entire Department of Chemistry – I am so proud to be part of this academic family. Thank you to my lab mates for your friendship through the years; it has been wonderful to pursue scientific knowledge alongside you all and I value the many good times we’ve had together both in and out of the lab! Thank you also to my departmental colleagues and good friends Rebecca, Kurt, and Jina, who I can always count on for friendship and advice. A special shout out to my BFF Hayley: thank you for being there for me to celebrate the good times, lift my spirits during the bad times, and make me laugh (until my muscles ache!) at any time.
Finally, a big thank you goes to the funding sources that supported the research in this thesis. In particular, I would like to thank the Harry P. Hopkins, Jr. Scholarship in Physical Chemistry for financial support during my graduate studies and the Barry Goldwater Scholarship for financial support during my undergraduate studies. Last but not least, thank you to the Department of Chemistry for providing graduate research funding through a research assistantship and a teaching assistantship.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... V

LIST OF TABLES ........................................................................................................................ XI

LIST OF FIGURES ....................................................................................................................... XII

LIST OF EQUATIONS .................................................................................................................. XIV

1 GENERAL INTRODUCTION ..................................................................................................... 1

1.1 DNA Minor Groove Ligands ................................................................................................. 1

1.1.1 A short history of minor groove binding compounds ..................................................... 1

1.1.2 Minor groove as target, part I: DNA structural aspects ................................................. 2

1.1.3 Minor groove as target, part II: Compound structural considerations ................. 4

1.2 Polyelectrolyte Theory and the Ionic Environment of DNA .......................................... 6

1.2.1 Counterion condensation: Manning’s limiting law approach ..................................... 6

1.2.2 The effects of high salt on DNA ................................................................................... 11

1.3 Thermodynamics of Compound/DNA Interactions ......................................................... 15

1.3.1 Monitoring binding by fluorescence anisotropy ........................................................... 16

1.3.2 Probing molecular hydration via osmotic stress ........................................................... 17

1.3.3 Salt as co-solute: coupling hydration effects to counterion release ............................ 21

1.4 Research Aim ....................................................................................................................... 22

1.5 Outline of the Thesis ........................................................................................................... 22
2 DNA RECOGNITION BY LINEAR INDOLE-BIPHENYL DNA MINOR GROOVE LIGANDS

2.1 Abstract ................................................................. 25

2.2 Introduction ............................................................. 25

2.3 Materials and Methods ................................................ 28

2.3.1 Compounds and DNA .............................................. 28

2.3.2 Binding experiments ............................................... 28

2.3.3 Steady-state fluorescence quenching ........................... 29

2.3.4 Data analysis .......................................................... 29

2.4 Results .................................................................... 30

2.4.1 Linear dications exhibit salt-sensitive and salt-insensitive binding modes 32

2.4.2 The binding modes of DB1804 are DNA sequence-specific ...................... 36

2.4.3 The binding modes of DB1804 are sensitive to inhibition by netropsin ..... 38

2.4.4 Preferential interactions with co-solutes modify DB1804 binding modes.. 39

2.4.5 Linear dications self-associate in the absence of DNA ......................... 41

2.5 Discussion ................................................................ 44

2.5.1 Structural determinants of binding mode selection by linear indole-biphényl cations ................................................. 45

2.5.2 Functional implications of multiple binding modes for dications as transcriptional inhibitors ........................................... 49
2.6 Conclusion......................................................................................................................... 50

2.7 Supplemental Data ............................................................................................................ 50

2.7.1 Supplemental methods ................................................................................................. 50

2.7.2 Supplemental table ......................................................................................................... 53

2.7.3 Supplemental figures ..................................................................................................... 54

3 STRUCTURE-BINDING RELATIONSHIPS OF DNA RECOGNITION BY
INDOLE-BIPHENYL MONOCATIONS ..................................................................................... 56

3.1 Abstract ............................................................................................................................... 56

3.2 Introduction ......................................................................................................................... 56

3.3 Materials and Methods ..................................................................................................... 58

3.3.1 Compounds and DNA .................................................................................................... 58

3.3.2 Steady-state fluorescence polarization binding assays ................................................. 59

3.3.3 Computational methods .................................................................................................. 59

3.4 Results .................................................................................................................................. 60

3.4.1 Hydration changes accompanying DNA binding by monocations ......................... 60

3.4.2 Thermodynamics of DNA binding by linear monocations .......................................... 62

3.5 Discussion ............................................................................................................................ 68

3.5.1 Structure-binding relationships of indole-biphenyl monocations .............................. 68

3.5.2 Intricacies of structure effects on DNA minor groove binding ................................. 71

3.5.3 A role for water beyond preferential hydration ......................................................... 72
4 SEQUENCE DEPENDENCE OF HYDRATION IN DNA BINDING BY AN
ISOHELICAL DIAMIDINE.......................................................................................... 74

4.1 Abstract .............................................................................................................. 74
4.2 Introduction ........................................................................................................ 74
4.3 Materials and Methods ...................................................................................... 76

4.3.1 Compound and DNA ..................................................................................... 76
4.3.2 Fluorescence polarization titrations ................................................................. 76

4.4 Results ............................................................................................................... 76

4.4.1 Probing preferential hydration in DNA site recognition by osmotic stress .... 77

4.5 Discussion ......................................................................................................... 80

4.5.1 Binding preferences are independent of preferential hydration properties .... 82

4.6 Conclusion ........................................................................................................ 83

5 OVERALL CONCLUSION .................................................................................... 84

5.1 Future Directions ............................................................................................... 84

REFERENCES ......................................................................................................... 87
LIST OF TABLES

Table 2.1 Goodness of fit of Hill analysis of DNA binding by linear indole-biphenyl cations. ................................................................. 53

Table 3.1 List of compounds investigated. ................................................................. 58

Table 3.2 Comparison of thermodynamic parameters. ............................................ 64
LIST OF FIGURES

Figure 1.1 Early DNA minor groove binding compounds. ............................................ 2
Figure 1.2 Structural parameters of B-DNA. ............................................................... 3
Figure 1.3 Properties of the DNA minor groove.......................................................... 4
Figure 1.4 Diversity in DNA minor groove binding compounds.................................... 5
Figure 1.5 Increasing divergence of the charge neutralization fraction at high salt as predicted by counterion condensation. ................................................................. 10
Figure 1.6 Differential effects of high salt concentrations on DNA duplex stability. ........ 12
Figure 1.7 Basic principle of fluorescence anisotropy binding assays. ............................ 17
Figure 1.8 Coupling of preferential hydration to electrostatics. ...................................... 22
Figure 2.1 Minor-groove binding ligands and target DNA used in this study. ..................... 27
Figure 2.2 NaCl unmasks two distinct binding modes for the ditetrahydropyrimidine DB1804. ........................................................................................................................................ 31
Figure 2.3 Salt dependence of DNA binding by indole-biphenyl mono- and dications....... 33
Figure 2.4 Stoichiometry of A₂T₂-bound complexes of linear cations. .............................. 34
Figure 2.5 DNA recognition by DB1804 is mechanistically described by an equilibrium distribution of 2:1 and 1:1 complexes. ................................................................. 36
Figure 2.6 The 2:1 and 1:1 binding modes exhibited by DB1804 are sequence-specific. ...... 37
Figure 2.7 Inhibition of A₂T₂-bound DB1804 by netropsin. ........................................... 38
Figure 2.8 The cryo-protectant 2-methyl-2,4-pentainediol enforces 1:1 binding by linear dications. ..................................................................................................................................... 40
Figure 2.9 Steady-state fluorescence quenching of linear indole-biphenyl cations in the absence of DNA. ............................................................................................................ 43
Figure 2.10 A model for dimer formation by linear indole-biphenyl dications. ......................... 48

Figure 2.11 Size-exclusion chromatography of A₂T₂ in the absence and presence of DB1804. ........................................................................................................................................ 54

Figure 2.12 Effect of DMSO and nicotinamide on DNA binding by DB1804. .......................... 55

Figure 3.1 Heterogeneity in DNA binding by linear monocations. ........................................ 61

Figure 3.2 Thermodynamic properties of biphenyl-indole monocations in complex with DNA................................................................................................................................................ 64

Figure 3.3 Total, polar, and hydrophobic surface areas of the compounds in their ground-
state equilibrium geometries. ........................................................................................................ 65

Figure 3.4 The relative number of conformers available to each compound in the unbound state. .................................................................................................................................................... 66

Figure 3.5 DB2783 is bulkier than DB2782. ........................................................................... 70

Figure 3.6 Comparison of binding affinity rank orders for similar compounds in the literature. ................................................................................................................................................... 71

Figure 4.1 Compound and DNA sequences used in this study. ............................................. 75

Figure 4.2 Perturbation with salt shows a net release of preferential hydration for the DB1976/DNA complexes............................................................................................................................................ 78

Figure 4.3 The effect of ionic strength on binding affinity of DB1976 to A₃T₃, A₂T₂, and A₂CGT₂. ....................................................................................................................................................... 79

Figure 4.4 Structural properties of the DNA oligonucleotides. ............................................. 81
LIST OF EQUATIONS

\[ \bar{g}_{el} = -(1-Z\theta)^2 \xi \ln (1-e^{-xb}) \] (1.1)

\[ \bar{g}_{mix} = \theta \ln \left(10^3 \theta V_p^{-1}/c_Z\right) \] (1.2)

\[ \frac{d}{d\theta} \left( \bar{g}_{el} + \bar{g}_{mix} \right) = \frac{d}{d\theta} \left[ \theta \ln \left(10^3 \theta V_p^{-1}/c_Z\right) - (1-Z\theta)^2 \xi \ln (1-e^{-xb}) \right] = 0 \]

\[ 1 + 2Z\xi (1-Z\theta) \ln (1-e^{xb}) + \ln \left(10^3/c_Z\right) = 0 \] (1.3)

\[ 1 + \ln \left(10^3 \theta V_p^{-1}/c_Z\right) = -2Z\xi (1-Z\theta) \ln (1-e^{-xb}) \] (1.4)

\[ \theta = 1 - \xi^{-1} \] (1.5)

\[ V_p = 41.1(\xi - 1)b^3 \] (1.6)

\[ c_{loc} = 10^3 \theta V_p^{-1} = 24.3(\xi b^3)^{-1} \] (1.7)

\[ \psi = \psi_e + \psi_z = (1-\xi^{-1}) + (2\xi)^{-1} = 1 - (2\xi)^{-1} \] (1.8)

\[ \psi_N = \psi - \frac{5.64}{N} \left(1-\xi^{-1}\right) \] (1.9)

\[ K_D = \frac{[DB][DNA]}{[DB:DNA]} \] (1.10)

\[ \langle r(\tau) \rangle = \frac{r_0}{1 + \tau/\phi} \] (1.11)

\[ \langle r \rangle = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}} \] (1.12)

\[ \left( \frac{\partial \log K_{obs}}{\partial \log a_L} \right)_{p,T,m_db,m_{DNA}} = V_L^{DB-DNA} - V_L^{DNA} - V_L^{DB} = \Delta V_L \] (1.13)

\[ RT \ln a_{w,L} = -\bar{V} \pi = -RT \frac{Osm}{55.5} \] (1.14)
\begin{align}
\frac{\partial \log K_D}{\partial \log Osm} &= -\frac{1}{55.5} \left( \frac{\partial \log K_D}{\partial \log a_{w,t}} \right) = -\left( \frac{\Delta v_w}{55.5} \right) \tag{1.15} \\
\frac{\partial \log K_D}{\partial \log a_\pm} &= -Z\psi + k = \Delta v_\pm \tag{1.16} \\
\frac{\partial \log K_D}{\partial \log a_\pm} &= \Delta v_\pm - \frac{2m_\pm}{55.5 \ln 10} \Delta v_w \tag{1.17} \\
\langle r \rangle &= F_b \left( \sum_{i=1}^{n} (\langle r_i \rangle - \langle r_0 \rangle) + \langle r_0 \rangle \right) = F_b \sum_{i=1}^{n} \Delta r_i + \langle r_0 \rangle \tag{2.1} \\
-\frac{\partial \log K_D}{\partial \log a_\pm} &= n_\pm = \psi Z = \left( \psi_w - \frac{2.53}{N} \right) Z \tag{2.2} \\
F_b &= f \frac{c_{n_{H,1}}}{K_{D,1} n_{H,1} + c_{n_{H,1}}} + (1 - f) \frac{c_{n_{H,2}}}{K_{D,2} n_{H,2} + c_{n_{H,2}}} \tag{2.3} \\
\frac{F_0}{F} &= \frac{\tau_0}{\tau} = 1 + K[Q] \tag{2.4} \\
\frac{1}{r} &= \frac{1}{r_0} + \frac{RT}{\eta V} \tag{2.5} \\
\frac{1}{r} &= \frac{1}{r_0} + \frac{RT}{\eta V \left( 1 + K_{SV}[Q] \right)} \tag{2.6} \\
K_1 &= \frac{[DB][DNA]}{[DB:DNA]} \tag{2.7} \\
K_2 &= \frac{[DB:DNA][DB]}{[DB_2:DNA]} = \omega K_1 \tag{2.7} \\
\text{[DB]}_i &= [DB] + [DB:DNA] + 2[DB_2:DNA] \tag{2.8} \\
\text{[DNA]}_i &= [DNA] + [DB:DNA] + [DB_2:DNA] 
\end{align}
\[ 0 = \varphi_0 + \varphi_1[DB] + \varphi_2[DB]^2 + \varphi_3[DB]^3 \]

\[
\begin{align*}
\varphi_0 &= K_1 K_2[DB]_t \\
\varphi_1 &= -K_1 K_2 - K_2[DNA]_t + K_2[DB]_t \\
\varphi_2 &= -2[DNA]_t - K_2 + [DB]_t \\
\varphi_3 &= -1
\end{align*}
\]  

(2.9)

\[
F_{b,1:1} = \frac{[DB:DNA]}{[DB]_t}
\]

(2.10)

\[
F_{b,2:1} = \frac{[DB_2:DNA]}{[DB]_t}
\]

\[
\langle r \rangle ([DNA]_t) = \sum_{i=1}^{n} \left[ \frac{[DB]_t}{[DB]_t} \left( \langle r_i \rangle - \langle r_0 \rangle \right) \right] + \langle r_0 \rangle
\]

(3.1)

\[- \left( \frac{\partial \log K_D}{\partial \log a_z} \right) = \Delta n_z - \frac{2m_z}{55.5 \ln 10} \Delta n_w
\]

(3.2)

\[- \log K_D = \log K_0 + \Delta n_z \log a_z - \Delta n_w \frac{2m_z}{55.5 \ln 10}
\]

(3.3)

\[
\Delta G = \Delta H - T \Delta S = \Delta H_{\text{bond}} + \Delta H_{\text{vdW}} + \Delta H_{\text{el}} + \Delta H_{\text{CC}} + \Delta H_{\text{SA}} - T \left( \Delta S_w + \Delta S_{\text{ion}} + \Delta S_{\text{conf}} + \Delta S_{\text{trat}} \right)
\]

(3.4)

\[
\Delta S_{\text{ion}} = -ZR \ln \left[ \text{Na}^+ \right]
\]

(3.5)

\[
S_{\text{conf}} = k_B \ln \Omega_{\text{conf}}
\]

(3.6)

\[- \left( \frac{\partial \log K_D}{\partial \log a_z} \right) = \Delta n_z - \frac{2m_z}{55.5 \ln 10} \Delta n_w
\]

(4.1)

\[- \log K_D = \log K_0 + \Delta n_z \log a_z - \Delta n_w \frac{2m_z}{55.5 \ln 10}
\]

(4.2)
1 GENERAL INTRODUCTION

1.1 DNA Minor Groove Ligands

In biological systems, deoxyribonucleic acid (DNA) is more often than not found in complex with proteins. In the nucleosome, transcriptionally-inactive DNA is wound around histone octamers, the positively-charged residues of which contact the DNA minor groove. Most other proteins, such as transcription factors (with sizes in the kDa range), bind primarily to the major groove of DNA with the occasional minor groove contact. By contrast, small molecules (10^2 Da) that bind non-covalently to DNA rarely do so in the major groove, instead intercalating themselves between DNA bases or else binding to the minor groove. Due to their inherently greater sequence specificity and tendency to be less perturbing to DNA structure than intercalating agents, minor groove binders are interesting systems to study that have numerous applications in biotechnology and medicine.

1.1.1 A short history of minor groove binding compounds

Decades prior to our conception of DNA as a double-helix with major and minor grooves, the scientific community had recognized and taken advantage of the anti-parasitic and antibiotic properties of small molecules such as synthalin, distamycin, and pentamidine (Figure 1.1) [1-5]. Following Watson and Crick’s report of the structure of DNA [6, 7], these compounds’ tendency to target the DNA minor groove was established, and their mode of action as drugs was attributed to this general property [8-17]. In the years since, thousands of analogues and their derivatives have been synthesized by different groups, many of which have shown potential as pharmaceutical agents [18-22].
Figure 1.1 Early DNA minor groove binding compounds. Distamycin is a natural product that was discovered in the late 1950s, while synthalin and pentamidine are synthetic compounds from the 1920s and 1930s.

1.1.2 Minor groove as target, part I: DNA structural aspects

 Canonical B-DNA is a right-handed double-helix 20 Å in diameter with 10 Watson-Crick base pairs (A=T, G=C) per helical turn, each of which is nearly perpendicular to the helical axis and has a twist angle of 36.0°, giving an axial rise per base pair of 3.4 Å. Consequent to the hydrogen bonding patterns in Watson-Crick base pairing, wherein the two bases’ deoxyribose groups are on the same side of the base pair, is the formation of a wider major groove and a narrower minor groove having widths of 11.6 Å and 6.0 Å, respectively (Figure 1.2) [23].

 DNA-targeting compounds, such as the heterocyclic diamidines, bind DNA by inserting themselves deep into the minor groove. Most such compounds preferentially target A/T-rich regions of the minor groove, though there are exceptions (this will be discussed further in the next section). Preference for the minor groove is partly driven by its groove width which, compared to that of the major groove, provides ideal distances for hydrogen-bonding interactions between
compound and base pair edges as well as for van der Waals interactions with the groove walls [24]. As minor groove width is sequence-dependent (Figure 1.3A) [25, 26], it comes as no surprise that the narrower A/T-rich minor groove regions are a particularly favorable target for many of these compounds. An additional sequence preference is conferred by a network of stable water molecules buried deep in the minor groove. In A/T-rich regions, these ordered water molecules (known as the spine of hydration) penetrate the groove more deeply than in G/C-rich regions (Figure 1.3B) [27]. The displacement of these waters provides an addition favorable entropic contribution to the binding free energy of DNA minor groove binders (MGBs). Finally, steric hindrance from the G-NH₂ group (which is not present on adenine; see Figure 1.2B) in the minor groove poses a barrier to binding that contributes to the broad A/T sequence specificity seen with many MGBs.

