UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Targeting unique nucleic acid structures with small molecules

Permalink

https://escholarship.org/uc/item/7mj8z8xc

Author

Tam, Victor Kin-man

Publication Date

2007

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Targeting Unique Nucleic Acid Structures with Small Molecules

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Chemistry

by

Victor Kin-man Tam

Committee in Charge:

Professor Yitzhak Tor, Chair
Professor Michael Burkart
Professor Jeffery Esko
Professor Thomas Hermann
Professor Peter Wolynes

Copyright

Victor Kin-man Tam, 2007.

All rights reserved.

The Dissertation of Victor Kin-man Tam is approved, and it is acceptable in quality and form for publication on microfilm.

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2007

DEDICATION

to my parents Tom and Patricia
&
my sisters Yvonne and Elaine

EPIGRAPH

"All that is necessary for the triumph of evil
is for good men to do nothing."

Edmund Burke, philosopher

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Figures x

List of Schemes xv

List of Tables xvi

Acknowledgements xvii

Vita… xx

Abstract xxi

Chapter 1: The Structural and Thermodynamic Considerations of Targeting Nucleic Acids 1

1.1 Targeting the Unique Structure of RNA 1

1.2 The Chemical Structure of Nucleic Acids 5

1.3 Modes of Binding 10

1.4 Nucleic Acid Polyelectrolyte Theory and Electrostatics 12

1.5 Hydration of Nucleic Acids 17

1.6 Thermodynamics of Binding 20

1.7 Small Molecule Strategies to Target Nucleic Acids 23

1.8 Targeting Triple Helices and Quadruplexes 35

1.9 The Need for RNA General Recognition Principles 38

1.10 RNA Structure and Groove Accessibility 40

1.11 RNA–Small Molecule Interactions: Aminoglycosides and the Ribosome 43

1.12 RNA Recognition Models 46

1.13 Concluding Remarks 54

1.14 References 57

Chapter 2: Extended Ethidium Bromide Analog as a Triple Helix Intercalator 68

2.1 The Importance of Triple Helices 68

2.2 The Structure of Triple Helical Nucleic Acids 70

2.3 Triple Helix Selective Binders 74

2.4 Design of an Extended Ethidium Bromide Scaffold 80

2.5 Synthesis of an Extended Ethidium Bromide Analog 82

2.6 Photophysical Properties of an Extended Ethidium Analog 83

2.7 Binding to Duplex and Triplex Nucleic Acids 85

2.8 Thermal Melting Experiments 88

2.9 Conclusions 91

2.10 Acknowledgement 91

2.11 Experimental 92

2.12 References 105

Chapter 3: Design and Synthesis of Neamine and 2-Deoxystreptamine Mimics 110

3.1 Prokaryotic Translation, A-site, and Aminoglycosides 110

3.2 Structure of Aminoglycosides 119
3.3 The Mechanism of Aminoglycoside Resistance 120
3.4 Aminoglycoside Mimetics 123
3.5 Targeting Unpaired A1408 with Neamine Mimics 133
3.6 Design and Synthesis of 2-Deoxystreptamine Homologues 138
3.7 Conclusions 141
3.8 Experimental 143
3.9 References 163
Chapter 4: Development of a Fluorescent HIV-1 DIS Binding Assay 169
4.1 The Impact of HIV on Humans 169
4.2 The HIV Life Cycle 171
4.3 Therapeutics Targeting the HIV Life Cycle 174
4.4 HIV-1 Dimerization: Structure and Importance 177
4.5 Design and Development of a HIV-1 DIS Fluorescence Assay 180
4.6 Controlling Kissing Loop and Extended Duplex Formation 183
4.7 Thermal Denaturation Studies 185
4.8 Monitoring Small Molecule Binding with DIS272(2AP) 186
4.9 Monitroing Small Molecule Binding with DIS273(2AP) 196
4.10 Fluorescent DIS Constructs Detect and Quantify Ligand Binding 198
4.11 Ligand Screening, Quantification and Discovery 201
4.12 Conclusions 202
4.13 Acknowledgement 203

4.14 Experimental.....205

4.15 References 209

LIST OF FIGURES

Figure 1.1: Structure of tRNA 3

Figure 1.2: Structures of common nucleosides found in nucleic acids 5

Figure 1.3: Canonical Watson-Crick base pairs 6

Figure 1.4: Examples of non Watson-Crick base pairing 7

Figure 1.5: Schematic of major sugar puckering modes found in DNA and RNA 8

Figure 1.6: Views of the major and minor grooves of a representative B-form DNA duplex 10

Figure 1.7: Views of the major and minor grooves of a representative A-form RNA duplex 11

Figure 1.8: Partial electrostatic charges of heteroatoms in Watson-Crick base pairs 13

Figure 1.9: Spine of hydration in the minor groove of the Dickerson-Drew dodecamer 19

Figure 1.10: Intercalation of ethidium bromide between dimeric C•G dinucleotides 23

Figure 1.11: Examples of simple intercalators 25

Figure 1.12: Examples of complex intercalators 26

Figure 1.13: X-ray crystal structure of a 2:1 nogalamycin:DNA complex 27

Figure 1.14: Examples of bisintercalators 28

Figure 1.15: NMR structures of bisintercalator bis-dauonrubicin binding into a DNA duplex 29

Figure 1.16: Examples of groove binders 30

Figure 1.17: X-ray crystal structure of a 1:1 netropsin:DNA complex 31

Figure 1.18: Schematic representation of hydrogen bonding contacts between distamycin and AATT bases in the minor groove of DNA..... 32

Figure 1.19: Schematic representation of the hydrogen bonding contacts between at PyPyIm “lexitropsin” and ATTG bases in the minor groove of DNA ..... .32

Figure 1.20: A 2:1 complex of ImPyPy with a DNA duplex in the minor groove..... 33

Figure 1.21: Schematic representation of the hydrogen bonding contacts between a sequence-selective hairpin polyamide in the minor groove of DNA 34

Figure 1.22: Hydrogen bonding pattern of PyPu×Py triplet bases and G-quadruplex base pairs. 36

Figure 1.23: Schematic structure of a CG×G antiparallel triple helix..... 36

Figure 1.24: Representative structures of triple helix binders 37

Figure 1.25: Schematic representation of different RNA secondary structures..... 42

Figure 1.26: RNA secondary structure of therapeutically targeted RNA constructs..... 42

Figure 1.27: Structure of different classes of aminoglycosides 44

Figure 1.28: Energy landscape of the “lock and key” and “conformational capture” model hypothesized to be associated with RNA–ligand recognition..... 47

Figure 1.29: Structures of different deoxytobramycin analogs..... 49

Figure 1.30: Structures of “amino-aminoglycosides”..... 51

Figure 1.31: Comparison of distanced between magnesium cations in the hammerhead ribozyme ammonium functional groups in neamine..... 53

Figure 2.1: Examples of base triplet orientations observed in triple helices..... 71

Figure 2.2: Nomenclature of grooves in a triple helix..... 72

Figure 2.3: Polyaromatic benzopyridoindole derivatives 75

Figure 2.4: Examples of naphthylquinoline-based derivatives 77

Figure 2.5: Examples of substituted amidoanthroquinones 78

Figure 2.6: Triplex selective dyes and antibiotics 79

Figure 2.7: Ethidium bromide **1** and extended ethidium bromide **2** 81

Figure 2.8: Photophysical properties of **1** and **2**. 84

Figure 2.9: Nucleic acid binding isotherms of **1** and **2** with ctDNA and poly r(A)•r(U) 86

Figure 2.10: Nucleic acid binding isotherms of **1** and **2** with duplex poly d(A)•d(T) and triple helix poly d(T)•d(A)•d(T) 87

Figure 2.11: Binding isotherm of **1** and **2** with triple helix poly r(U)•r(A)•r(U) 88

Figure 2.12: Thermal melting profiles of **1** and **2** with dT_{19}•dA_{19}•dT_{19} 89

Figure 2.13: Thermal melting profiles of **1** and **2** with rU_{19}•rA_{19}•rU_{19} 90