![Diagram of DNA structural parameters](image)

**Figure 1.2 Structural parameters of B-DNA.**

A. Cartoon and space-filling representation of the Dickerson dodecamer (PDB: 1BNA) showing the helical parameters pitch and rise as well as the major (dashed curve) and minor (dotted curve) grooves. B. The wide major groove (dashed curve) and the narrow minor groove (dotted curve) are a consequence of the orientation of the bases relative to their linked deoxyribose (sugar) groups in the Watson-Crick arrangement.
**Figure 1.3 Properties of the DNA minor groove.**
*A, Plot of predicted minor groove widths for various DNA sequences showing the narrower width, in general, of AT-bp regions over GC-bp regions. Calculations were done using the Rhos Lab DNASHAPE Tool. B, The spine of hydration in the minor groove of the 5'-AATT-3' region of the Dickerson dodecamer (PDB: 455D).*

### 1.1.3 Minor groove as target, part II: Compound structural considerations

The structural diversity of MGBs is remarkable: cationic compounds of different lengths and bulkiness, harboring fused ring systems or flexible linkers, and having curvature from that exceeding the DNA minor groove to none at all have all been demonstrated to bind to the DNA minor groove (Figure 1.4). This structural variability of MGBs paints a seemingly simple picture wherein electrostatic forces between the cationic compounds and the negatively-charged DNA phosphate backbone are the primary drive for compound/DNA binding. At the same time, even small variations to a compound’s structure can drastically alter its DNA recognition mechanisms. For example, while most MGBs bind preferentially to A/T-rich sequences, DB2277 and DB2528 bind to mixed sequence (e.g., AAAGTTT) DNA [28, 29]. The asymmetric furamidine derivative DB293 also recognizes G/C-containing sequences, but does so as a dimer [30], while the natural product distamycin and the cyanine dye DiSC2(5) dimerize as well but in the minor groove of a
5’-AAATT-3’ site or alternating A/T sequences, respectively [31, 32]. Meanwhile, the A/T-specific linear compound DB921 utilizes an interfacial water molecule to mediate binding to the curved minor groove of DNA [33]. While all these compounds bind to DNA in the minor groove, the details of their structures reveal corresponding subtleties in how they bind DNA. Since generalized electrostatic forces alone cannot account for such heterogeneity in binding, other factors must also be at play. Progress in this area has identified a number of such factors, including specific water-mediated interactions (as with DB921), the thiophene sigma-hole interaction (positive electrostatic potential acting as a hydrogen-bond acceptor to facilitate interactions with the G-NH$_2$ in mixed base pair sequences) [34], and the formation of dimers [35].

![Figure 1.4 Diversity in DNA minor groove binding compounds](image)

**Figure 1.4 Diversity in DNA minor groove binding compounds.**

DNA minor groove recognition has been demonstrated by compounds having diverse structures (differing in features such as degree of isohelicity, charge number, flexibility, etc.), such as the examples shown here. At the same time, compounds with apparently minor structural variations may exhibit very different binding patterns (e.g., sequence specificity, stoichiometry of binding, involvement of structural waters).
1.2 Polyelectrolyte Theory and the Ionic Environment of DNA

The polyelectrolyte nature of DNA – that is, its occurrence in aqueous solution as a polymer of repeating charge units – has long been called upon to explain empirical observations of the behavior of DNA in solution (such as the winding-together of two like-charged phosphate backbone strands; its migration toward one end of an electric field; and its susceptibility to precipitate out of solution at high ionic strengths in the presence of organic co-solvents). The theoretical basis for this behavior is found in multiple models. Discussed herein is Manning’s well-known limiting law treatment of counterion condensation, which we apply in our analysis of the data in Chapters 2-4. An additional discussion on the effects of high salt on DNA in the context of the oligonucleotides used in these chapters follows.

1.2.1 Counterion condensation: Manning’s limiting law approach

A well-established description of the behaviors of polyelectrolytes is given by Manning’s counterion condensation model (referred to henceforth as CC), which offers an analytical solution to polyelectrolyte theory. Counterion condensation describes the phenomenon whereby a DNA molecule (or other polyelectrolyte) in aqueous solution is surrounded by a “cloud” of cations (or counterions) that are territorially bound (as opposed to site-specifically bound) to the DNA (or other polyelectrolyte). These counterions were directly observed in a series of $^{23}$Na NMR studies in the 1970s and ‘80s as well as in a later gel electrophoresis study [36-40]. More recently, in an elegant manipulation of the electrostatic environment of DNA, the extreme stability of the layer of condensed counterions most proximal to the DNA has been demonstrated [41, 42]. Experimental evidence notwithstanding, counterion condensation itself, and the consequent charge neutralization of the polyelectrolyte, is a simple physical fact resulting from the tendency of two like-charged objects (such as two negatively-charged DNA strands) to repel one another.
As will be shown, CC predicts the extent of charge neutralization, the concentration of condensed counterions in the proximity of the DNA, and the distance from the DNA surface within which the condensed counterions reside. Two key insights stemming from the CC regime are of particular importance: (1) Even at the fullest extent of counterion condensation, the DNA molecule retains a characteristic fraction of its original negative charge. (2) At bulk salt concentrations approaching zero, a local concentration of condensed counterions is maintained such that the DNA still experiences the characteristic fractional charge neutralization, which is equal in magnitude to the fractional charge neutralization of DNA in excess salt. The limiting law approach used by Manning [43-50] to arrive at these conclusions is summarized below for the case of a small, monovalent ion (e.g., Na\(^+\)) as the counterion species.

A polyelectrolyte with an average axial charge spacing \(b\) has a reduced linear charge density \(\xi = b_{B}/b\), where \(b_{B}\) is the Bjerrum length given by \(b_{B} = e^{2}/\varepsilon k_{B}T\) (\(e\) is the electronic charge, \(\varepsilon\) is the dielectric constant). In the presence of counterions with valence \(Z\), a critical value \(\xi_{\text{crit}} = |Z|^{-1}\) exists. For the case of a monovalent ion (\(Z = 1\)), \(\xi_{\text{crit}} = 1\). When \(\xi > \xi_{\text{crit}}\), the resulting thermodynamic instability of the system acts as the driving force for counterion condensation, which occurs to the extent required to bring the charge density of the polyelectrolyte-counterion “complex” to \(\xi_{\text{crit}}\). In other words, counterion condensation will occur if \(Z\xi > 1\). For water at room temperature, \(b_{B} = 7.15\) Å. Therefore, under these conditions, counterion condensation will occur for polyelectrolytes having charge density \(b \leq 7.15\) Å, which is to say nearly all polyelectrolytes and certainly B-DNA (\(b = 1.7\) Å) [46].

In quantifying the extent of charge neutralization, Manning showed [51] that the fraction of polyelectrolyte charge neutralized by counterion condensation is given by \(Z\theta = 1 - (Z\xi)^{-1}\).
which he derived as follows. First, we define the total free energy of the DNA polyelectrolyte system. Our interest here is only in the contributions to the free energy that depend on the extent of counterion association by condensation ($\theta$). Thus, the total free energy is the sum of an electrostatic component $\bar{g}_{el}$ (the energy associated with attractive interactions between counterions and DNA phosphates and with repulsive interactions between the phosphates on one DNA strand and those on the other) and an entropic contribution $\bar{g}_{mix}$ (the energy of mixing of free and bound counterions and solvent) [51]:

$$\bar{g}_{el} = -(1 - Z\theta)^2 \xi \ln (1 - e^{-xb})$$

$$\bar{g}_{mix} = \theta \ln \left(10^3 \theta V_p^{-1} / c_Z\right)$$

where $\kappa$ is the Debye-Hückel screening parameter, $V_p$ is the volume (in units of mL/mole phosphate) surrounding the DNA within which counterions are considered to be “condensed,” and $c_Z$ is the bulk concentration (in molarity) of the counterion $M^{Z+}$. We make the assumptions that condensed counterions are free to move within the volume $V_p$ and that the interactions of solvent molecules with condensed versus free counterions are indistinguishable. Under these assumptions, $\bar{g}_{el}$ and $\bar{g}_{mix}$ are the only contributions dependent on $\theta$ to the free energy of the system. The free energy minimum is given by:

$$d\left(\bar{g}_{el} + \bar{g}_{mix}\right) / d\theta \left[ \theta \ln \left(10^3 \theta V_p^{-1} / c_Z\right) - (1 - Z\theta)^2 \xi \ln (1 - e^{-xb}) \right] = 0.$$  \hspace{1cm} (1.3)

$$1 + 2Z\xi (1 - Z\theta) \ln (1 - e^{-xb}) + \ln \left(10^3 / c_Z\right) = 0$$

Therefore, at the free energy minimum, the following relation holds:

$$1 + \ln \left(10^3 \theta V_p^{-1} / c_Z\right) = -2Z\xi (1 - Z\theta) \ln (1 - e^{-xb}).$$  \hspace{1cm} (1.4)

At the low salt limit ($c_Z \to 0$), expressions for $\theta$ and $V_p$ that satisfy Eq. (1.4) are obtained via numerical iteration, defined here for the case of a monovalent salt [51]:
\[ \theta = 1 - \xi^{-1} \]  
\[ V_p = 41.1 (\xi - 1) b^3. \]  

With the parameters \( \theta \) and \( V_p \) in hand, one can predict the concentration of counterions in the condensed layer, \( c_{loc} \) [51]:

\[ c_{loc} = 10^3 \theta V_p^{-1} = 24.3 (\xi b^3)^{-1} \]

where the coefficient \( 10^3 \) allows \( c_{loc} \) to be framed in units of molarity. For B-DNA, \( c_{loc} \approx 1.2 \text{ M Na}^+ \). Note that the characteristic extent of counterion condensation as formulated by Manning is dependent only on the polyelectrolyte’s structure (namely, the axial charge spacing \( b \)), the solvent conditions (dielectric constant, temperature), and the valency of the counterion. Accordingly, the concentration of condensed counterions is also dependent on only these parameters and, more significantly, independent on both bulk salt concentration as \( c_N \rightarrow 0 \) and the identity of the monovalent salt. The result is that B-DNA in aqueous solution \((b = 1.7 \text{ Å}, \xi = 4.2)\) undergoes a 76% reduction in charge \(( \theta = 1 - 0.24 = 0.76)\) due to counterion condensation by Na\(^+\) ions \((Z = 1)\). An equivalent interpretation that becomes useful in applications of polyelectrolyte theory to DNA-ligand binding is that, on average, 0.76 Na\(^+\) ions are territorially bound to DNA per phosphate.

CC is a limiting law theory in the sense that it was derived and is therefore valid for the limit of zero bulk salt concentration. Indeed, while the theory predicts a fairly constant, non-zero value for \( \theta \) at the limit \( c_Z \rightarrow 0 \) (a non-trivial result given that \( c_{loc} \) is accordingly ~1.2 M even when bulk salt is not in excess), inconsistencies in the theory at higher bulk salt concentrations [46] lead to deviations in \( \theta \) as \( c_Z \rightarrow \infty \). The validity of the limiting law approach is nevertheless supported by the excellent agreement of its predicted low-salt value for \( \theta \) with experimental values reported.
by Anderson, et al. [36] in a study that also demonstrated the marked stability of $\theta$ in bulk salt concentrations at least up to 0.5 M Na$^+$ (Figure 1.5).

\[
\psi = \psi_\epsilon + \psi_s = (1 - \xi^{-1}) + (2\xi)^{-1} = 1 - (2\xi)^{-1}
\]  \hspace{1cm} (1.8)

**Figure 1.5 Increasing divergence of the charge neutralization fraction at high salt as predicted by counterion condensation.**

The charge neutralization fraction $\theta$ is independent of bulk salt concentration. Open circles (○), theoretical values of $\theta$ computed as in Ref. 51 by numerical iteration of Eq. (1.3) as a function of $c_Z$. The limiting law nature of CC is evident by the relative invariance of $\theta$ at low salt; a breakdown of the theory at higher Na$^+$ concentrations coincides with deviations in $\theta$. Red dotted line (···), experimental value for the charge neutralization fraction ($\theta = 0.76$) determined by Anderson, et al. (Ref. 36) via $^{23}$Na NMR linewidth measurements. This experimental evidence suggests that $\theta$ remains constant at least up to ~0.5 M Na$^+$ despite the prediction by CC that $\theta \rightarrow 1$ at higher salt concentrations.

As is evident from the above discussion, ~24% of the polyelectrolyte charge remains unneutralized after counterion condensation. A second and more mobile population of ions, which by definition reside within a volume greater than $V_P$ from the DNA surface, screens the remaining polyelectrolyte charge fractions from one another. Record, et al. defined a thermodynamic ion association parameter $\psi$ that takes both counterion condensation and this additional screening effect into account to give the fraction of “thermodynamically bound” counterions [52]:

\[
\psi = \psi_\epsilon + \psi_s = (1 - \xi^{-1}) + (2\xi)^{-1} = 1 - (2\xi)^{-1}
\]  \hspace{1cm} (1.8)
where \( \psi_c \) is an alternative notation for \( \theta \) as used above (charge neutralization by counterion condensation; equivalently, extent of “binding” by condensed counterions) and \( \psi_s \) is a parameter defining charge neutralization by screening effects. Thus, for B-DNA, \( \psi = 0.88 \).

The calculation of \( \psi = 0.88 \) for B-DNA (i.e., \( \tilde{\zeta} = 4.2 \)) carries with it the underlying assumption that the DNA is of infinite length, which allows a complete disregard for any end-effects. While this assumption is valid for sufficiently long DNA polymers, where end-effects (i.e., the partial unraveling of the double helix) are negligible, it is an inaccurate representation of oligomeric DNA, for which base pairs involved in end-effects constitute a considerable fraction of the total length and the DNA can therefore no longer be characterized by \( \tilde{\zeta} = 4.2 \). To address this limitation, Record and Lohman [53] extended the original formulation of \( \psi \) to account for end effects, thereby enabling the application of CC to oligonucleotides of length \( N \) phosphates:

\[
\psi_N = \psi - \frac{5.64 \left(1-\tilde{\zeta}^{-1}\right)}{N} \frac{b}{b}
\]  

(1.9)

where \( \psi = 0.88 \), \( b = 1.7 \) Å, and \( \tilde{\zeta} = 4.2 \) for B-DNA.

### 1.2.2 The effects of high salt on DNA

In the chapters that follow, the polyelectrolyte nature of DNA will be drawn on to probe the hydration and electrostatics of DNA minor groove binding by small compounds. As will later become clear, this is achieved by perturbing the compound/DNA system with salt out to very high ionic strengths (ca. 3.5 M). Apropos of this, the effects of high monovalent salt concentrations on DNA structure and stability as reported in the literature are reviewed briefly in this section.

Perhaps the most well-studied property of DNA as a function of salt, especially at the high salt limit, is duplex stability as inferred from the \( T_m \) of the helix-coil transition. At typical experimental salt concentrations (< 1 M), the stabilizing effect of salt is well known: \( T_m \) increases
Figure 1.6 Differential effects of high salt concentrations on DNA duplex stability. 
A, The helix-coil transition of duplex DNA is in most cases stabilized (higher $T_m$) by high salt concentrations, as shown at left. High concentrations of certain salts (in particular, salts of $Cs^+$ and $ClO_4^-$) cause a decrease in $T_m$ as shown in the panel at right, especially for DNA of high GC content. Symbols represent different salts: LiCl (squares), NaCl (circles), TMA-Cl (up triangles), KCl (diamonds), Na$_2$SO$_4$ (left triangles), LizSO$_4$ (asterisks), Cs$_2$SO$_4$ (+), CsCl (X), NaClO$_4$ (hexagons), TEA-Cl (stars). Colors represent different DNA sequences: sea urchin (light purple), salmon sperm (light blue), E. coli (gray), calf thymus (black), poly dAT (purple), T4 (light green), M. lysodeikticus (green), Cl. perfringens (blue), M. luteus (orange), D. pneumoniae (red). Data is compiled from Refs. 54-60. Note: the $T_m$ for salmon sperm DNA in TMA (light blue, up triangle) is given for TMA in molal units (15.8 m). 

B, GC content of the DNA sequences from panel A.

linearly with the logarithm of monovalent salt concentration. A number of groups in the 1960s and 70s showed that, as should be expected, this linearity is not maintained at higher salt concentrations. Perhaps more surprising, however, is that high concentrations of certain salts seem to destabilize the DNA duplex in a cation- and anion-specific manner. Furthermore, GC-bps are disproportionately destabilized by high salt concentrations relative to AT-bps such that the well-
established direct relationship between T\textsubscript{m} and GC content at lower salt concentrations no longer holds [54-56]. **Figure 1.6A** shows these salt- and GC-content-dependent effects with data compiled from seven different studies spanning ten different types of salts and ten DNA sequences (the GC contents of which are shown in **Figure 1.6B**) [54-60]. Tetraethylammonium chloride (TEA-Cl) is strongly destabilizing to DNA of all GC contents for which there is data (31-72%), while tetramethylammonium chloride (TMA-Cl), Cs\textsubscript{2}SO\textsubscript{4}, and CsCl are only destabilizing to DNA having high (>50-70%) GC content. Importantly to Chapters 2-4 of this thesis, NaCl (as well as KCl, LiCl, Na\textsubscript{2}SO\textsubscript{4}, and Li\textsubscript{2}SO\textsubscript{4}) has no destabilizing effect even at concentrations as high as 4 M and regardless of GC content.

Alongside the thermodynamic stability of the helix-coil transition, much attention has been given to the effects of solvent environment, including salt concentrations, on the structural integrity of duplex DNA. A classic high-salt effect is the B-Z equilibrium transition observed for the alternating copolymers poly d(IC) and poly d(CG), wherein canonical right-handed B-DNA undergoes a major transition to its left-handed Z-DNA isoform in the presence of high salt concentrations [61-63]. Compared to B-DNA, the negatively-charged phosphate backbones of each strand in a Z-DNA duplex are closer together. Additionally, the left-handed Z-DNA helix places the nucleotides in alternating conformations of syn (which is adopted more readily by purines than by pyrimidines [64]) and anti [65]. It is therefore easy to see why this conformation is favored over the B isoform at high ionic strengths and also why it forms almost exclusively with alternating purine/pyrimidine DNA sequences such as poly d(CG) [66]. Not all alternating purine/pyrimidine sequences, however, adopt the Z-conformation as readily as the specific case of poly d(CG). The sequence poly d(CA)-poly d(TG), for example, adopts the Z-conformation only at much higher salt concentrations than those required to drive the B-Z transition for poly d(CG)
Sequences of d(AT), meanwhile, do not undergo this transition at all under any of the conditions sufficient for inducing it with either sequence above [69, 70]. The great instability of AT base pairs over GC base pairs in Z-DNA is attributed to the loss of two water molecules involved in a bridged hydrogen-bonding system in the case of d(CG) sequences [71, 72].