Figure 3.1: The translation initiation complex 112

Figure 3.2: First stage of the elongation phase of translation 113

Figure 3.3: Second stage of the elongation phase of translation 114

Figure 3.4: Third stage of the elongation phase of translation 114

Figure 3.5: Termination of translation 115

Figure 3.6: Contacts made between paromomycin and the prokaryotic A-site ...116

Figure 3.7: Structure of the prokaryotic A-site docked with a tRNA, paromomycin and mRNA 118

Figure 3.8: Structures of different aminoglycoside families.....120

Figure 3.9: Modification sites of various resistance enzymes on kanamycin B ..122

Figure 3.10: Examples of aminoglycoside and 2-DOS mimics synthesized by Thomas Hermann and colleagues 124

Figure 3.11: Examples of aminoglycoside mimics synthesized by Chi-Huey Wong and colleagues 126

Figure 3.12: Examples of additional aminoglycoside mimics synthesized by Chi-Huey Wong and colleagues 128

Figure 3.13: Examples of aminoglycoside mimics synthesized by Isis 129

Figure 3.14: Representative pyranmycin and pyrankacin structures....130

Figure 3.15: Modified aminoglycosides synthesized by Tor and colleagues 132

Figure 3.16: Proposed 2-DOS-pyrimidine conjugates....134

Figure 3.17: Proposed triazine-2-DOS homologue conjugates....139

Figure 4.1: HIV life cycle 172

Figure 4.2: Structures of nucleoside reverse transcriptase inhibitors 175

Figure 4.3: Structures of non-nucleoside reverse transcriptase inhibitors 176

Figure 4.4: Structures of protease inhibitors 176

Figure 4.5: Schematic representation of the 5′-end of the HIV RNA genome and the DIS 179

Figure 4.6: Structural comparison of the prokaryotic A-site, DIS kissing loop, and DIS extended duplex 181

Figure 4.7: Putative base-flipping dynamics of purines 272 and 273 182

Figure 4.8: Sequence and secondary structure of 23mer RNA DIS constructs ...182

Figure 4.9: Native PAGE used to monitor the association of the RNA hairpins .184

Figure 4.10: Normalized thermal melting profile of RNA construcuts....185

Figure 4.11: Structures of the different classes of ligands tested....186

Figure 4.12: Fluorescence titrations with paromomycin 187

Figure 4.13: Fluorescence titrations with neomycin B 189

Figure 4.14: Fluorescence titrations with ribostamycin.....190

Figure 4.15: Fluorescence titrations with butirosin A 191

Figure 4.16: Fluorescence titrations with neamine.....192

Figure 4.17: Fluorescence titrations with **1a**.....193

Figure 4.18: Fluorescence titrations with **1d**.....193

Figure 4.19: Fluorescence titrations with **2**.....194

Figure 4.20: Fluorescence titrations with apramycin 195

LIST OF SCHEMES

Scheme 2.1: Synthesis of extended analog **2**..... 82

Scheme 3.1: Synthesis towards **1**.....135

Scheme 3.2: Synthesis towards **2**.....136

Scheme 3.3: Synthesis of **3** and **4**.....137

Scheme 3.4: Synthesis of **5**.....138

Scheme 3.5: Synthesis towards **18**.....140

Scheme 3.6: Synthesis towards **28**.....141

LIST OF TABLES

Table 1.1: Table of physical characteristics between B-form DNA and A-form RNA. 9

Table 1.2: Table of Dervan's pairing rules for sequence selective targeting with polyamides 35

Table 1.3: Table of rate constants for hammerhead ribozyme cleavage 52

Table 2.1: Nucleic acid affinities of **1** and **2** 88

Table 2.2: Thermal melting points of a DNA 19mer duplex and triplex 90

Table 2.3: Thermal melting points of a RNA 19mer triple helix 91

Table 4.1: Table of EC_{50} values (μM) for DIS ligand binding 197

ACKNOWLEDGEMENTS

The completion of this thesis and my doctorate degree is the product of many years of support from friends and family. First of all, I would like to thank my parents Tom and Patricia, and sisters Yvonne and Elaine for always being supportive, loving and instilling within me the value of an education. I am forever grateful for all that my family has given me.

I have been blessed to know many friends throughout my lifetime, and I am a reflection of who they are. I am thankful to have crossed paths with the always lovely and supportive Tina Ng, as well as Stephanie Cordova, Michelle Kim, Thuy Trang, Kimberly Dauz, Tammy Sirich, Almita Heramia, Chris Noble and John Dise. There are so many others I would like to thank, but it would be impossible to list everyone that has affected and changed my life. To them all, please know that I am forever in your debt.

Next, I must thank my collaborators, of which, none of this work would be possible. Charles (Qi) Liu was an excellent teacher during my first years in graduate school and always encouraging whenever I stumbled. Denise Kwong and Baia Lasky were fantastic chemistry partners that only enriched my UCSD experience, and I thank them for their warmth and friendship. I would also like to acknowledge Paola Castaldi for always being helpful during the latter years of graduate school and sharing her great sense of humor with me.

Additionally, I have been able to work with and meet many brilliant people whom are not only my co-workers but lifelong friends. I would like to thank my partner-in-crime in the chemistry lab Nicholas Greco, who has always been a great sounding board and source of advice. I also must acknowledge other graduate students and postdoctoral scholars, for being my scientific older brothers and sisters and teaching me with patience and enthusiasm. They include Derek Fisher, Dave Jaramillo, Kenneth Blount, S. G. Srivatsan, Renatus Sinkeldam, Beth Wilson, Phoebe Glazer, Dave Heidary, Michelle Hysell, Grace Yang, Damien Jouvenot and Robyn Drake Swanland.

Next, in going full circle to the beginning, I would like to thank my high school chemistry teacher Charles Miltenberger who started me in science and my undergraduate advisor Matthew B. Francis. I thank them for the scientific foundation they provided and the passion they instilled in me to continue my scientific career.

Finally, I would like to thank Thomas Hermann for always being an insightful and supportive collaborator, and Yitzhak Tor for always ensuring I was a well-balanced scientist. Initially, I was narrowly focused on the research fields I wanted to pursue, but Yitzhak insisted that I become more interdisciplinary in order to fully appreciate and be more proficient in the scientific world. Even with some bumps in the road, Yitzhak has always been encouraging to tackle new endeavors, and I am a better person for it.

In regards to the material in this dissertation, please note the following:

Portions of Chapter 2 have been previously published in the article "Extended Ethidium Bromide Analogue as a Triple Helix Intercalator: Synthesis, Photophysical Properties and Nucleic Acids Binding," *Chem. Commun.* **2006**, 2684–2686 by V. Tam, Q. Liu and Y. Tor. The dissertation author was the primary investigator and author of this paper.

Portions of Chapter 4 have been previously published in the article "Fluorescent HIV-1 Dimerization Initiation Site: Design, Properties, and Use for Ligand Discovery," *J. Am. Chem. Soc.* **2007**, *129*, 3257–3266 by V. Tam, D. Kwong and Y. Tor. The dissertation author was the primary investigator and author of this paper.

VITA

EDUCATION

University of California, San Diego
Ph.D., Chemistry; September 2007

University of California, San Diego
M.S., Chemistry; March 2004

University of California, Berkeley
B.S., Chemistry; May 2002

AWARDS, HONORS AND FELLOWSHIPS

Teddy Traylor Graduate Fellowship, UC San Diego (2006)
University-wide AIDS Research Program Pre-Doctoral Scholar, UC San Diego (2004)
Teaching Assistant Excellence Award, UC San Diego (2004)
George C. Pimentel Award for Undergraduate Research, UC Berkeley (2002)
UC Leadership Excellence Through Advance Degrees (LEADS) Scholar, UC Berkeley (2000)

PUBLICATIONS

1. Tam, V. K.; Kwong, D.; Tor, Y. "Fluorescent HIV-1 Dimerization Initiation Site (DIS): Design, Properties and Use for Ligand Discovery." *J. Am. Chem. Soc.* **2007**, 3257–3266.

2. Tam, V. K.; Liu, Q.; Tor, Y. "Extended Ethidium Bromide Analogue as a Triple Helix Intercalator: Synthesis, Photophysical Properties and Nucleic Acids Binding." *Chem. Commun.* **2006**, 2684–2686.