Z-DNA and its strong association with alternating d(CG) sequences is but one specific case of salt-induced changes to DNA structure, and the strong preference of other sequences for B-DNA over Z-DNA is not pathognomonic for the stability of their B-form against other structural alterations. Salt-induced conformational transitions toward a number of non-B-forms or, in other cases, more minor alterations to helical parameters have been described for a range of DNA sequences. Many of these structural effects appear to have a degree of ion-specificity; that is, they are driven not merely by high salt concentrations, but by high concentrations of a given type of salt. The cesium cation, especially when present as its fluoride salt, seems particularly adept at inducing structural perturbations to DNA. At concentrations of 3.7 M, CsF was observed by circular dichroism to cause a conformational transition of poly d(CA)∙poly d(GT) away from B-form [68]. In a study that shortly followed, a similar transition observed for poly d(AT) was determined via $^{31}$P NMR to reflect a rearrangement of the base positions in (and an unwinding of) the double helix, with increased base stacking for the ApT linkage and decreased base stacking for the TpA linkage [70]. By comparison, NaCl as well as a number of other salts have been shown to pose little threat to the structural integrity of B-DNA [56, 67, 68].

The varying degrees to which the aforementioned salts perturb B-DNA structure and destabilize the double helix provide an interesting comparison with the Hofmeister series, which empirically classifies anions and cations by their effects on protein stability and solubility [73, 74]. In order of increasing ability to precipitate proteins, the Hofmeister series for the cations and anions
discussed herein are, respectively, \( \text{Li}^+ < \text{K}^+ \approx \text{Na}^+ < \text{NR}_4^+ \) and \( \text{ClO}_4^- < \text{Cl}^- < \text{F}^- < \text{SO}_4^{2-} \) [75, 76].

The mechanism by which salts in the Hofmeister series are more or less solubilizing to proteins is thought to be due to specific (preferential) interactions of the ions with protein groups and solvating water molecules; salts appearing earlier in the series (e.g., \( \text{Na}^+\), \( \text{ClO}_4^- \)) favor interactions with hydrophobic protein residues and increase the solubility of nonpolar molecules, thereby perturbing higher-order structures of the protein by diminishing the significance of the hydrophobic effect on protein folding [76]. Considering that DNA bases are relatively nonpolar compared to the phosphate backbone, one might expect that these salts would be more destabilizing to the DNA duplex than salts appearing later in the series. Instead, as shown in Figure 1.6, the opposite seems to be true: the chloride salts of sodium, lithium, and potassium, as well as the sulfate salts of sodium and lithium, are overtly stabilizing to B-DNA. Meanwhile, the cesium salts are increasingly destabilizing to B-DNA in the order \( \text{Cl}^- < \text{F}^- \), and the tetra-alkyl chloride salts follow the order \( \text{TMA}^+ < \text{TEA}^+ \). The limitation of Hofmeister series predictions for DNA melting has been proposed to be a result of the difference in composition of the surface exposed on DNA melting (≈35% hydrocarbon) compared to protein unfolding (≈65% hydrocarbon) [77].

Clearly, salt effects on DNA duplex stability and structure depend on both the type of salt (cation and anion identities) and the DNA sequence composition. In the context of the oligonucleotides used in the studies in Chapters 2-4, which contain A/T-tracts and vary in composition between 50-67% GC, salt (particularly NaCl) at concentrations up to at least 3.5 M is expected neither to significantly destabilize nor alter the structural integrity of the DNA.

### 1.3 Thermodynamics of Compound/DNA Interactions

A typical 1:1 compound/DNA binding interaction can be modeled as follows:

\[
\text{DB} + \text{DNA} \leftrightharpoons \text{DB:DNA}
\]
where “DB” is the DNA binding compound. As with any receptor-ligand interaction of this type, we can define a dissociation constant $K_D$ that describes how readily the complex forms:

$$K_D = \frac{[DB][DNA]}{[DB:DNA]}$$

which is related to the binding free energy $\Delta G = \Delta H - T\Delta S$ by $\Delta G = RT \ln K_D$. Experimental methods abound for determining dissociation constants, from kinetic to steady-state measurements, those that require immobilization of one component to those that are immobilization-free, and those yielding a breakdown of the free energy terms to those that do not.

### 1.3.1 Monitoring binding by fluorescence anisotropy

The intrinsic blue fluorescence of the DB compounds discussed herein can be exploited in the form of a fluorescence polarization binding assay that is label-free, immobilization-free, and broadly amenable to the manipulation of solvent conditions. In brief, fluorescence polarization titrations involve the incremental addition of a receptor into a smaller, fluorescent ligand. When an immobilized fluorophore is excited with plane-polarized light, the light it emits as it relaxes to the ground state will be polarized in the same direction. Because the fluorophore is free to tumble in solution rather than immobilized, the emitted light detected will have some degree of depolarization (i.e., it will have some isotropic character) depending on the rate of tumbling of the fluorophore. As the heavier binding partner is titrated into the fluorophore and the receptor-ligand complex begins to form, the rate of tumbling, and thus the degree of polarization, will be affected (Figure 1.7). In general, heavier object tumble more slowly in solution; the result is an increase in polarization (i.e., increased anisotropy) of the emitted light. The relationship between anisotropy $\langle r \rangle$ and tumbling rate $\phi$ (rotational correlation coefficient) is described by the Perrin equation

$$\langle r(t) \rangle = \frac{r_0}{1 + \tau/\phi}$$
Figure 1.7 Basic principle of fluorescence anisotropy binding assays. The light emitted by a fluorophore is polarized in the same direction as the incident (excitation) light. However, when tumbling freely in solution, the detected light will not appear fully polarized. Since, in general, heavier objects tend to tumble more slowly in solution, the binding of a fluorophore to a ligand is observed as increasing anisotropy of the emitted light. (Excerpted from p. 23 of Ref. 73).

where $\tau$ is the fluorescence lifetime and $r_0$ is the intrinsic anisotropy of the fluorophore. The anisotropy can be monitored over the course of a titration to produce a binding curve from which a $K_D$ can be extracted. In this type of experiment, $\langle r \rangle$ is measured according to the ratio of the parallel ($I_{VV}$) and perpendicular ($I_{VH}$) components of the total fluorescence intensity

$$\langle r \rangle = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}$$

where $G$ is a grating factor specific to the instrument that accounts for differences in the optical efficiencies of the polarized components. For a more in-depth discussion of the physical basis of fluorescence anisotropy, the reader is referred to p. 24-25 of this author’s undergraduate Honors thesis [78].

1.3.2 Probing molecular hydration via osmotic stress

As with the measurement of binding constants, numerous approaches exist to quantify molecular hydration in its many forms, each one presenting its own limitations and advantages. For example, crystallographic studies can be used to identify exceptionally stable water molecules
that play a structural role; volumetric studies offer a means of directly measuring a molecule’s partial molar volume via density measurements; high pressure studies employ the use of hydrostatic or osmotic pressure to obtain hydration numbers linked to volume changes; nuclear magnetic resonance (NMR) studies allow for a quantitative description of the dynamics of hydration waters; and molecular dynamics simulations can lend insight into all of these hydration behaviors, tying together experimental insights from these and other techniques. In this work, the primary approach used to probe hydration is via osmotic stress (OS) experiments, an indirect method for tracking changes in preferential hydration accompanying the binding of a ligand to its receptor. The focus of this section will therefore be on the theory and applications of OS as it pertains to the compound-DNA systems studied herein.

OS theory is simple to understand conceptually. Suppose the binding of a ligand to its receptor requires a net uptake of hydration waters at the receptor-ligand interface. If water molecules are removed from the environment such that they are unavailable to participate in the binding interaction, the ligand will bind less readily to its receptor. The thermodynamic equivalent of “removing water” is to reduce the water activity (or effective water concentration) of the sample. As the name suggests, the OS method employs the use of osmolytes (typically non-ionic cosolutes) to reduce the water activity, which is inversely related to solution osmolality.

The physical basis of this phenomenon is rooted in Wyman’s theoretical description of receptor-ligand binding [79], which takes the form of a linkage equation. The general Wyman linked function demonstrates the thermodynamic coupling between a shift in a receptor-ligand binding equilibrium and the alteration of the chemical potentials of receptor and ligand on the addition of a second component to the system. For simplicity, we will use the case of the
compound-DNA binding system in this discussion. We will denote the compound as DB and the second component as L. The general Wyman linkage for this system is then given by

\[
\left( \frac{\partial \log K_{obs}}{\partial \log a_L} \right)_{P,T,m_{DB},m_{DNA}} = v_L^{DB-DNA} - v_L^{DNA} - v_L^{DB} = \Delta v_L. \tag{1.13}
\]

In the above equation, \( K_{obs} \) is the binding constant for the DB-DNA interaction, \( a_L \) is the activity of L, \( m_{DNA} \) or \( m_{DB} \) is the concentration of DNA or DB, and \( v_L \) is the number of bound L (the superscript indicates the component to which L is bound – either the free DNA, the free DB, or the DB-DNA complex). Thus, at a given ligand concentration \( m_L \) at constant temperature and pressure, the change in binding by L (\( \Delta v_L \)) is given by the change in the DB-DNA binding constant with respect to the change in activity of L. From this linkage relationship, we arrive at two important conclusions central to the Wyman theory of binding. First, the addition of L to the system will shift the equilibrium toward formation of the DB-DNA complex only if the number of L bound to the DB-DNA complex is greater than the summed number of L bound to the free DNA and L bound to free DB. In other words, L will drive DB-DNA binding only if the L has a higher affinity for DB-DNA than for DNA and DB in their free states (\( \Delta v_L > 0 \)), and vice-versa (\( \Delta v_L < 0 \)) for driving DB-DNA dissociation. Significantly, this is true regardless of the various affinities of the components or of their binding mechanisms. Second, the change in ligand binding between the two states (free DNA and DB versus bound DB-DNA complex) can be quantified from the slope of a log-log plot of \( K_{obs} \) versus \( a_L \).

The modeling of OS is a specific case of a Wyman linked function in which L is the osmolyte and where L “interacts” with DNA and DB-DNA only by virtue of its effect on water activity (and thus the interaction of water with DNA and DB-DNA). The relationship between solution osmolality and water activity is
\[ RT \ln a_{w,L} = -\bar{V} \pi = -RT \frac{Osm}{55.5} \] (1.14)

where \( R \) is the gas constant, \( T \) is the thermodynamic temperature, \( a_{w,L} \) is the activity of water in a solution of osmolyte \( L \) having osmotic pressure \( \pi \), \( \bar{V} \) is the molar volume of water, and \( Osm \) is the concentration of osmolyte in Osmol/kg. The osmolyte therefore acts as a probe for the physical binding interaction between water and DNA or DB-DNA and, in this way, the change in water “binding” (preferential hydration) between the free and bound DNA states can be quantified. The Wyman linkage for the stoichiometric coupling of the change in hydration (\( \Delta v_w \)) with compound-DNA binding is

\[ -\left( \frac{\partial \log K_D}{\partial \log Osm} \right) = -1 \frac{55.5}{\left( \frac{\partial \log K_D}{\partial \log a_{w,L}} \right)} = -\left( \frac{\Delta v_w}{55.5} \right) \] (1.15)

where \( \Delta v_w < 0 \) and \( \Delta v_w > 0 \) indicate, respectively, a net release or net uptake of water on binding.

A key assumption of the OS method is that the osmolytes are “inert” in the sense that they act only to increase the osmotic pressure of the system and otherwise do not interact with receptor or ligand. This assumption is critical, as is evident from a reflection on the implications to the Wyman linkage in its absence: preferential interactions with the osmolyte can drive a receptor-ligand binding event even if water is preferentially excluded in the bound state (and vice-versa). In some cases, the assumption of the osmolytes as “inert” is valid, as demonstrated by the qualitative and quantitative uniformity with which osmolytes of different chemical structures and physical properties perturb binding [80, 81]. In other cases, however, this assumption is decidedly not valid due to the presence of preferential interactions for which a theoretical basis does not exist. Importantly, the latter case has been demonstrated for DNA minor groove binding systems similar to those of interest to this work [82, 83]. Without a clearly defined physical basis for the
preferential interactions of non-ionic co-solutes, the presence of such interactions confounds the
data and poses significant limitations toward its interpretability.

1.3.3 Salt as co-solute: coupling hydration effects to counterion release

One straightforward method of circumventing the difficulty presented by ill-defined preferential interactions is to use as osmolyte a co-solute whose preferential interactions are understood in detail and have a well-characterized theoretical basis, thus enabling the preferential interactions to be quantitatively taken into account in the interpretation of the data. To that end, monovalent salts such as NaCl are convenient co-solutes with which to perturb the interactions of small compounds with DNA, as the theoretical basis of polyelectrolytes and their ionic environments is very well-established (see Section 1.2). As shown by Record, et al. [52], the number of thermodynamically-bound counterions displaced on binding of a ligand to DNA is $\Delta \nu_\pm$:

$$-\left(\frac{\partial \log K_D}{\partial \log a_\pm}\right) = -Z\psi + k = \Delta \nu_\pm$$

(1.16)

where $a_\pm$ is the mean ionic activity of a salt whose cation is of valence $Z$, $\psi$ is the thermodynamic ion association parameter for DNA that includes contributions from counterion condensation and screening effects, and $k$ is the equivalent parameter for the case of the positively-charged ligand to which anions from the salt are thermodynamically bound. At high ($\geq 1$ M) salt concentrations, hydration effects become significant. The extension of this Wyman linkage by Tanford [84] to include these effects allows preferential hydration to be determined using salt as an osmolyte. The Tanford-Wyman formalism for coupling of hydration changes to electrostatics [85] is

$$-\left(\frac{\partial \log K_D}{\partial \log a_\pm}\right) = \Delta \nu_\pm - \frac{2m_\pm}{55.5\ln 10} \Delta \nu_w$$

(1.17)

where $a_\pm$, $\Delta \nu_\pm$, and $\Delta \nu_w$ are defined as in Eqs. (1.15) and (1.16) and $m_\pm$ is the molality of the salt. As before, the sign of $\Delta \nu_w$ indicates the disposition of preferential hydration on binding of
compound to DNA (Figure 1.8). If binding is not coupled to a change in hydration, then $\Delta \nu_w = 0$ and Eq. (1.17) takes the form of Eq. (1.16).

![Figure 1.8 Coupling of preferential hydration to electrostatics. Simulated fits to the integrated form of Eq. (1.17) with $\Delta \nu_\pm = -0.88$ are shown for the case of $\Delta \nu_w = -50$ (net release of preferential hydration on binding), $\Delta \nu_w = +50$ (net uptake of preferential hydration on binding), and $\Delta \nu_w = 0$ (no link between preferential hydration and binding).]

1.4 Research Aim

This thesis attempts to define relationships between the structures of various DNA minor groove binding compounds and their DNA binding properties such as affinities, hydration effects, binding stoichiometries, and DNA recognition specificities.

1.5 Outline of the Thesis

In Chapter 2, an investigation of the DNA binding properties of four structurally-related minor groove binding compounds is presented. The compounds share a biphenyl-indole core
structure but differ in charge (+1 or +2) and substitution at the cationic termini. This study reports a previously unrecognized binding mode for two of the compounds in which they bind to DNA as dimers. This binding mode is specific to the A₂T₂ DNA binding site and seems to be a feature of only the dications. Possible mechanisms for facilitating self-association of the dications but not the monocations are discussed, as are the biophysical and translational implications of this finding.

Chapter 3 presents a subsequent study in which the hydration properties of the two monocationic compounds from Chapter 2 are investigated alongside three additional novel monocationic compounds harboring the indole-biphenyl scaffold. The compounds studied are the monoamidine DB1944 and the tetrahydropyrimidine (THP), isopropyl-amidine, dimethyl-THP, and 2-imidazoline derivatives of DB1944.

Chapter 4 contains results from a similar study on the role of hydration in DNA recognition specificity by an extended heterocyclic diamidine. Unlike the compounds discussed in Chapters 2 and 3, the cationic termini of this compound hold between them a selenophene-bis-benzimidazole scaffold that is isohelical with the minor groove. In this chapter, the hydration properties of the compound are analyzed in the context of its complex with three different DNA sequences.

This thesis concludes with Chapter 5, which attempts to consolidate the findings from the three preceding studies into some general conclusions about the structure-activity and structure-hydration relationships of compounds with a given structure. In this chapter, I also outline a number of unanswered questions and propose future directions aimed at providing answers to these questions.
2 DNA RECOGNITION BY LINEAR INDOLE-BIPHENYL DNA MINOR GROOVE LIGANDS


Elsevier’s Permissions of Use:

“Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.” https://www.elsevier.com/about/policies/copyright/permissions [86].

Author’s contribution: Noa Erlitzki performed all fluorescence anisotropy and fluorescence quenching experiments, prepared samples for size-exclusion chromatography, analyzed the data, and contributed to the preparation of the manuscript.
Abstract

Linear heterocyclic cations are interesting DNA minor groove ligands due to their lack of isohelical curvature classically associated with groove-binding compounds. We determined the DNA binding properties of four related dications harboring a linear indole-biphenyl core: the diamidine DB1883, a ditetrahydropyrimidine derivative (DB1804), and their monocationic counterparts (DB1944 and DB2627). These compounds exhibit heterogeneity in binding in accordance with their structures. Whereas the monocations exhibit salt-sensitive 1:1 binding to the duplex 5′-CGCGAATTCGCG-3′ (A₂T₂), the dications show a marked preference for a salt-insensitive 2:1 complex. The two binding modes are differentially modulated by salt and specific non-ionic cosolutes. For both dications, 2-methyl-2,4-pentanediol enforces 1:1 binding as observed crystallographically. Fluorescence quenching studies show self-association without DNA in a relative order that is correlated with preference for the 2:1 complex. The data support a structure-binding relationship in which favorable cation-π interactions drive dimer formation via antiparallel stacking of the linear indole-biphenyl cation motif.

Introduction

Studies into the properties of DNA-binding ligands inform the design of DNA-targeting drugs and extend our understanding of DNA recognition more broadly. Linear compounds that target the DNA minor groove are of particular interest, as they deviate from the conventional notion of curvature matching that of the groove (isohelicity) as a requirement for high-affinity DNA binding [87, 88]. A classic example is the anti-trypanosomal agent CGP 40215A, a linear symmetric diamidine that binds as well as its curved analog Berenil to AT-rich DNA [89, 90]. The diamidine DB921, which harbors a linear benzimidazole-biphenyl core, binds the AT-rich minor groove with ~10-fold higher affinity than its isohelical analog DB911 [91]. Co-crystal structures
of these and other compounds with DNA [92, 93] all reveal water molecules bridging the gaps between ligand and minor groove. These examples establish the ability of non-isohelical compounds of diverse structures to bind to DNA and a role for hydration in facilitating complementarity with the DNA minor groove.