ABSTRACT OF THE DISSERTATION

Targeting Unique Nucleic Acid Structures
with Small Molecules

by

Victor Kin-man Tam

Doctor of Philosophy in Chemistry

University of California, San Diego, 2007

Professor Yitzhak Tor, Chair

Nucleic acids are the central biomolecules of life, encoding all the genetic information necessary for cellular metabolism and replication. Binding of small molecules to DNA or RNA can perturb their function as well as interactions with proteins or other nucleic acids. Disrupting these key intermolecular interactions with low molecular weight ligands is a promising strategy for the development of new therapeutics. Our efforts have been focused on three different areas. First, the formation of triple helices has been shown to affect the normal binding of duplex nucleic acids with enzymes such as DNA polymerase. Stabilizing these triple helices with ligands that exploit π–π stacking interactions can potentially assist in selective modulation of gene expression. We have demonstrated that an extended

phenanthridinium core selectively binds to DNA triple helices. Second, a diminishing number of last resort antibiotics are effective against the continuing emergence of drug resistant bacteria. Mimicking the characteristics of current antibiotics with new molecular scaffolds is one strategy to overcome antimicrobial resistance mechanisms. With this in mind, we have initiated the synthesis of a small library of heterocycle–deoxystreptamine conjugates to potentially mimic the efficacy of aminoglycosides, a class of RNA-binding antibiotics. Finally, the rapid screening of low molecular weight ligands is critical to identifying new, potential nucleic acid binders. We have developed a novel, real-time fluorescence-based assay for monitoring ligand binding to the HIV-1 DIS, which has led to the discovery of new natural and semi-synthetic binders to this interesting viral RNA target. Taken together, our research highlights the potential benefits of targeting nucleic acids: the selective modulation of gene expression, overcoming current antibiotic resistance mechanisms, and inhibiting replication of retroviruses.

Chapter 1

The Structural and Thermodynamic Considerations of Targeting Nucleic Acids

---

1.1 TARGETING THE UNIQUE STRUCTURE OF RNA

The transfer of hereditary information is fundamental to the continuation of life. In most organisms, DNA (deoxyribose nucleic acid) stores all the genetic information necessary for cell replication and function. The central dogma of biology states that DNA transcribes its genetic code into RNA (ribose nucleic acid), which in turn is processed at the ribosome for protein synthesis. Since the discovery of the double helix, molecular biologists has been focused on understanding the role nucleic acids played in cellular function, with the focus mostly on DNA, while RNA was relegated to the background as a passive messenger molecule. Two events, in particular, challenged this view. First, the discovery of retroviruses, where RNA acts as the main genetic storage biomolecule, expanded the traditional view of information flow in the cell and expanded RNA's role in molecular biology.[1] Second, Cech and Altman further enhanced the understanding of RNA's capabilities with the discovery of catalytic RNA (ribozymes) in the 1980's, suggesting that RNA possesses functional capabilities on the order of proteins.[2–4] These two paradigm shifts have altered the understanding of importance in RNA's cellular function. Due to the significance of

these discoveries, it is clear that to comprehend fully the pivotal roles of RNA, particular attention must be paid to its structure and interactions with effector ligands.

Unlike DNA, RNA is found in a multitude of secondary structures beyond the ordinary double helix described by Watson and Crick in 1953.[5,6,7] In the cell, RNA is predominantly in a globular form connected by double-stranded regions.[8] Within the globular region, the single stranded RNA folds upon itself to form a complex architecture of hairpin loops, symmetric and asymmetric bulges, and junctions that bend to create an infinite array of secondary structures.[9] The first X-ray crystal structure of tRNA, yeast alanine transfer RNA ($tRNA^{Ala}$), provided a prime example of the diverse motifs RNA can adopt (Figure 1.1).[10] Hairpin loops compose the D, TΨC, and anticodon arms, while tRNA's characteristic cloverleaf shape is formed by a central stem junction. Purine-purine and pyrimidine-pyrimidine non-canonical base pairs form internal loops and bulges that deviate from a canonical double helix. The structure of tRNA mediates interactions with proteins and other nucleic acids, suggesting that folds, clefts, grooves and pockets can act as RNA structural recognition motifs.[11]