More recently, the co-crystal structures of the asymmetric linear dications DB1804 and DB1883 with DNA were reported [94]. DB1883 is the indole-biphenyl analog of DB921, while DB1804 is a carbocyclic derivative of DB1883. Both DB1883 and DB1804 bind AT-rich DNA with similar affinities as DB921. With the report of a mono-amidine derivative of DB1883 exhibiting weak binding to the same DNA target, we became interested in the structure-affinity relationships between charge density and substitution at the termini of these linear minor groove binders. In our investigations, we found that these compounds were heterogeneous in their DNA binding properties. Both DB1804 and DB1883, but not their monocationic counterparts, form 2:1 complexes in preference over the 1:1 complexes observed in co-crystal structures with the same AATT-bearing DNA target. The discrepant binding properties exhibited by the same compounds and DNA between solution and crystal suggest that preferential interactions with other solutes play an important role in directing the binding modes of these minor groove ligands. While “non-standard” binding modes of minor groove-binding compounds and their DNA sequence requirements have been extensively described [reviewed in 33], the physical chemistry of this behavior, which requires experimental characterization of the properties of the unbound compounds, is less well understood. We therefore interrogated a set of four related indole-biphenyl compounds consisting of DB1804, DB1883, DB2627, and DB1944 (Figure 2.1) in both their DNA-bound and free states. DB2627 and DB1944 are the monocationic derivatives of DB1804 and DB1883, wherein the amidine and tetrahydropyrimidine (THP) at the phenyl ends are
uninstalled. The data show a structure-binding relationship for an indole-biphenyl cation core that is sensitive to the DNA minor groove as well as the physicochemical environment in the absence and presence of DNA.

**Figure 2.1 Minor-groove binding ligands and target DNA used in this study.** The minor-groove binders consist of the linear indole-biphenyl amidine DB1944 and three inter-related compounds, shown in the ionization state expected at pH 7.5. DB2627 is the tetrahydropyrimidine (THP) analogue. DB1883 and DB1804 are dicationic variants of DB1944 and DB2627, respectively. The two DNA targets are the standard AATT dodecamer (A2T2) and an isomeric sequence in which the AT-tract is interrupted (A2CGT2).
2.3 Materials and Methods

2.3.1 Compounds and DNA

The syntheses of DB1883, DB1804, and DB1994 have been previously described [94]. The synthesis of DB2627 is detailed in Supplemental Methods. DNA oligonucleotides encoding A\textsubscript{2}T\textsubscript{2} and A\textsubscript{2}CGT\textsubscript{2} (Figure 2.1) were synthesized by Integrated DNA Technologies (Coralville, IA). Lyophilized DNA was dissolved at 1 mM in 10 mM TrisHCl (pH 7.5) containing 1.0 M NaCl and then dialyzed (MWCO 3.5 kDa) extensively against 10 mM TrisHCl (pH 7.5). Duplex DNA concentration was measured by UV absorption at 260 nm using the nearest-neighbor extinction coefficients 191,511 M\textsuperscript{-1} cm\textsuperscript{-1} for A\textsubscript{2}T\textsubscript{2} and 190,127 M\textsuperscript{-1} cm\textsuperscript{-1} for A\textsubscript{2}CGT\textsubscript{2}. All other reagents were obtained at ACS grade or higher purity and used without further purification.

2.3.2 Binding experiments

DNA binding at equilibrium was monitored by steady-state polarization of the intrinsic blue fluorescence of the compounds as previously described [82, 95]. In brief, each compound was titrated with DNA in 10 mM TrisHCl (pH 7.5) with or without other co-solutes as stated in the text and measured using a Perkin-Elmer LS 55 instrument. The excitation and emission maxima were established at 328/456 (DB1883), 338/454 (DB1944), 337/438 (DB2627), or 320/438 nm (DB1804). Steady-state anisotropy was computed using a grating factor as determined under the conditions of each measurement. With excitation and emission slit widths of 15 and 20 nm centered at these wavelengths, each compound was sampled at the lowest concentration sufficient to acquire signal for anisotropy measurements after blank subtraction: 50 nM for DB1883, 20 nM for DB1944, 10 nM for DB2627, 20 nM for DB1804, unless otherwise indicated in the text.
2.3.3 **Steady-state fluorescence quenching**

Each compound was titrated at a constant concentration of 200 nM in water with NaI, acrylamide, and nicotinamide. Total intensity and steady-state anisotropy at $\lambda_{\text{max}}$ were adjusted for volume changes, blank-subtracted, and analyzed by linear regression.

2.3.4 **Data analysis**

The signal from DNA titration experiments represented the fractional bound compound ($F_b$), scaled by the limiting anisotropies of the ensemble of $n$ (typically 1 or 2) bound states $\langle r \rangle_i$ and unbound state $\langle r_0 \rangle$ as follows:

$$\langle r \rangle = F_b \left( \sum_{i=1}^{n} \langle r \rangle_i - \langle r_0 \rangle \right) + \langle r_0 \rangle = F_b \sum_{i=1}^{n} \Delta r_i + \langle r_0 \rangle$$ (2.1)

$F_b$ is described empirically by the Hill equation or a mechanistic binding model as described in the text with total DNA concentration taken as independent variable. For salt-dependent analysis, mean ionic activity was calculated from molal concentration and literature values of the mean ionic activity coefficient $a_{\pm}$ in water [96]. The dependence of DNA-binding affinities on $a_{\pm}$ is analyzed in terms of polyelectrolyte theory to estimate the number of neutralized DNA backbone phosphates $Z$ from the net number of displaced ions $n_{\pm}$:

$$-\frac{\partial \log K_{12}}{\partial \log a_{\pm}} = n_{\pm} = \psi Z = \left( \psi_{\infty} - \frac{2.53}{N} \right) Z$$ (2.2)

where $\psi$ reflects screening and condensation interactions between backbone phosphates in B-DNA and their ion atmosphere [85]. The assigned value of $\psi = 0.67$ includes an end-effect correction for our $N = 12$ bp\(^1\) oligonucleotide duplexes relative to polymeric DNA ($\psi_{\infty} = 0.88$) [53].

\(^{1}\) Correction: The variable $N$ represents the number of phosphates, not base-pairs. For a synthetic DNA dodecamer, $N = 20$ and $\psi = 0.75$. 
2.4 Results

Binding to the target duplex A\textsubscript{2}T\textsubscript{2} in solution was determined at equilibrium by titration with DNA via the large change in anisotropy of the intrinsic blue fluorescence of the compounds. This technique, which obviates the need for extrinsic labeling or immobilization of the DNA, reverses the more common approach of titrating the DNA with compound. The titrations were designed such that DNA concentration was varied over five or more decades while keeping the dilution of compound to less than ~5\%, which was sufficiently fixed for one-dimensional analysis [95]. At pH 7.5, as a function of increasing NaCl concentration from 0.010 to 0.750 M, the titration profiles for DB1804 exhibited an increasingly biphasic appearance (Figure 2.2A), while those for the other compounds remained monophasic (Figures 2.2B to D). As a first step to parameterize the two binding modes for DB1804, we fitted the data empirically with a sum of two Hill equations:

\[ F_b = f \frac{c^{n_{H,1}}}{K_{D,1}^{n_{H,1}} + c^{n_{H,1}}} + (1 - f) \frac{c^{n_{H,2}}}{K_{D,2}^{n_{H,2}} + c^{n_{H,2}}} \]  

where \( n_{H,i} \) is the Hill coefficient and \( K_{D,i} \) is the DNA concentration at half maximal occupancy for binding mode \( i = 1 \) or 2. The scaling factor \( f \) represents the fractional contribution to the total anisotropy change from each binding mode. To statistically infer the extent to which the salt-dependent binding profiles exhibited biphasic character, we compared the fits of each dataset by Eq. (2.3) (with \( f \) floating) relative to a single term (\( f \) fixed at 1) using the Fisher F-test on the residual sums of squares. Across the full range of NaCl concentrations tested, the two-term Hill model afforded significantly better fits to the DB1804 data than a single term (\( p < 0.05; \) Table 2.1, Supplemental Data). Thus, the binding properties of DB1804 (the THP dication) exhibited two spectroscopically distinguishable binding modes at equilibrium. In contrast, titration profiles for its mono-THP counterpart (DB2627) as well as the di- (DB1883) and mono-amidine (DB1944) were monophasic across the entire salt range as confirmed by \( F \)-testing (Table 2.1).
Figure 2.2 NaCl unmasks two distinct binding modes for the ditetrahydropyrimidine DB1804. Rows A to D show representative $A_2T_2$-into-compound titrations for each species in the presence of 10 to 750 mM NaCl. Curves represent fits by either a one- (gray) or two-term Hill equation (red) as given in Eq. (2.3). Compounds were present at $10^{-9}$-$10^{-8}$ M as described in Materials and Methods.
2.4.1 Linear dications exhibit salt-sensitive and salt-insensitive binding modes

Examination of the empirical affinities ($K_{D,i}$) revealed distinct trends in salt dependence for the four related compounds (Figure 2.3). For DB1804, the high-affinity binding mode ($K_{D,2}$) exhibited no salt dependence, while the low-affinity binding mode ($K_{D,1}$) varied linearly with mean ionic activity with a log-log slope of -0.63 ± 0.10. For the diamidine DB1883, the titration curves were insensitive to salt, similar to the high-affinity mode of DB1804. We considered the possibility that the high-affinity modes might appear to be insensitive to salt due to titrant depletion, i.e., binding was tight such that $K_{D,1}$ reflected the (fixed) compound concentrations in the titrations rather than binding affinities. However, the apparent dissociation constants corresponding to the high-affinity mode ($K_{D,2}$) were ~10-fold lower than the concentrations of the dications used (20 nM for DB1804, 50 nM for DB1883). Moreover, the dissociation constants for both dications agree closely with values measured by surface plasmon resonance, an altogether different experimental configuration [94]. Therefore, the apparent salt insensitivity could not be significantly attributed to depletion. We concluded that the salt insensitivity of the single binding mode of DB1883 and the high-affinity mode of DB1804 was intrinsic to their binding properties.

For the monocations, the THP (DB2627) bound A$_2$T$_2$ with ~3-fold higher affinity than the amidine DB1944, and both compounds gave identical salt dependence at -0.60 ± 0.10 (DB1944) and -0.64 ± 0.03 (DB2627) in log-log slope. Interpreting this slope by polyelectrolyte theory [85], given by Eq. (2.2) for oligonucleotides [53], binding of A$_2$T$_2$ by DB2627 and DB1944 corresponded to the neutralization of one DNA phosphate.

The salt dependence data suggested that the salt-insensitive high-affinity binding mode for DB1804 to A$_2$T$_2$ was similar to DB1883. Likewise, the salt-sensitive low-affinity mode for DB1804 was similar to the two monocations. Since the titrations used DNA as titrant, increasing
Figure 2.3 Salt dependence of DNA binding by indole-biphenyl mono- and dications. Empirical dissociation constants of titration curves obtained at 10 to 750 mM NaCl were estimated by a Hill analysis according to Eq. (2.3). DB1804, DB1883, DB2627, DB1904. Closed and open symbols refer respectively to the apparent high- and low-affinity modes, i.e., $K_{D,1}$ and $K_{D,2}$ in Eq. (2.3), observed with DB1804 that were absent with the other compounds.

DNA concentration was expected to drive the equilibria in the direction of decreasing stoichiometric order with respect to compound. Combining these clues, we hypothesized that the biphasic binding by DB1804 reflected a distribution between two distinct stoichiometric complexes with DNA. To test this hypothesis, we determined the stoichiometry of the apparent complexes for all four compounds under depleting conditions at the low NaCl concentration of 5 mM (Figure 2.4). The binding curves revealed that the high-affinity binding mode for DB1804 and the single binding mode for DB1883 represented the dications in 2:1 excess to DNA, while their monocationic counterparts exhibited equimolar binding. To rule out the formal possibility that these complexes might consist of DNA at multiple equivalents, we probed DB1804-bound
A₂T₂ by size-exclusion chromatography (Figure 2.11, Supplemental Data). At DNA concentrations sufficient for UV detection (5 µM), we detected no species consisting of two or more A₂T₂ duplexes. The evidence therefore showed that the spectroscopically distinct modes of DB1804 binding to A₂T₂ both involved a single duplex.

Figure 2.4 Stoichiometry of A₂T₂-bound complexes of linear cations. Binding was measured at 5 mM NaCl. Compound concentrations ranged from 10 to 140 nM.

To mechanistically analyze the binding properties of DB1804, we modeled its titration profiles according to Scheme I:

\[
\begin{align*}
\text{DNA} & \rightleftrarrows K_1 \text{ DB : DNA} \rightleftrarrows K_2 \text{ DB}_2 : \text{DNA} \\
\end{align*}
\]

Scheme I
where the equilibria are written (from left to right) in the opposite direction as our titrations. $K_1$ and $K_2$ are the intrinsic stepwise dissociation constants for the 1:1 (low-affinity) complex and 2:1 (high-affinity) complex, respectively. The functional form of Scheme I is given in Supplemental Methods. As shown in Figure 2.5A, Scheme I described DB1804 binding to $A_2T_2$ across the full range of NaCl concentrations tested and captured the distinct salt dependence of the 1:1 and 2:1 complexes. To generalize the salt-dependent data, we tested the effect of Na$_2$SO$_4$ in place of NaCl on DB1804 binding. The titration profiles for Na$_2$SO$_4$ showed similarly biphasic properties that were also described by Scheme I, although the two binding modes were not as well-resolved as for NaCl at matching Na$^+$ concentrations. However, when cast as a function of mean ionic activity [96], perturbation of binding by Na$_2$SO$_4$ fell in line with the data for NaCl (Figure 2.5B). Thus, the salt-induced divergence of the two binding modes of DB1804 was consistent with an equilibrium distribution of a 2:1 and a lower-affinity 1:1 complex, only the latter of which was sensitive to the ionic environment (-0.75 ± 0.18 in log-log slope). Moreover, the independence of this perturbation from anion identity confirmed that release of condensed DNA counter-ions, as described by polyelectrolyte theory, accounted for the disposition of ions in salt-sensitive binding.

Two features of Scheme I are of note. First, since the final state in the titration was the 1:1 complex, the model assigned a higher steady-state fluorescence anisotropy to the 1:1 complex than the 2:1 complex in the titrations (on the order of 10%). Second, the equilibrium constants are formulated with unbound compound as monomers. While it is possible to incorporate additional equilibria for the self-association of the unbound compound, the parameters for these interactions, such as the anisotropy of the DNA-free species, are not well defined by the titration data. Although these details were not included in Scheme I, the scheme afforded a satisfactory fit to the
experimental data and captured the salient details of the system, namely the salt-sensitive low-affinity binding mode and the salt-insensitive high-affinity binding mode.

![Figure 2.5 DNA recognition by DB1804 is mechanistically described by an equilibrium distribution of 2:1 and 1:1 complexes.](image)

A, Titration data for DB1804 for the three NaCl concentrations (green) shown in Figure 2.2A was globally fitted to Scheme I with $\Delta <r>$ for the two complexes shared, c.f. Eq. (2.1). DNA titrations in the presence of Na$_2$SO$_4$ instead of NaCl, shown here in orange at matching Na$^+$ concentrations, were also fitted with this model. Curves are offset vertically for presentation. B, Salt-dependence of the dissociation constants for the low- (open symbols) and high-affinity (closed symbols) transitions obtained by fitting the binding data to Scheme I. Lines of best fit for $K_1$ (dashed) and $K_2$ (solid) were obtained by globally fitting the data for NaCl (green) and Na$_2$SO$_4$ (orange).

### 2.4.2 The binding modes of DB1804 are DNA sequence-specific

Having established a 2:1 complex as the high-affinity binding mode for DB1804 to A$_2$T$_2$, we asked whether this behavior was specific to the DNA sequence (Figure 2.6). To address this question, we permuted the dodecameric A$_2$T$_2$ to generate an isomeric sequence harboring 5’-AACGTT-3’ (A$_2$CGT$_2$; Figure 2.1). DB1804 bound A$_2$CGT$_2$ more weakly than A$_2$T$_2$ and
exhibited more than one binding mode, although the corresponding anisotropies did not coincide. At 10 mM Na\(^+\), the high-affinity mode for A\(_2\)CGT\(_2\) was comparable to that for A\(_2\)T\(_2\), but the low-affinity mode was ~100-fold weaker. In contrast with A\(_2\)T\(_2\), the high-affinity mode for A\(_2\)CGT\(_2\) was salt-sensitive, becoming ~100-fold weaker in 200 mM Na\(^+\). In addition to the apparent affinities, the anisotropies associated with the binding modes for A\(_2\)CGT\(_2\) progressively diverged from those associated with A\(_2\)T\(_2\) with increasing Na\(^+\) concentration. These changes suggested DNA-dependent dynamics, photophysical properties of different binding modes or, more likely, the development of additional nonspecific modes with A\(_2\)CGT\(_2\). Moreover, as the salt dependence of binding was described by DNA counter-ion condensation over the experimental salt concentrations (*vide supra*), the observed sequence specificity was expected to be general with respect to cation identity [37]. In summary, the two binding modes observed with the A\(_2\)T\(_2\) site were sequence-specific and therefore relevant to this high-affinity DNA that dications are generally known to target.

**Figure 2.6** The 2:1 and 1:1 binding modes exhibited by DB1804 are sequence-specific. DB1804 was titrated with the interrupted AT-tract (A\(_2\)CGT\(_2\); circles, blue) in 10, 200, and 750 mM NaCl. The data is plotted alongside corresponding data for the specific site A\(_2\)T\(_2\) from Figure 2.2A (squares, red) to facilitate comparison. Curves represent empirical fits by the two-term Hill equation, Eq. (2.3). \(K_{D,1}\) and \(K_{D,2}\) are indicated by dashed and dotted drop lines, respectively. At 750 mM NaCl, additional binding modes appeared likely for A\(_2\)CGT\(_2\).
2.4.3 The binding modes of DB1804 are sensitive to inhibition by netropsin

To further define the high- and low-affinity DNA-bound states of DB1804, we challenged the DB1804:A₂T₂ complexes with netropsin, a well-established minor groove ligand for A₂T₂. Competition titrations with (non-fluorescent) netropsin were performed at DNA concentrations corresponding to saturated and various levels of sub-saturated binding by DB1804 in 10 or 750 mM NaCl (Figure 2.7). Displacement of DB1804 from A₂T₂ by netropsin was indicated by a decrease in the apparent anisotropy of DB1804. As expected for the salt-sensitive binding of

Figure 2.7 Inhibition of A₂T₂-bound DB1804 by netropsin.
Netropsin was titrated into A₂T₂-bound DB1804 at 10 mM (top left) and 750 mM NaCl (bottom left). DB1804 was present at 20 nM in all cases. At both Na⁺ concentrations², DB1804 was complexed with 2 nM (triangles), 8 nM (circles), and 700 nM DNA (squares). Binding between DB1804 and netropsin was detected above ~10⁻⁶ M netropsin in the absence of DNA (gray). Unbound DB1804 is marked by ×. Data from Figure 2.2A showing A₂T₂ titrations at the corresponding salt concentrations is shown in the right panels for comparison.

² Correction: At the lower Na⁺ concentration (10 mM), the 2 nM DNA condition was not tested.
netropsin to A₂T₂ [97], it competed for the DNA more strongly at the low salt condition. At all DB1804 concentrations, netropsin was initially observed to displace the bound DB1804 from the DNA. Interestingly, control experiments showed an interaction between DB1804 and netropsin above 10⁻⁶ M concentration of the latter in the absence of DNA. Though this unexpected behavior obscured a full competition profile, it remained apparent that both binding modes could be inhibited by netropsin and supported the minor groove as the binding site of DB1804 in both modes.

2.4.4 Preferential interactions with co-solutes modify DB1804 binding modes

To better understand the physicochemical basis of the different binding modes exhibited by DB1804, we examined the effects of non-ionic co-solutes on its DNA binding equilibria. We initially focused on dimethyl sulfoxide (DMSO) and nicotinamide, two common solubilizing agents for low-molecular weight compounds. At up to 20% v/v DMSO (2.8 m), the biphasic transition persisted, but the apparent affinities for both binding modes were attenuated (Figure 2.12A, Supplemental Data). In contrast, nicotinamide at up to 0.10 M (0.09 m) exerted opposing effects on the two binding modes, slightly favoring the high-affinity mode while significantly destabilizing the low-affinity mode (Figure 2.12B, Supplemental Data). The qualitatively different effects of DMSO and nicotinamide on the two binding modes indicated that they arose from specific preferential interactions with the co-solutes and were not due to viscosity or colligative effects such as hydration alone.

In light of the 1:1 complexes observed in co-crystal structures of DB1804 and DB1883 with A₂T₂, and the sensitivity of the binding modes to co-solutes, we considered whether crystallization conditions might contribute to the different binding behaviors observed for these dications in solution. As the reported crystallization conditions [94] employed 2-methyl-2,4-
pentanediol (MPD) as a cryo-protectant and precipitant at up to 50% v/v during equilibration, we performed titrations of DB1804 and DB1883 with $A_2T_2$ in the presence of 50% v/v MPD. Although the high viscosity of MPD strongly perturbed the observed anisotropy of both compounds, it gave equimolar binding by both DB1804 and DB1883 at 750 mM Na$^+$ (Figure 2.8). MPD therefore suppressed the 2:1 complex and enforced the alternative 1:1 complex, even under high-salt conditions in which it would be otherwise disfavored, to yield the stoichiometry seen in the co-crystal structures with $A_2T_2$.

![Graph showing steady-state anisotropy vs. $A_2T_2$ compound ratio for DB1804 and DB1883](image)

**Figure 2.8** The cryo-protectant 2-methyl-2,4-pentanediol enforces 1:1 binding by linear dications. Binding of $A_2T_2$ to 100 nM DB1804 or 740 nM DB1883 at 750 mM NaCl was measured in the presence of 50% MPD as used in the crystal structures. The error bars used during fitting, which average ±0.05, have been omitted in Panel B for clarity. The small change in anisotropy for DB1883 compared to DB1804 reflects the higher unbound anisotropy associated with this compound (c.f., Figure 2.2). Stoichiometric analyses indicate equimolar complexation of each compound with $A_2T_2$ under these conditions. At right, the crystal structures for $A_2T_2$-bound DB1804 (PDB ID: 3U05) and DB1883 (PDB ID: 3U0U) show the two bound species forming 1:1 complexes.
2.4.5 Linear dications self-associate in the absence of DNA

An important feature of the dications not directly addressed by the binding experiments was whether the dications dimerize in the absence of DNA. In one possibility, dication monomers are induced to dimerize onto DNA. Alternatively, unbound dications assemble into pre-formed dimers (or other oligomers) that persist upon binding to DNA. A pre-organized dimer would decrease the entropic barrier for forming the 2:1 complex, offering a possible explanation for the higher affinity of this binding mode relative to the 1:1 complex. To probe whether the compounds self-associate in the unbound state, we determined the effect of quenching agents on the intrinsic fluorescence of each compound. We tested three chemically distinct quenchers (NaI, acrylamide, and nicotinamide) at a common compound concentration of 200 nM to allow for sufficient fluorescence detection upon quenching. We analyzed the decay in total fluorescence intensity upon incremental additions of quencher according to the Stern-Volmer relationship (Figures 2.9A to C):

$$\frac{F_0}{F} = \tau_0 \frac{\tau}{\tau} = 1 + K[Q]$$

(2.4)

where $F_0$ and $F$ represent fluorescence in the absence and presence of quencher Q, and the (positive) slope $K$ is the Stern-Volmer constant $K_{SV}$ in the case of dynamic quenching or the association constant $K_a$ in the case of static quenching. To differentiate between dynamic and static contributions to the observed intensity quench, we simultaneously determined the steady-state anisotropy at each concentration step as a proxy for fluorescence lifetime, $\tau$. The two quantities are related by the Perrin equation:

$$\frac{1}{r} = \frac{1}{r_0} + \frac{RT}{r_0 \eta V} \tau$$

(2.5)
where \( r_0 \) is the limiting anisotropy, \( R \) is the gas constant, \( T \) is the absolute temperature, \( \eta \) is viscosity, and \( V \) is the molar hydrated volume. Dynamic quench is modeled by substitution of the Stern-Volmer relationship into \( \tau \) [98]:

\[
\frac{1}{r} = \frac{1}{r_0} + \frac{RT}{r_0 \eta V} \frac{1}{1 + K_{SV}[Q]}.
\] (2.6)

As Figures 2.9D to F show, the anisotropies exhibited small changes with no systematic trends, even in cases where fluorescence intensity changed significantly. To constrain the analysis, we used a relatively low and narrow range of quencher concentrations (up to 6 mM) to minimize viscosity and preferential hydration effects on the observed anisotropies. As it was then improbable that any change in viscosity and hydration would compensate so similarly for the three chemically disparate quenchers if \( K_{SV} \) were substantial, we concluded that the weak dependence of the observed anisotropy on quencher concentration reflected a small \( K_{SV} \). We therefore reject collisional relaxation of the excited state as a significant contributor to the intensity quench, which must therefore represent quencher-specific interactions of ground state species at equilibrium (static quench).

Having established preferential interactions of the ground-state compounds with the quenchers as the basis of the intensity quench, several overarching observations presented themselves. First, nicotinamide was the most efficient quencher for each compound, consistent with its reputation as a hydrotrrope with strong preferential interaction properties. Second, regardless of quencher identity, each monocation exhibited a higher \( K_a \) (i.e., was more sensitively quenched) than its dicationic analogue. Since it was improbable that the additional charge in the dications would result in less favorable interactions with all three chemically distinct quenchers, we interpreted this behavior as self-association of the dications. Thus, dimeric DB1804 and DB1883 presented significantly less accessible surface areas (or lower effective concentrations) to
the quenchers. Third, the diamidine showed a significantly larger difference in intensity quench relative to the mono-amidine than the corresponding THPs. Along the same line of reasoning, we interpreted this difference in terms of a stronger preference by the unsubstituted dications for self-association (i.e., DB1883 > DB1804). As A₂T₂ DNA induced 1:1 binding by DB1804, but not DB1883 (Figure 2.3), the relative tendency of the linear dications to self-associate appeared to correlate with their preference for 2:1 binding over the 1:1 complex.³

![Figure 2.9](image)

**Figure 2.9 Steady-state fluorescence quenching of linear indole-biphenyl cations in the absence of DNA.**

Total fluorescence intensity (A to C) and steady-state anisotropy (D to F) were measured for each compound at 200 nM in the presence of NaI, acrylamide, and nicotinamide.

---

³ We note that the quenching studies themselves do not provide a stoichiometric evaluation of self-association in the unbound state (i.e., formation of dimers may not necessarily correspond to a two-fold reduction in quenching sensitivity).
2.5 Discussion

The iconic binding mode of minor groove ligands, originally observed in the co-crystal structure of netropsin and A₂T₂ [99], is insertion as a monomer deep into the minor groove of duplex B-DNA, usually with a strong preference for AT-rich regions. Depending on the DNA sequence context, other binding modes have also been described. For example, netropsin binds in two molar equivalents to the minor groove of the self-complementary duplex C₅I₅ as an end-to-end dimer [100]. More subtle are thermodynamically or spectroscopically distinguishable netropsin complexes with AT-rich DNA of different configurations, such as sequence variations [101, 102] or the presence of a hairpin [103]. Beyond the double helix, netropsin also binds the minor groove of a DNA triplex [104, 105]. Reciprocally, specification of binding mode results in conformational selection in the bound DNA, in some cases with dramatic effect [106, 107].

Beyond netropsin, DNA-dependent binding modes have also been described among designed minor groove binders. For example, furamidine (DB75, a diphenylfuran diamidine) and related analogs can intercalate or bind to the minor groove of duplex DNA depending on sequence and the relative positions of the substituents in the compounds [108]. DB293 (a phenyl-furan-benzimidazole diamidine) binds to A₂T₂ as a monomer but to mixed (e.g., 5’-ATGA-3’) sequences cooperatively as a stacked antiparallel dimer [30, 109]. DB1003, a difuran-benzimidazole derivative of DB293, binds 5’-AATT-3’ as a monomer but 5’-TTAA-3’ as a positively cooperative dimer [110]. The present compounds based on the indole-biphenyl scaffold add to the growing diversity of DNA recognition by low-MW ligands and highlight a role for interactions of the compounds in the unbound state in modulating the selectivity of one binding mode over another.
### 2.5.1 Structural determinants of binding mode selection by linear indole-biphenyl cations

The four compounds examined in this study constitute an internally consistent set of analogues from which the structural bases of the multiple binding modes may be inferred. In dilute solution, the dications DB1804 and DB1883 preferentially bind AT-rich DNA as dimers. In the case of DB1804, the dimer succumbs to a 1:1 complex when DNA is in large excess. While the high-affinity 2:1 complex is salt-insensitive, the low-affinity 1:1 complex is destabilized with increasing Na\(^+\) concentration. Thus, for DB1804, the titrations become increasingly biphasic with increasing salt due to the differential sensitivity of the two binding modes to bulk salt concentrations. In the case of DB1883, the corresponding low-affinity mode is presumably too low in affinity to detect at the highest concentration of DNA used (10^{-4} M).

The binding modes exhibited by the dications DB1804 and DB1883 in solution are not reflected in their crystal structures, which consist only of 1:1 complexes. In contrast, DNA binding by their monocationic analogs occurs exclusively as 1:1 complexes. The relationship between the two pairs of amidines (DB1883/DB1944) and THPs (DB1804/DB2627) pose an interesting contrast with an earlier study on DB183 and DB185 [111]. DB185 is a dibenzimidazole-phenyl diamidine that binds 5'-TTAA-3' as a monomer while DB183, a monocationic derivative in which the phenyl end is replaced by an (uncharged) hydroxyl, targets the same site as a positively cooperative dimer. The linear indole-biphenyl compounds in this study also demonstrate a clear relationship between charge number and dimeric binding but in a manner opposite the isohelical pair DB183/DB185. Examples are therefore accumulating that suggest charge number as a parameter for multiple binding modes by heterocyclic cations in the DNA minor groove.
The fluorescence quenching data indicate that both of the dications self-associate at low concentrations in the absence of DNA to a significantly higher extent than the monocations. Structurally, the aromatic indole-biphenyl core suggests a stacking mechanism for dimer formation. The major driving force for stacking, which includes a favorable entropic component due to hydrophobic dehydration of stacked surfaces, should scale with the removal of solvent-accessible surface area. If one side of the indole-biphenyl core becomes inaccessible to solvent in a stacked dimer, stacking would exclude 225 Å² of surface area per monomer in the dimer. Assuming that the π-stacked termini remain fully solvent-accessible, this reduction represents a lower limit of 26% and 32% of the total solvent-accessible surface area of DB1804 (di-THP) and DB1883 (diamidine), respectively. However, the monocations DB2627 and DB1944 are similar in aromaticity to their dicationic counterparts, differing from DB1804 and DB1883 only in the absence of a cationic terminus. Some charge-based interactions must therefore provide the additional driving force to favor dicationic dimers over their monocationic counterparts.

Given the propensity of indoles and phenyl rings to engage in cation-π binding, the data suggest a role for cation-π stabilization in the self-association of dicationic compounds. Thus, one might envision an anti-parallel stacked dimer in which the cation at the phenyl end of one dication stabilizes the π interactions near the indole end of another (Figure 2.10A). Experimental studies with model low-MW compounds estimate the free energy contribution of single cation-π stack on the order of 10 kJ/mol at 25°C, equivalent to 1 to 2 hydrogen bonds in liquid water [112]. Each dication dimer contains two distinct cation-π stacks. The steric complementarity and symmetry of the linear dications arranged in an antiparallel configuration would further favor association.

Cation-π stacking explains why the cation at the phenyl end is essential for dimer formation: the charge at the indole end is off-axis relative to the linear aromatic core and cannot
achieve similar alignment without steric or charge clashes (Figure 2.10B). Monocations harboring substituents at the phenyl end are therefore expected to behave similarly as the non-substituted monocations DB2627 and DB1944. The cation-π model also accounts for the stronger dimeric preference of the diamidine (DB1883) over the di-THP (DB1804), as the π-interacting charge density would be attenuated by the carbocyclic ring in DB1804. Finally, the cation-π stacking provides a basis for the dimer as a self-limiting unit that does not readily associate into insoluble aggregates, as attested by the solubility of the dications up to $10^{-4}$ M in aqueous solution. Indeed, the requirement for charge in dimerization by these linear dications contrasts with, for example, the heterocyclic polyamides for which aggregation is inhibited by increasing charge density [113].

Given the stacked-dimer model of minor groove occupancy proposed for other heterocyclic dications such as DB293 and DB183 [30, 109-111], a dimeric dication in the unbound state may also be related to its DNA-bound conformation in the 2:1 complex. Structural consistency between the DNA-free and DNA-bound dimers would account for: 1) the absence of a 1:1 complex for DB1883 due to its preference for the dimeric state relative to DB1804, and 2) the absence of 2:1 binding by the monocations. The stacked dimer also presents a plausible symmetry argument for a lower apparent anisotropy (increased dynamics) for the 2:1 complex relative to the 1:1 state. The dynamic ensemble for a 2:1 complex, harboring a symmetrized stacked dimer bound to self-complementary DNA, is expected to sample symmetry-related configurations that are absent for the asymmetric 1:1 complex. In addition, a stacked dimer that removes substantial low-polarity surfaces from the solvent is compatible with the solute-specific preferential interactions (DMSO, nicotinamide, MPD) observed with DB1804. The formal possibility of a 2:1 complex in the major groove is discounted by the susceptibility of the 2:1 complex to inhibition by netropsin, an
established minor groove ligand, and the lack of evidence for an allosteric mechanism of inhibition.

Figure 2.10 A **model for dimer formation by linear indole-biphenyl dications.** As illustrative examples, the diamidine DB1883 and its mono-amidine analogue DB1944 are shown here, offset vertically in the plane of the page for presentation in two dimensions. A, An antiparallel stacked dimer of DB1883 is stabilized by π-stacking and cation-π interactions with the amidinium at each phenyl end (colored in blue), as well as molecular symmetry. B, Illustrative configurations of stacking by two DB1944 monocations. The equivalent cation-π interactions to those of DB1883 cannot be achieved by the off-axis amidinium at the indole end of DB1944 without less optimal geometries (marked by double-headed arrows) or steric clashes and electronic repulsion (red exclamation symbols).
The 5’-AATT-3’ motif, as found in A2T2, narrows into the midpoint of the sequence at the minor groove. To accommodate the 2:1 complex, induced perturbation in the structure of both DNA minor groove and ligand is therefore likely, with the possible consequence that additional charge neutralization is needed to maintain a compatible level of axial charge density for the double helix. If an uptake of anions by the compound is coupled to cation (Na+) release from neutralization of DNA phosphates (as demonstrated by the monocations), the compensation could explain the apparent salt insensitivity of high-affinity binding to A2T2 by the two dications. Additional supporting evidence for mutual structural adjustment by DNA and ligand is found in the ~5-fold lower affinity of the 2:1 mode for DB1883, the more facile dimer in the absence of DNA, relative to DB1804. Such structural perturbations may not be needed for the suboptimal DNA site in A2CGT2, for which the minor groove is expected to be wider relative to A2T2. Accordingly, A2CGT2 binding by DB1804 is salt sensitive and nearly as strong as binding to A2T2 at low salt (Figure 2.6).

2.5.2 Functional implications of multiple binding modes for dications as transcriptional inhibitors

Translationally, minor groove-binding heterocyclic cations are promising agents as antimicrobials and other therapeutics [114], including recent success as inhibitors of transcription factors of major oncologic interest [21, 115, 116]. Their therapeutic potential in transcriptional regulation depends, mechanistically, on their ability to competitively inhibit protein/DNA interactions at the minor groove. In this respect, the susceptibility of both binding modes of DB1804 to inhibition by netropsin at equilibrium is significant, as the thermodynamic nature of the data assures that the reciprocal action also holds. In other words, minor groove binding as realized by netropsin can also be inhibited by both the high- and low-affinity binding modes of
DB1804. The 2:1 complex is therefore a functionally effective binding mode as far as minor groove inhibition is concerned, even if the structure of dimeric DB1804 (and, by extension, DB1883) in the minor groove is not definitively defined. Indeed, the high affinity of 2:1 binding (near \(10^{-9}\) M under physiologically saline conditions) relative to the two monocationic analogs supports the 2:1 mode as a desirable characteristic in the molecular design of inhibitors with this scaffold.

### 2.6 Conclusion

We evaluated the solution binding modes of the two linear indole-biphenyl dications DB1804 (ditetrahydropyrimidine) and DB1883 (diamidine) as well as their respective monocations DB2627 and DB1944. In dilute solution, a dimeric DNA binding mode is accessible only to the dications. The monocationic analogues, which differ from the dications only by the absence of one cationic terminus, exhibit a single 1:1 binding mode typical of minor groove binders of this general class. These structure-binding relationships reflect cation-\(\pi\) stacking of a linear indole-biphenyl cation core that is not typically observed with isohelical heterocyclic compounds.

### 2.7 Supplemental Data

#### 2.7.1 Supplemental methods

*Synthesis of DB2627.*

\[
\text{NC} \quad \xrightarrow{\text{i- HCl gas/EtOH}} \quad \text{ii- 1,3-diaminopropane} \quad \xrightarrow{\text{DB 2627}}
\]

The starting cyanooindole derivative (1 mmol; [117]) was dissolved in saturated ethanolic HCl and stirred at room temperature for 2 weeks, isolated from air and moisture. Dry ether was added and the solid, which formed, was filtered, dried under vacuum for 30 min and then dissolved in absolute ethanol. 1,3-Diaminopropane (4.2 mmol) was added to the reaction mixture while cooling and the resulting solution was stirred at room temperature for 4 days. Anhydrous ether was
added and the precipitated crystals (HCl salt) were filtered. The diamidine was purified by neutralization with 1 M sodium hydroxide solution, the solid that formed was filtered, washed with water and dried. Finally, the free base was stirred with ethanolic HCl for 2 days, diluted with ether, and the crystals which formed were filtered and dried to give the diamidine HCl salt.

White solid (0.18 g, 44%), mp > 300 °C. \(^1\)HNMR (DMSO-\(d_6\)): \(\delta\) 12.43 (s, 1H), 9.91 (s, 2H), 8.08 (d, \(J = 8.4\) Hz, 2 H), 8.85 (m, 2H), 7.82 (br s, 1 H), 7.76 (d, \(J = 8\) Hz, 2 H), 7.73 (s, 1H), 7.50 (m, 2H), 7.40 (m, 1 H), 7.35 (dd, \(J = 8.4, 1.2\) Hz, 1 H), 7.13 (d, \(J = 1.2\) Hz, 1 H), 3.52 (s, 4H), 2.01 (m, 2H); ESI-HRMS: m/z calculated for \(C_{24}H_{22}N_3\): 352.1808, found: 352.1781 (Double charged amidine base \(M^+ + 2\)). Anal. Calc. for \(C_{24}H_{21}N_3\). HCl. 1.25H\(2\)O: C, 70.30; H, 6.02; N, 10.25. Found: C, 70.40; H, 5.94; N, 10.25.

**Numerical analysis of Scheme I.** Starting with Scheme I as stated in the main text:

\[
\begin{align*}
\text{DNA} & \rightleftharpoons DB : DNA \\
\text{DB} : DNA & \rightleftharpoons DB_2 : DNA
\end{align*}
\]

**Scheme I**

the two stepwise dissociation constants describing this interaction are:

\[
K_1 = \frac{[DB][DNA]}{[DB:DNA]} \\
K_2 = \frac{[DB:DNA][DB]}{[DB_2:DNA]} = \omega K_1
\]  

where \(K_2\) may alternatively be expressed by the cooperativity parameter \(\omega\). In addition to the equilibrium constants, the system is constrained by the following equations of state for both compound and DNA:

\[
\begin{align*}
[DB]_i &= [DB] + [DB:DNA] + 2[DB_2:DNA] \\
[DNA]_i &= [DNA] + [DB:DNA] + [DB_2:DNA]
\end{align*}
\]  

(2.8)
In direct titrations of compound by DNA, since the observed anisotropy change represented the summed contributions of the two complexes, the most efficient approach is to build the binding polynomial in terms of the unbound compound. The solution, which is cubic in [DB], is:

\[
0 = \varphi_0 + \varphi_1[\text{DB}] + \varphi_2[\text{DB}]^2 + \varphi_3[\text{DB}]^3
\]

\[
\begin{align*}
\varphi_0 &= K_1K_2[\text{DB}]_t \\
\varphi_1 &= -K_1K_2 - K_2[\text{DNA}]_t + K_2[\text{DB}]_t \\
\varphi_2 &= -2[\text{DNA}]_t - K_2 + [\text{DB}]_t \\
\varphi_3 &= -1
\end{align*}
\] (2.9)

[DB] was solved numerically from Eq. (2.9) using the cubic solver routine (c02akc) provided by the NAG Library in the Origin 2018b environment (Northampton, MA), rather than analytically via the cubic formula, to avoid failure due to loss of significance. With [DB] in hand, [DB:DNA] and [DB2:DNA] were computed from the equilibrium expressions (2.7) and the corresponding equations of state (2.8). The fraction bound (F_b) for each complex is then given by:

\[
F_{b,1:1} = \frac{[\text{DB:DNA}]}{[\text{DB}]_t}
\]

\[
F_{b,2:1} = \frac{[\text{DB2:DNA}]}{[\text{DB}]_t}
\] (2.10)
### 2.7.2 Supplemental table

**Table 2.1 Goodness of fit of Hill analysis of DNA binding by linear indole-biphenyl cations.** The relative goodness of fit of a one- vs. two-term Hill equation to titration data in Figure 2.2 in the main text using Fisher’s $F$-test on the residual sum of squares, setting $p = 0.05$ per pairwise comparison. Testing of multiple hypotheses ($n$) for each compound over the full salt range is conservatively controlled by applying Bonferroni’s correction to the family-wise error rate $\alpha = p/n$.

<table>
<thead>
<tr>
<th>[NaCl, M]</th>
<th>Residual Sum of Squares</th>
<th>$F$-value</th>
<th>$p$-value</th>
<th>Significant at $\alpha = p/n$?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single-term</td>
<td>Two-term</td>
<td></td>
</tr>
<tr>
<td>DB1804</td>
<td>0.01</td>
<td>1.0 × 10²</td>
<td>3.3</td>
<td>1.3 × 10²</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>14</td>
<td>1.0</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>14</td>
<td>1.0</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>0.150</td>
<td>20</td>
<td>1.8</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>2.3 × 10²</td>
<td>22</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>0.250</td>
<td>13</td>
<td>1.0</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0.750</td>
<td>59</td>
<td>6.3</td>
<td>33</td>
</tr>
<tr>
<td>DB1883</td>
<td>0.005</td>
<td>17</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>44</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>8.7</td>
<td>8.7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.065</td>
<td>5.6</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>11</td>
<td>11</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.750</td>
<td>12</td>
<td>9.3</td>
<td>1.5</td>
</tr>
<tr>
<td>DB2627</td>
<td>0.005</td>
<td>7.9</td>
<td>7.3</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>1.2</td>
<td>0.77</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>6.6</td>
<td>6.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.150</td>
<td>35</td>
<td>20</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>2.8</td>
<td>2.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.750</td>
<td>2.4 × 10⁻⁵</td>
<td>8.5 × 10⁻⁶</td>
<td>4.7</td>
</tr>
<tr>
<td>DB1994</td>
<td>0.005</td>
<td>14</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>1.2</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>6.8</td>
<td>6.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.080</td>
<td>5.0</td>
<td>3.4</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>4.6</td>
<td>2.3</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>0.750</td>
<td>5.8</td>
<td>4.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>
2.7.3 Supplemental figures

Figure 2.11 Size-exclusion chromatography of A₂T₂ in the absence and presence of DB1804.
DNA (5 μM) in the absence or presence of an equimolar concentration of DB1804 was eluted isocratically in Tris-buffered saline from a Superdex 75 100/300 GL column (GE) at 0.5 mL/min. under the control of a Bio-Rad NGC FPLC instrument. The eluate was monitored by UV absorption at 260 nm (black) and 370 nm (blue) to follow DNA and DB1804, respectively. The elution volumes of the DNA with and without DB1804 were 13.2 mL and 13.1 mL, respectively.
Figure 2.12 Effect of DMSO and nicotinamide on DNA binding by DB1804.
DNA titrations of $A_2T_2$ into 20 nM DB1804 at 1 M NaCl in the presence of 10% and 20% (v/v) DMSO ($A$, purple circles and triangles) and in the presence of 0.05 and 0.10 M nicotinamide ($B$, orange circles and triangles). The titration data from Figure 2A in the main text at 750 mM NaCl in the absence of co-solutes is shown in gray for reference. In general, the presence of co-solutes perturbed the observed anisotropy values (particularly unbound compound) in accordance with their known effects on solution viscosity (nicotinamide > DMSO, [112, 113]). Drop lines in Panels $A$ and $B$ indicate $K_{D,1}$ (dashed lines) and $K_{D,2}$ (dotted lines) estimated by Hill analysis i.e., Eq. (2.3) in the main text. $C$, Stoichiometric analysis for the first transition for DB1804/A$\_2T_2$ binding in the presence of 0.10 M nicotinamide at 750 mM NaCl shows that it represents the 2:1 DB1804:A$\_2T_2$ complex. The error bars from Panel $C$ have been omitted for clarity.
3 STRUCTURE-BINDING RELATIONSHIPS OF DNA RECOGNITION BY INDOLE-BIPHENYL MONOCATIONS

3.1 Abstract

The linear indole-biphenyl dications DB1804 and DB1883 were recently shown by us to target the minor groove of an A2T2 binding site as dimers in a 2:1 complex that is sensitive to its cosolute environment. The stacked-dimer model proposed for these linear dications may pre-organize them into an orientation that is more isohelical with the minor groove. By contrast, DB1944 and DB2627, the respective monocationic analogues of DB1804 and DB1883, also recognize the A2T2 site but were found to do so as monomers. Given our ongoing interest in the role of hydration in DNA binding by similar compounds, together with previous reports of interfacial water molecules facilitating minor groove binding by completing the curvature of non-isohelical compounds, we wondered whether a link exists between heterogeneity in DNA binding by the linear monocations and their hydration profiles. We therefore evaluated the electrostatic and hydration properties of DNA binding by the indole-biphenyl monocations DB1944, DB2627, and three novel monocationic analogues (DB2782, DB2783, DB2784) having various substitutions at the cationic terminus. While an inverse relationship between hydration release and binding affinity was observed for some of the compounds, our results imply that hydrophobicity and bulkiness are at least as important in determining binding affinities for these linear indole-biphenyl monocations.

3.2 Introduction

Despite ongoing interest over the last several decades in the physicochemical properties of DNA minor groove binders, and continued refinement of our understanding of their structure-activity relationships, progress in the context of hydration as a structural feature (or as a
consequence accompanying other structural features) has been less prolific. The studies that have been reported in this area, however, have had great impact. In a collection of paradigm-shifting examples, the presence of interfacial hydration waters was found to facilitate minor groove binding by linear compounds, with the water molecules acting as a molecular prosthetic to lend the compound curvature matching that of the minor groove [33, 94, 120]. Prior to these reports, isohelicity with the minor groove was thought to be an essential determinant of binding affinity. As these studies demonstrate, a more thorough understanding of the role of hydration in DNA minor groove binding is critical for an accurate description of the factors influencing binding heterogeneity. From an applications perspective, the development of structure-hydration relationships can be exploited as an additional degree of freedom with which to confer target specificity in the design of novel minor groove binding compounds. As several such compounds are currently being investigated for their potential as pharmaceutical agents [21, 114, 121], the ability to direct binding more sensitively and specifically is likely to have important translational implications.

In a recent attempt to characterize the hydration properties of DNA minor groove binding by four related mono- and dications sharing a linear indole-biphenyl scaffold, we uncovered unexpected binding modes for two of the compounds. More specifically, we determined that the dications (but not the monocations) preferentially bind to DNA as dimers in a 2:1 complex that is sensitive to its cosolute environment [83]. This interesting finding adds to a growing body of information about the implications of compound structural variations on heterogeneity in DNA recognition mechanisms. Nevertheless, the self-association of the dications precluded insights into their hydration properties by the methods used. Returning now to our original aim, we report here an investigation on structure-hydration relationships for a set of five internally consistent DNA
minor groove binding compounds varying in structure at the cationic terminus only (Table 3.1). As dimerization appears to be a feature limited to the dicationic variety of linear indole-biphenyl compounds, those chosen for the present study are all monocationic variants of the same indole-biphenyl core investigated previously.

Table 3.1 List of compounds investigated.
Excitation and emission wavelengths are reported at the maxima recorded from independent measurements of fluorescence spectra of the compounds. Abbreviations and graphical representations used throughout the text are shown below for reference.

<table>
<thead>
<tr>
<th>Cmpd.</th>
<th>Structure</th>
<th>Ex/Em WL (nm)</th>
<th>Abbreviation</th>
<th>Graphical Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB1944</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>338/454</td>
<td>Am</td>
<td><img src="image2.png" alt="Graphical Representation" /></td>
</tr>
<tr>
<td>DB2627</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>337/438</td>
<td>THP</td>
<td><img src="image4.png" alt="Graphical Representation" /></td>
</tr>
<tr>
<td>DB2782</td>
<td><img src="image5.png" alt="Structure" /></td>
<td>336/445</td>
<td>iPr-Am</td>
<td><img src="image6.png" alt="Graphical Representation" /></td>
</tr>
<tr>
<td>DB2783</td>
<td><img src="image7.png" alt="Structure" /></td>
<td>337/430</td>
<td>DM-THP</td>
<td><img src="image8.png" alt="Graphical Representation" /></td>
</tr>
<tr>
<td>DB2784</td>
<td><img src="image9.png" alt="Structure" /></td>
<td>348/452</td>
<td>2-Im</td>
<td><img src="image10.png" alt="Graphical Representation" /></td>
</tr>
</tbody>
</table>

3.3 Materials and Methods

3.3.1 Compounds and DNA

The synthesis of DB1944 and DB2627 have been previously described [83, 94]. The synthesis of DB2782, DB2783, and DB2784 will be published elsewhere. All compounds were dissolved in water to concentrations of 2 to 80 μM. The DNA dodecamer 5’–CGCGAATTTCGCG–
3’ (“A₂T₂”) was synthesized by Integrated DNA Technologies (Midland, IA). Lyophilized DNA was dissolved in 10 mM TrisHCl (pH 7.5) containing 1.0 M NaCl, annealed, and dialyzed (MWCO 3.5 kDa) extensively against 10 mM TrisHCl (pH 7.5). Duplex DNA concentration was measured by UV absorption at 260 nm using the nearest-neighbor extinction coefficient 191,511 M⁻¹ cm⁻¹. All other reagents were obtained at ACS grade or higher purity and used without further purification with the exception of stock solutions of NaCl and KCl, which were filtered through a 0.45 μm syringe filter before use.

### 3.3.2 Steady-state fluorescence polarization binding assays

DNA binding by compounds was evaluated at equilibrium via the large change in steady-state fluorescence polarization of the compounds as described previously [83, 122, 123]. Titrations of A₂T₂ into the intrinsically-fluorescent compounds were performed at ambient temperatures (except where otherwise indicated in the text) using a PerkinElmer LS-55 fluorimeter with excitation and emission wavelengths set at the maxima for each compound (Table 3.1). Compound concentrations were between 10⁻⁸ and 10⁻⁹ M and remained constant within no more than ~10% dilution over the course of the titration.

### 3.3.3 Computational methods

Compounds were built and energy minimized in Spartan ’16 or Spartan ’18 using the 6-31G* basis set at the B3LYP DFT level of theory in a water environment (dielectric constant of 78) to obtain the equilibrium ground state geometry and structural information such as polar and nonpolar surface areas. Conformer distributions were calculated with the oB97X-V/6-311+G(2df,2p)[6-311G*] basis set using oB97X-D/6-31G* geometry at the DFT level of theory. Conformers with energies greater than 15 kJ/mol were pruned.
3.4 Results

3.4.1 Hydration changes accompanying DNA binding by monocations

DNA binding by the compounds was monitored by the increase in their fluorescence anisotropies on the addition of DNA. The titrations afforded a range of DNA concentrations spanning about six orders of magnitude. To obtain dissociation constants for the compound-DNA complexes, plots of anisotropy $\langle r \rangle$ as a function of total DNA concentration $[\text{DNA}]_t$ were fitted to a one-site binding model

$$
\langle r \rangle([\text{DNA}]_t) = \sum_{i=1}^{n} \frac{[\text{DB}_i]_b}{[\text{DB}_i]} \left( \langle r_i \rangle - \langle r_0 \rangle \right) + \langle r_0 \rangle
$$

(3.1)

where $\langle r_0 \rangle$ is the anisotropy of the unbound compound and $\langle r_i \rangle$ is the intrinsic anisotropy of the compound in the $i$-th bound state having a DNA-bound concentration $[\text{DB}_i]_b$. To estimate the attendant changes in counterion release and preferential hydration, we analyzed log-log plots of binding affinity as a function of mean ionic activity in the context of a linked Wyman function for coupled electrostatics and hydration effects on DNA binding (Figure 3.1B):

$$
-\left( \frac{\partial \log K_D}{\partial \log a_\pm} \right) = \Delta n_\pm - \frac{2m_\pm}{55.5 \ln 10} \Delta n_w
$$

(3.2)

where the gradient of the dissociation constant $K_D$ with respect to mean ionic activity $a_\pm$ corresponding to a salt molality of $m_\pm$ reflects the number and disposition of counterions ($\Delta n_\pm$) and waters of preferential hydration ($\Delta n_w$) on DNA binding by compound. Data for all five compounds was fitted globally with the counterion displacement parameter $\Delta n_\pm$ shared across the data sets. To further validate our experimental salt slope, which is expected to be independent of specific ion effects if in accordance with the molecular picture of polyelectrolyte theory, we repeated the salt perturbation titrations in KCl with a model compound. We chose DB2627 for this purpose on the
A, Representative steady-state fluorescence polarization titrations of A2T2 into compound at 15 mM (squares), 200 mM (circles), and 2.5 M NaCl (triangles; in the case of DB1944, NaCl concentration was 2.4 M). Data is fitted to Eq. (3.1). B, Summaries of salt-dependence of DNA binding measured in NaCl (closed circles, solid lines) and KCl (open circles, dashed line). Data is fitted globally to Eq. (3.2) with the shared parameter $\Delta n = -1.14 \pm 0.04$. The concave upward curvature indicates a net release of preferential hydration on complex formation. C and D, Comparison of binding affinities ($K_0$ in NaCl) and hydration numbers, respectively, extracted from the fits to the data in Panel B.

basis of its binding affinity to A2T2 (which was the strongest of all the compounds tested) and its behavior in the absence and presence of DNA having been previously characterized [83]. The resulting salt slope of $\Delta n = -1.14 \pm 0.04$ was somewhat higher than theoretical predictions (-0.75 for the binding of a monocation to a 12-bp oligonucleotide, [53]). The excess counterion number of -0.39 likely represents a fractional contribution from the release of Cl$^-\$ anions associated with the positively-charged termini of the unbound compounds. Binding by the monocations spanned ~1.5 orders of magnitude (Figure 3.1C) and captured variations in the disposition of preferential hydration on binding at a resolution of ~2 water molecules (Figure 3.1D). As a point of comparison for the binding affinities of the five compounds, we chose to look at the formal value log $K_0$, which is the constant of integration obtained through integration of Eq. (3.2) and represents the logarithm of the affinity at near unit salt activity:
\[-\log K_P = \log K_0 + \Delta n_{w} \log a_{w} - \Delta n_{w} \frac{2m_{w}}{55.5 \ln 10}.\] (3.3)

DNA binding by all of the compounds involved a net release of preferential hydration ($\Delta n_{w} < 0$), but agreement in the sign of $\Delta n_{w}$ was matched by variations in its magnitude. The hydration spectrum was bounded by the highest-affinity compound, DB2627, at the low end ($\Delta n_{w} = 25 \pm 2$) and the lowest-affinity compound, DB2783, at the high end ($\Delta n_{w} = 32 \pm 2$). Except for these two extremes, the differential hydration properties of the full set of compounds in the context of their rank-ordered binding affinities revealed no straightforward relationship between the two parameters, suggesting instead that both hydration and binding affinity are governed by a more complex interplay of structural and physicochemical properties.

### 3.4.2 Thermodynamics of DNA binding by linear monocations

The binding free energy $\Delta G$ is a combination of various enthalpic and entropic contributions. Assuming these contributions are additive and barring any additional factors, $\Delta G$ can be modeled as a sum of the following energetic contributions:

$$\Delta G = \Delta H - T\Delta S = \Delta H_{\text{bond}} + \Delta H_{\text{vdW}} + \Delta H_{\text{el}} + \Delta H_{\text{CC}} + \Delta H_{\text{SA}} - T\left(\Delta S_{\text{w}} + \Delta S_{\text{sam}} + \Delta S_{\text{conf}} + \Delta S_{\text{rot}}\right).$$ (3.4)

Major contributions to the binding enthalpy include hydrogen-bonding interactions between donor atoms on the compound and acceptor atoms on the DNA bases ($\Delta H_{\text{bond}}$), van der Waals interactions between compound and the floor of the minor groove ($\Delta H_{\text{vdW}}$), electrostatic interactions between the negatively-charged phosphate backbone and the cationic terminus of the compound ($\Delta H_{\text{el}}$), a salt-dependent electrostatic enthalpy term from counterion condensation ($\Delta H_{\text{CC}}$), and the burial of hydrophobic surface area ($\Delta H_{\text{SA}}$). The terms $\Delta H_{\text{el}}$ and $\Delta H_{\text{CC}}$ are not expected to differ considerably between the five compounds. Theoretical predictions for $\Delta H_{\text{CC}}$ at 25 °C, from calculations based on both a Poisson-Boltzmann model (which has not been previously discussed in this thesis but is
described in [124]; the interested reader is referred to [125] for an in-depth comparison between the Poisson-Boltzmann and CC models) and Manning’s counterion condensation model (discussed in Section 1.2.1; for very detailed discussion, see Refs. 44 and 51 therein as well as [126]), give a low-salt (1 μM) estimate of around -9 kJ/mol which increases linearly to around -2 kJ/mol at ~0.6 M salt [127].

The major favorable entropic contributions arise from the release of water into bulk solvent (ΔS_w) and an entropy of mixing from the release of Na⁺ counterions into bulk solvent (ΔS_ion). As the conformational freedom of a compound in the bound state is expected to be constrained relative to that in the unbound state, binding to DNA presumably imposes a configurational entropic penalty, ΔS_conf. Similarly, a loss of translational degrees of freedom on complex formation imposes an entropic penalty of ca. -6 to -15 kJ/mol at 25 °C by way of a salt-dependent cratic term, ΔS_crat [128, 129]. From our salt-perturbation data, we could obtain experimental values for ΔG from the relationship ΔG = RT ln(K_B) and approximate values for ΔS_w given an entropic gain of about +8 kJ/mol per water molecule at 25 °C [130]. The entropic gain from the release of condensed counterions ΔS_ion can be estimated according to polyelectrolyte theory from

\[ ΔS_\text{ion} = -ZR \ln \left[ \text{Na}^+ \right] \]  

where Z = 1 for a monocationic compound. Values for these and other parameters are listed in Table 3.2. The remaining contributions to the binding free energy could not be determined without additional experimental and computational procedures. There was thus an energy contribution

\[ ΔE = ΔH_{\text{bond}} + ΔH_{\text{vdW}} + ΔH_{\text{el}} + ΔH_{\text{SA}} - TΔS_{\text{conf}} \]  

for which we could not account from our experiments alone (Figure 3.2).
Table 3.2 Comparison of thermodynamic parameters.
Values of $\Delta G$ and $-T\Delta S_{\text{ion}}$ are given for the 50 mM salt condition. Energies are given in units of kJ/mol. The contribution to the binding free energy that remains unaccounted for is given by $\Delta E$. Values of $\Delta \Delta G$ and $\Delta \Delta E$ are relative to DB1944.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G$</th>
<th>$\Delta n_w$</th>
<th>$-$T$\Delta S_w$</th>
<th>$-$T$\Delta S_{\text{ion}}$</th>
<th>$\Delta E$</th>
<th>$\Delta G$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB1944</td>
<td>$-44.2 \pm 0.2$</td>
<td>$-29 \pm 2$</td>
<td>$-229 \pm 16$</td>
<td>$-16$</td>
<td>201</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DB2627</td>
<td>$-47.8 \pm 0.3$</td>
<td>$-25 \pm 2$</td>
<td>$-198 \pm 16$</td>
<td>$-16$</td>
<td>167</td>
<td>$-3.6$</td>
<td>$-35$</td>
</tr>
<tr>
<td>DB2782</td>
<td>$-44.9 \pm 0.2$</td>
<td>$-32 \pm 2$</td>
<td>$-252 \pm 13$</td>
<td>$-16$</td>
<td>224</td>
<td>$-0.7$</td>
<td>$+23$</td>
</tr>
<tr>
<td>DB2783</td>
<td>$-41.1 \pm 0.2$</td>
<td>$-32 \pm 2$</td>
<td>$-257 \pm 13$</td>
<td>$-16$</td>
<td>232</td>
<td>$+3.1$</td>
<td>$+31$</td>
</tr>
<tr>
<td>DB2784</td>
<td>$-46.5 \pm 0.2$</td>
<td>$-29 \pm 2$</td>
<td>$-234 \pm 18$</td>
<td>$-16$</td>
<td>204</td>
<td>$-2.3$</td>
<td>$+2.8$</td>
</tr>
</tbody>
</table>

Figure 3.2 Thermodynamic properties of biphenyl-indole monocations in complex with DNA.
The free energy of binding to DNA (light gray) includes a large favorable entropic contribution from the attendant release of preferential hydration waters and condensed counterions (blue). Based on the magnitudes of $\Delta G$ and the hydration changes on binding, complex formation by the monocations is driven by additional, substantial enthalpic and/or entropic contributions, the sum of which poses an energetic penalty to binding (red).
Seeking nevertheless to better understand the structural implications of the compounds on their DNA binding energetics, we turned our attention toward the physicochemical properties of the compounds in their unbound states as a possible explanation for their differential behaviors in the bound state. Namely, in the absence of experimental values for $\Delta \Delta H_{SA}$ and $\Delta \Delta S_{conf}$ we compared the hydrophobic surface areas (Figure 3.3) and the number of relative configurational degrees of freedom of each compound (Figure 3.4). We calculated the hydrophobic surface areas from the total surface areas (from the space-filling models) and polar surface areas (defined as the area due to N, O, and hydrogens attached to either N or O) generated in Spartan for the energy-minimized compounds. As expected, the hydrophobic surface area of each compound compared to the others matched the relative number of C atoms at its cationic terminus. The general tendency of heterocyclic minor groove binding compounds is to be inserted deep into the DNA minor groove. Therefore, with the exception of DB1782, the iPr group of which is free to rotate such that it faces out of the minor groove, the positioning of the compounds deep in the groove is expected

![Figure 3.3 Total, polar, and hydrophobic surface areas of the compounds in their ground-state equilibrium geometries.](image)

The burial of hydrophobic surface area confers a favorable enthalpic contribution to $\Delta G$ and may be more significant for DB2627, DB2782, and DB2783, which have larger hydrophobic surface areas in the unbound state.
to result in a non-negligible burial of hydrophobic surface area. Given that the burial of hydrophobic surface area has been reported to confer a favorable energetic contribution of about \(-63\ \text{J/mol per \(\AA^2\)}\) [131], even the modest increase in hydrophobic surface area for DB2784 over DB1944 (+50 \(\AA^2\), the burial of which corresponds to an energetic gain of -3 kJ/mol) could, all else being equal, potentially account in full for the difference in binding energies between the two compounds.

**Figure 3.4** *The relative number of conformers available to each compound in the unbound state.*

Conformer distributions were calculated based on the number of rotatable bonds and ring flip conformations. Conformers with energies \(\leq 15\ \text{kJ/mol}\) were kept. Aromatic bonds (and, thus, the indole-biphenyl scaffold common to all the compounds) are treated as non-rotatable. Therefore, the various conformers shown here arise only from configurational freedom at the cationic termini of the compounds. The indole-biphenyl motifs are omitted from the images for clarity. Barring large differences between the configurational degrees of freedom of the DNA-bound complexes, higher configurational entropy in the unbound state may be indicative of a more significant unfavorable contribution to \(\Delta G\) due to a more pronounced loss of configurational freedom.
As the compounds all share the same phenyl-phenyl-indole scaffold, the most significant difference in configurational freedom arises from their different cationic termini. Since greater configurational freedom lends to a larger sampling of various available microstates, a statistical mechanical treatment gives a configurational entropy $S_{\text{conf}}$ according to the Boltzmann definition

$$S_{\text{conf}} = k_B \ln \Omega_{\text{conf}} \quad (3.6)$$

for a compound whose geometry allows it to occupy a number of configurational microstates $\Omega_{\text{conf}}$.

While the scope of the present study eludes such quantitative descriptions of each compounds’ configurational microstates for either the unbound or the DNA-bound states, a qualitative comparison for the unbound state is quite readily ascertained. At the most basic level, compounds harboring a ringed structure at the terminus (DB2627, DB2783, DB2784) are more rotationally constrained than compounds with a non-cyclic terminus (DB1944, DB2782). Among the ringed-terminus compounds, those with six-membered rings (DB2627, DB2783) have greater conformational freedom than DB2784 with its five-membered ringed terminus. Finally, the isopropyl group emanating from one of the amidine N atoms of DB2782 confers increased configurational freedom not only due to the greater number of rotationally-unconstrained bonds but also by virtue of its asymmetric structure (i.e., positioning of the isopropyl on either face of the compound relative to the indole N).\(^4\) To allow a semi-quantitative comparison of relative configurational entropies, we used as a first approximation the number of conformers of the cationic termini obtained from an in silico conformer distribution search (Figure 3.4). Since the search treats aromatic bonds as non-rotatable, these values ignore rotation about the bonds connecting the phenyl, indole, and amidine (or substituted amidino) groups. As expected, the iPr-

---

\(^4\) We note that another consequence of the structure of DB2782 is that the missing degree of symmetry relative to the other compounds (which all have symmetric termini) imposes an additional entropic penalty on binding that is not present for the other compounds.
Am of DB2782 has the greatest number of conformers (four), while the unsubstituted amidine of DB1944 and the 2-Im of DB2784 each have only a single conformer. While the absolute number of conformers obtained in this way is not an accurate reflection of the various low-energy microstates each compound might occupy, it nonetheless offers a good point of comparison for the most significant differences in the configurational flexibility of the unbound compounds.

3.5 Discussion

The role of water in facilitating DNA minor groove binding by small compounds is an interesting question to study from both a biophysics perspective and a translational perspective given the promise of some such compounds as drugs and the highly hydrated environment of physiological systems. Another area of ongoing investigation attempts to define the intricate relationships between the structures of such compounds and their DNA binding activities. Here, to provide further insight on both of these questions, we looked at the DNA-binding properties of a set of five linear, indole-biphenyl monocations identical in structure except for variations at the cationic terminus. Our results point to the release of preferential hydration as a significant driving force for binding as well as to the presence of a large energetic penalty of between +170 to +230 kJ/mol, depending on the compound, that compensates for the very large entropic gain of hydration and counterion release relative to the overall free energy of binding. The incongruent rank orders for binding affinity and preferential hydration release among the various compounds imply that the other contributions to the binding free energy differ between the compounds and, in this way, suggests a structure dependence on the interplay of factors that influence binding.

3.5.1 Structure-binding relationships of indole-biphenyl monocations

The structure-binding relationship most apparent from our data is that of hydrophobicity. We found, for some of the compounds, a correlation between their hydrophobic surface areas and
their binding affinity to DNA in a way that was independent from the trend in preferential hydration. Namely, the hydrophobic surface area of DB2627 is higher than that of DB2784, which in turn is higher than that of DB1944; this trend follows the rank order for binding affinity among these compounds: DB2627 > DB2784 > DB1944. Perhaps significantly, all three of these compounds have in common a structure that allows the molecule as a whole to adopt a roughly planar geometry.

The trend between hydrophobicity and binding affinity did not hold for DB2782, which has moderately higher hydrophobic surface area than DB2627 but binds more weakly than either DB2627 or DB2784, and DB2783, which has the highest hydrophobic surface area but binds with lowest affinity to DNA. One possible explanation is that the configurationally-constrained dimethyl group emanating from the THP ring of DB2783 results in a geometry that is not easily accommodated by the DNA minor groove. Meanwhile, since the isopropyl group of DB2782 rotates freely, DB2782 is likely to have a greater number of DNA-bound configurations, some of which would place the isopropyl in an orientation approximately in-plane with the rest of the molecule (Figure 3.5). This geometry is inaccessible to DB2783 due to the rotationally-constrained dimethyl group (Figure 3.5), lending to a lower binding affinity relative to DB2782. This explanation, which rests also on the ability of DB2782 to adopt a number of less-favorable DNA-bound configurations (e.g., with the isopropyl positioned out-of-plane with the rest of the molecule), would also account for the lower binding affinity of DB2782 relative to DB2784 and DB2627.

DB2782 is unique among this set of five compounds as the only one for which rotation about the amidine alters the positioning of the attached hydrocarbon groups relative to the minor groove. In one orientation, the iPr of DB2783 faces into the minor groove, while in another it faces
outward. Since the outward-facing orientation exposes a larger hydrophobic surface area to bulk solvent than does the first, the distribution of the two orientations clearly affects the contribution to $\Delta G$ from $\Delta H_{SA}$. If the burial of hydrophobic surface area confers a large favorable contribution to the binding free energy, then one would expect the most favorable DNA-bound conformation for DB2783 to have the iPr facing into the minor groove. Interestingly, DB2782 displaces the same number of waters on binding to DNA as does DB2783, both of which displace more waters of preferential hydration than any of the other compounds, which may support a groove-facing orientation for the iPr group of DB2782. This orientation is likely to be further favored by enabling van der Waals interactions with the floor of the minor groove that would be inaccessible to an outward-facing iPr group.

Figure 3.5 **DB2783 is bulkier than DB2782.** The compounds are depicted as stick structures and space-filling models and shown from two views. Rotation of the iPr group of DB2782 allows the molecule as a whole to adopt a more planar geometry. By contrast, the dimethyl group in DB2783 is rotationally constrained, lending to the overall greater bulk of the compound.
3.5.2  Intricacies of structure effects on DNA minor groove binding

The present study offers an interesting comparison with previous investigations of minor groove binding compounds having similar substitutions at their cationic termini but different charge numbers (dicationic rather than monocationic) and linking structures between the cationic termini (Figure 3.6). The rank order we report here for binding affinity between the DB1944 (the Am compound) and DB2782 (the iPr-Am) is in agreement with that reported for a dicationic Am compound and a dicationic iPr-Am compound whose cationic termini flank an aromatic phenylene.

---

**Figure 3.6** Comparison of binding affinity rank orders for similar compounds in the literature.

Compounds are represented by their cationic termini, with nitrogens indicated by light blue circles. **Row A**, Relative binding affinities to A₂T₂ for the monocationic compounds used in this study. **Rows B and C** show relative binding affinities to poly(dA)-poly(dT) for a series of dicationic compounds from the study in Ref. 121, while **Row D** shows the same for a series of dicationic pentamidine analogues from the study in Ref. 122. The linker between the cationic termini of compounds in Row B and the core indole-biphenyl structure of compounds in Row A are continuously aromatic, while the linkers between the compounds in Rows C and D are not.
bis(amidinobenzimidazole) linker [132]. Interestingly, this rank order almost completely reverses itself for the Am, iPr-Am, 2-Im, and THP compounds when the linking structure is non-aromatic [132, 133]. It should be noted that hydrophobic effects are likely less significant for dicationic compounds compared to the monocationic compounds studied herein, which may also influence the differences between the rank orders. Though the present investigation was focused on structure-binding relationships in the context of the structure of the cationic terminus only, when taken together with these previous studies, it adds to the existing body of evidence for the influence of the rest of the molecule’s structure on the dependence of binding on substitutions at the termini.

3.5.3 A role for water beyond preferential hydration

Since perturbation by salt as a means of probing hydration changes captures only changes in preferential hydration, this type of experiment yields an incomplete estimate of the full hydration picture. One might envision, for example, the formation of water-mediated hydrogen bonds between compound and DNA that pose both an entropic penalty and an enthalpic gain, neither of which would have been accounted for by our present analysis. This type of structural water bridge has been reported for some non-isohelical compounds as a means of lending increased curvature to match that of the minor groove [33, 94, 120]. As the compounds studied herein are also linear, it is possible that complex formation by these monocations is likewise mediated by interfacial waters that play a structural role in facilitating minor groove binding. The implications, if any, of the various terminal substitutions on this type of hydration (and, in turn, on ΔH_{bond}) are unclear at present. Future studies implementing the use of calorimetric techniques to measure binding to DNA sequences having a range of minor groove widths and in solvents of various hydrogen-bonding cohesiveness (e.g., H₂O vs. D₂O) may prove useful in teasing apart the contributions to the binding free energy from ΔH_{bond}, ΔH_{vdW} and (as a probe for ΔH_{SA}) ΔCₚ.
Together with molecular dynamics simulations to determine $\Delta S_{\text{conf}}$, the complete thermodynamic and hydration profile could be comprehensively determined for DNA binding by these linear indole-biphenyl monocations.

3.6 Conclusion

Using NaCl to perturb DNA binding by five structurally-related biphenyl-indole monocations, we probed the influence of their structures on the hydration and electrostatic properties of their DNA-bound complexes. In general, high-affinity binding seemed to be favored by the addition of hydrocarbon groups to the cationic terminus so long as the presence of these groups did not cause drastic deviations from the overall planarity of the rest of the molecule. In other words, compounds with increased hydrophobic surface area but minimally increased bulkiness (thereby allowing them to still be readily accommodated by the narrow minor groove) bind more strongly. This finding highlights the importance of hydrophobic and van der Waals interactions, alongside the energetically-favorable release of preferential hydration, in facilitating DNA minor groove binding by linear, heterocyclic monocations.
4 SEQUENCE DEPENDENCE OF HYDRATION IN DNA BINDING BY AN
ISOHELICAL DIAMIDINE

4.1 Abstract

DB1976 is a symmetric selenophene bis-benzimidazole diamidine with demonstrated pharmaceutical activity against leukemia in cellular and mouse-model systems. In vitro, DB1976 binds A/T-rich DNA at the minor groove with nM affinity and has a bright, intrinsic fluorescence that largely persists on binding to DNA. It is therefore an ideal model compound for studying the physicochemical properties of drug-like DNA minor groove binders. In a previous investigation of the electrostatic and hydration properties of DB1976 [82], we reported a net release of preferential hydration accompanying its binding to the A₂TA₂ minor groove of the physiologically-relevant DNA sequence 5’–GCGAATAAGAGGAAATGACG–3’. Since compounds like DB1976 tend to recognize A/T-bp sequences preferentially over mixed-bp sequences, we asked whether the release of preferential hydration would be correlated with sequence specificity. To explore this question, we interrogated the hydration properties of DB1976 complexes with three DNA dodecamers differing in GC content and bp positions via osmotic stress experiments.

4.2 Introduction

Small heterocyclic diamidines are emerging as useful modulators of transcriptional activity in vivo and, as such, present an attractive therapeutic approach for diseases arising from transcriptional deregulation. Since insufficient target recognition specificity by such compounds would preclude their use in the clinic, there is an ongoing, collaborative effort to understand the physicochemical basis of their DNA recognition properties. One such property is molecular hydration, which is now recognized to play an important thermodynamic and structural role in
facilitating DNA binding by compounds [94, 120] as well as by the DNA-binding proteins these compounds may be capable of inhibiting [134, 135]. Given the growing body of evidence for molecular hydration in conferring target specificity and as a driving force for high-affinity binding, we became interested in exploring the role of water in DNA binding by DB1976 (Figure 4.1), a model bis(benzimidazole) diamidine with demonstrated inhibitory potential against the transcription factor PU.1, a master regulator of hematopoiesis [21, 123].

We previously reported the preferential hydration changes accompanying DNA binding by DB1976 to a 23-bp oligodeoxynucleotide containing the 5′–AGAGGAAGTG–3′ consensus site for PU.1. Having established a net release of preferential hydration waters for DB1976 binding to the minor groove of this physiologically important DNA sequence [82], we asked whether the magnitude of this hydration release would be conserved among various DNA targets or whether
hydration might play a role in conferring sequence-specificity in DNA recognition by DB1976. To that end, we interrogated the hydration properties of the complexes formed by DB1976 with three oligodeoxynucleotides containing various GC contents and AT-tract lengths (Figure 4.1) by way of salt-perturbation osmotic stress studies.

4.3 Materials and Methods

4.3.1 Compound and DNA

The synthesis of DB1976 has been previously described [115]. DNA oligomers were synthesized by IDT (Midland, IA) and obtained as a lyophilized solid. Lyophilized DNA was dissolved in 10 mM sodium cacodylate buffer (pH 6.8) containing 1.0 M NaCl, annealed, and exhaustively dialyzed against the same buffer absent additional NaCl. DNA concentrations were determined by UV absorption at 260 nm using nearest-neighbor extinction coefficients of 186,075 M⁻¹cm⁻¹ for 5’–CGAAATTTGCG–3’ (“A₃T₃”), 191,511 M⁻¹cm⁻¹ for 5’–CGCAATTTCGCG–3’ (“A₂T₂”), and 190,127 M⁻¹cm⁻¹ for 5’–CGCAACGTTGCG–3’ (“A₂CGT₂”) [136].

4.3.2 Fluorescence polarization titrations

DNA binding by DB1976 was determined as previously described [82]. In brief, equilibrium titrations of A₃T₃, A₂T₂, and A₂CGT₂ into 10⁻⁹ M DB1976 (Ex/Em = 365/445 nm) were performed in 10 mM sodium cacodylate containing various concentrations of NaCl at ambient temperature using a PerkinElmer LS-55 fluorimeter. The large change in the steady-state fluorescence anisotropy of DB1976 upon binding to DNA as a function of total titrant concentration was fitted to a 1:1 binding model from which dissociation constants were extracted.

4.4 Results

We investigated the hydration properties of DNA binding by the heterocyclic diamidinium DB1976 (dicationic at pH 7) to the well-studied dodecamer 5’–CGCAATTTCGCG–3’ (A₂T₂) and
two additional dodecameric analogues, A₃T₃ (5’–CGCAATTTGC–3’) and A₂CGT₂ (5’–CGCAACGTGCG–3’). A₃T₃ introduces a longer AT-tract, expected to afford increased binding affinity to DB1976, while A₂CGT₂ is an isomer of A₂T₂ in which the AT-tract is interrupted by a CG and is therefore expected to be less readily bound by DB1976. Steady-state fluorescence polarization titrations of DNA into DB1976 at varying ionic strengths (Figure 4.2A) captured dissociation constants spanning four orders of magnitude, from ~10⁻¹⁰ to ~10⁻⁶ M. Additional titrations at 50 mM NaCl under depleting conditions confirmed that DB1976 bound each DNA sequence with 1:1 stoichiometry (Figure 4.2B). At all NaCl concentrations tested, DB1976 bound each of the three DNA sequences with affinities matching the expected rank order of A₃T₃ > A₂T₂ > A₂CGT₂ in line with A/T-tract lengths (Figure 4.2C).

4.4.1 Probing preferential hydration in DNA site recognition by osmotic stress

Log-log plots of the binding affinities as a function of mean ionic strength (Figure 4.3) allowed the change in preferential hydration Δnₓw accompanying binding to be estimated in the context of its linkage with the well-established displacement of counterions Δnₓ by

\[-\left( \frac{\partial \log K_D}{\partial \log a_x} \right) = \Delta n_x - \frac{2m_x}{55.5 \ln 10} \Delta n_w \]  \hspace{1cm} (4.1)

where the molal salt concentrations mₓ at which dissociation constants Kₓ are determined have mean ionic activity aₓ. Functionally, the parameters were computed by fitting the data in Figure 4.3 to the integrated form of Eq. (4.1):

\[-\log K_D = \log K_0 + \Delta n_x \log a_x - \Delta n_w \frac{2m_x}{55.5 \ln 10} \]  \hspace{1cm} (4.2)

We obtained a shared counterion number of Δnₓ = -2.2 ± 0.1, in excellent agreement with that obtained from our previous investigation of DNA binding by this compound [82]. While the binding affinities of DB1976 to the three DNA sequences varied drastically, the attendant changes
Figure 4.2 Perturbation with salt shows a net release of preferential hydration for the DB1976/DNA complexes.
A, Representative fluorescence polarization titrations of DNA into DB1976 at NaCl concentrations bracketing and within the range tested (arrows). B, Stoichiometric analysis of DNA binding by DB1976 at 50 mM NaCl shows that all three complexes form with 1:1 stoichiometry. C, Binding affinities (represented by log $K_0$ from fits of the data shown in Figure 4.3 to Eq. (4.2)) for DB1976/DNA binding decrease with decreasing A/T-tract length.
in preferential hydration were remarkably similar to one another. The largest difference observed was between A₃T₃ (Δnₓ = -27 ± 2) and A₂T₂ (Δnₓ = -31 ± 3), the latter of which released only four waters of preferential hydration more than the former. The very low affinity of DB1976 to A₂CGT₂ made it difficult to evaluate binding at very high salt concentrations due to DNA consumption; as such, the full hydration picture for DB1976 binding to this DNA sequence was occluded. Nevertheless, simulated fits to the data strongly suggested a hydration number within the range of -20 > Δnₓ > -40. Taking the experimentally-obtained value of Δnₓ = -30 ± 6 for the A₂CGT₂ complex, the rank order for preferential hydration was A₂T₂ ~ A₂CGT₂ > A₃T₃.

**Figure 4.3** The effect of ionic strength on binding affinity of DB1976 to A₃T₃, A₂T₂, and A₂CGT₂.

Perturbation with salt shows a net release of preferential hydration for the DB1976/DNA complexes. Global fits to the data with a shared ion parameter Δnₓ = -2.2 ± 0.1 gave preferential hydration numbers of Δnₓ = -27 ± 2, Δnₓ = -31 ± 3, and Δnₓ = -30 ± 6 for A₃T₃, A₂T₂, and A₂CGT₂, respectively. Simulated fits to the A₂CGT₂ data with Δnₓ fixed at -10 (light gray dotted line), -20 (light blue dotted line), -30 (blue dashed line), -40 (light blue dashed line), and -50 (light gray dashed line) suggest that extension of the salt-perturbation data to higher salt would have yielded a preferential hydration number between -20 and -40.
4.5 Discussion

DB1976 is an extended diamidine whose structure lends a curvature matching that of the DNA minor groove. Like most compounds in its class, DB1976 preferentially targets AT-rich regions of the minor groove with sub-nM affinities to DNA sequences containing four or more A/T base pairs in a row. To further understand the physical chemistry of DNA sequence recognition by this model compound, we interrogated its binding to three DNA sequence variants (A₃T₃, A₂T₂, and A₂CGT₂) at a range of NaCl concentrations. As a convenient means of comparing the binding affinities of complex formation with the three DNA sequences, we looked at values of log $K_0$ obtained from fitting the salt-dependent binding data to Eq. (4.2), noting that log $K_0$ is the integration constant from integrating Eq. (4.1) and is thus only a formal value. As expected, based on the A/T-tract lengths, binding of DB1976 was strongest to A₃T₃ ($K_0 = 5.4 \times 10^8$ M⁻¹), seven times weaker to A₂T₂ ($K_0 = 7.5 \times 10^7$ M⁻¹), and three orders of magnitude weaker to A₂CGT₂ ($K_0 = 1.3 \times 10^5$ M⁻¹). In light of this trend, we compared our current results to those we reported previously for DB1976 binding to the A₂TA₂ tract of a 23-bp oligonucleotide [82], expecting the rank order for binding affinity to be preserved (that is, A₃T₃ > A₂TA₂ > A₂T₂ > A₂CGT₂). Instead, we noticed that DB1976 binds A₂TA₂ with equal affinity ($K_0 = 6.7 \times 10^7$ M⁻¹) to A₂T₂, despite the longer A/T-tract of the former. This rank order seems to be a consequence of the combined effects of minor groove widths, electrostatic potential in the minor groove, and entropy. As shown in Figure 4.4A and B, the minor groove width and electrostatic potential (predictions from the Rhos lab’s DNAphi tool, [137]) at the center of the A/T-tracts increase in order with decreasing binding affinity for A₃T₃, A₂T₂, and A₂CGT₂. However, both the minor groove width and electrostatic potential of A₂T₂ are smaller than those of A₂TA₂, which would be consistent with a rank order for binding affinity of A₂T₂ > A₂TA₂. Meanwhile, since DB1976 is expected to contact four base-
pairs and the DNA sequences are asymmetric, there are two distinct ways for DB1976 to bind A₂TA₂ but only one way for DB1976 to bind A₂T₂ (Figure 4.4C). From a statistical mechanics perspective, the A₂TA₂ complex has higher entropy than the A₂T₂ complex, which would be consistent with a binding affinity rank order of A₂TA₂ > A₂T₂ and thereby counterbalances the opposing combined effects of minor groove width and electrostatic potential. It is presumably the interplay between these three properties that leads to the observed binding affinity rank order of A₃T₃ > A₂TA₂ ~ A₂T₂ > A₂CGT₂.

**Figure 4.4 Structural properties of the DNA oligonucleotides.** Minor groove widths (A) and electrostatic potential (B) of the DNA sequences at each base position (5' → 3'). Electrostatic potential is reported as multiples of k_BT/e, where k_B is the Boltzmann constant, T is the thermodynamic temperature, and e is the electronic charge. C, Positional microstates available to DB1976 bound to A₃T₃ (three microstates), A₂TA₂ (four microstates due to asymmetry of the full DNA sequence), and A₂T₂ (one microstate). Arrows represent nucleobase contacts with DB1976.
4.5.1 Binding preferences are independent of preferential hydration properties

To probe the changes in preferential hydration coupled to the release of counterions on DNA binding by DB1976, we took an osmotic stress approach in which we perturbed the binding interaction with NaCl. Binding by DB1976 to all three DNA sequences involved a net release of hydration. All three compound/DNA complexes showed similar hydration profiles, with only minor differences in preferential hydration release in the order A2T2 ~ A2CGT2 > A3T3. The value of $\Delta n_w$ for binding to the three DNA sequences ranged between -27 and -31 waters, which is the same as the number we previously reported (-29 ± 2) for DB1976 binding to the longer A2TA2-containing DNA mentioned earlier. Clearly, while the release of preferential hydration is an important entropic driving force for DNA minor groove binding, it plays a very minor role in the differential DNA recognition properties for oligomeric DNA sequences containing A/T-tracts of different lengths.

It should be noted that the osmotic stress method as employed herein uses the well-established phenomenon of counterion condensation and release to infer the coupled change in preferential hydration via the effects of salt perturbation on binding. As such, hydration numbers obtained in this way include only those waters of preferential hydration that are excluded from the entire compound/DNA binding system on complex formation. Moreover, the Wyman linkage model we use to quantitatively evaluate preferential hydration (see Eq. (4.2)) does not account for the contribution to $\Delta n_w$ from the released counterions themselves and assumes that $\Delta n_\pm$ and $\Delta n_w$ remain constant across the full range of salt concentrations (which is non-physical). As discussed in section 1.2.1, experimental data suggest a constant value for $\psi$ and, therefore, $\Delta n_\pm$ (through the relation $\Delta \psi = -Z\psi$ where $Z$ is the ligand charge) at salt concentrations up to ~0.5 M NaCl [36]. Our experiments, however, employ Na$^+$ at concentrations well above 0.5 M (as high as 3.5 M, in
the case of A₃T₃). While experimental values for \( \psi \) and \( \Delta n_\pm \) at such high salt concentrations are lacking, the salt number \( \Delta n_\pm \) most likely increases with salt activity at sufficiently high bulk salt concentration in a manner similar to that with which the theoretical prediction of the salt number \( \Theta \) diverges (see Figure 1.5) [37, 38]. Thus, our reported values of the coupled hydration number \( \Delta n_w \) likely underestimate, in a sequence-dependent manner, the actual release of preferential hydration that occurs on complex formation [138-141]. Further examination of the hydration changes by direct methods such as volumetric measurements and MD simulations, which include the contribution of all hydration events involving changes to the hydrogen bonding network of water, would complement the present study by painting a more complete picture of molecular hydration in the context of DNA site recognition.

4.6 Conclusion

In this investigation, we determined the hydration properties of DB1976 binding to the three DNA sequence variants A₃T₃, A₂T₂, and A₂CGT₂ by osmotic stress experiments using only NaCl. The osmotic stress approach suggested similar patterns of preferential hydration for all three compound/DNA complexes despite drastic differences in binding affinities, which suggests that additional contributions to the binding free energy are a major driving force for the differential binding preferences. One possibility is the formation of water-mediated hydrogen bonding interactions (a favorable enthalpic gain) between DB1976 and A₃T₃ or A₂T₂ – the higher-affinity targets – which are absent in the much lower affinity DB1976/A₂CGT₂ complex. Efforts to explore this and other possibilities are currently underway.
5 OVERALL CONCLUSION

In this thesis, the DNA-binding properties of eight heterocyclic amidines and amidine derivatives were investigated. First, from a set of four linear indole-biphenyl compounds, we found that the two dications, but not their monocationic analogues, bind the DNA minor groove as a dimer in what we propose to be an antiparallel, stacked geometry. As the two monocationic analogues were found to bind DNA in canonical 1:1 fashion, we then investigated their preferential hydration properties alongside three novel indole-biphenyl monocations in order to infer the effects of compound structure on both DNA binding and the accompanying change in hydration. From this second study, we concluded that substitution at the monocationic terminus of such compounds does not correlate with any major hydration pattern. However, we identified a potential major role for structure effects – namely, hydrophobicity and bulkiness – in directing DNA binding affinities by these non-isohelical compounds. Finally, in a peripheral project, we probed the role of hydration in DNA sequence recognition by an isohelical diamidine, and found that its preference for longer, uninterrupted A/T-tracts was largely independent of changes in preferential hydration on binding.

5.1 Future Directions

The studies within this thesis motivate a number of new research questions while leaving others still unanswered. In light of the tendency discovered for linear dications, but not linear monocations, to dimerize in the DNA minor groove and self-associate in the unbound state, one might wonder whether the trend between charge number and dimerization would be maintained if the compounds were isohelical with the minor groove. An investigation akin to that in Chapter 2 but using isohelical analogues of those compounds would be useful in answering such a question.
In Chapter 3, we demonstrated that DNA binding by five different indole-biphenyl monocations is accompanied by a release of preferential hydration corresponding to a very large, favorable entropic contribution to the binding free energy. When considered in the context of the overall free energy of binding, which is much smaller in magnitude than the entropic contribution from water and counterion release, it became clear that a significant unfavorable contribution to the binding free energy is present. What accounts for this energetic penalty? Future studies capable of teasing apart the thermodynamics of binding in detail (such as isothermal titration calorimetry experiments) seem a crucial next step to understanding structure-binding relationships by non-isohelical monocations. When coupled with surface plasmon resonance studies or other experimental approaches to probe the kinetics of binding, detailed DNA binding mechanisms and their structure dependencies could be defined to yield a complete picture of DNA recognition by these compounds.

Finally, we show in Chapter 4 that, while the isohelical diamidine DB1976 shows a clear DNA binding preference for longer A/T tracts over short A/T tracts or those interrupted by G/C base-pairs, this DNA sequence specificity is not explained by the rather similar hydration properties of the various DNA/compound complexes. What, then, are the other (non-preferential) hydration events that accompany DNA binding by DB1976, and do these hydration events correlate with DNA sequence specificity? Current efforts to answer these questions are being pursued via direct volumetric studies and molecular dynamics simulations. Additionally, as no crystal nor NMR structures exist for DB1976 in complex with DNA, attempts to obtain such a structure may prove valuable by offering yet another perspective for its hydration properties.

In the process of completing this thesis, it came to the attention of this author that there is a severe lack of experimental evidence for the stability (or lack thereof) of the polyelectrolyte
charge fraction (see Chapter 1.2.1) at salt concentrations higher than ~1.5 M. Considering the rarity with which experimentalists require such high (and, admittedly, non-physiological) salt concentrations, this absence of data is not altogether surprising. However, in light of our present use of NaCl concentrations at ionic strengths up to ~3.5 M to make inferences about properties that may have biological relevance, answering this question appears to have become exceedingly important. Thus, I propose an effort to revisit a decades-old question and extend the experimental support for polyelectrolyte theory to the high-salt limit via 23-sodium NMR studies and the Poisson-Boltzmann approach for analysis of the charge fraction. This project would shed light on an interesting biophysics question and may be of relevance to not only the nucleic acids and physical chemistry communities but also polymer chemists and material scientists working with polyelectrolyte compounds.
REFERENCES


[78] N. Erlitzki, THE ROLE OF WATER IN DNA MINOR GROOVE BINDING BY HETEROCYCLIC CATIONS, Chemistry, Georgia State University, ScholarWorks @ Georgia State University, 2017, pp. 96
[85] M.T. Record, Jr., C.F. Anderson, T.M. Lohman, 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, Quarterly reviews of biophysics, 11 (1978) 103-178.
[86] Permissions, Elsevier.