The interesting structural topology of RNA functions as binding sites for proteins and effector molecules. Binding can act as a signal for the initiation or control of specific cellular processes; therefore, targeting these secondary and tertiary nucleic acid structures with small molecules to disrupt normal biological function is a logical avenue for therapeutics development.[11,12–21] Examples of small molecules inhibiting protein–RNA binding or RNA function have been highly cited in

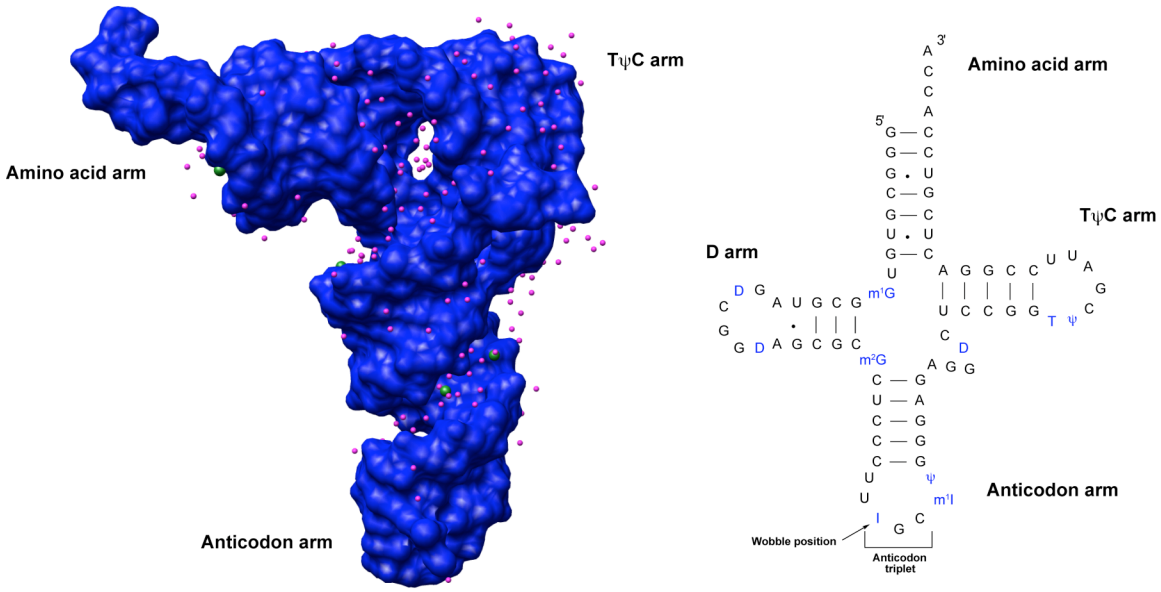


Figure 1.1. Structure of tRNA. *Left*. The three dimensional x-ray crystal structure of $tRNA^{Phe}$ at 2.0 Å resolution stabilized by magnesium ions (green) and water molecules (magenta). PDB ID: 1EVV. (Jovine, L., Djordjevic, S., Rhodes, D. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg^{+2} in 15-year old crystals. *J. Mol. Biol.* **2000**, *301,* 401–414). *Right*. Base sequence of $tRNA^{Ala}$ with modified bases indicated (blue). Modified bases: D, 5,6-dihydrouridine; ψ, psuedouridine; I, inosine; T, ribothymidine; m^1I, 1-methylinosine; m^1G, 1-methylguanosine; m^2G, N^2-dimethylguanosine. (Figure adapted from Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Lehninger principles of biochemistry*. 3rd ed.; Worth Publishers: New York, 2000; p. 1038).

literature.[22–24] In the human immunodeficiency virus (HIV), the retrovirus responsible for AIDS, particular attention has been paid to important lifecycle interactions between proteins and RNA. Examples include the Rev protein and Rev Response Element (RRE) RNA binding site,[25,26] the Tat protein and segment of RNA termed the Tat transactivation region (TAR),[27,28] and, more recently, the NCp7 encapsidation protein and the Dimerization Initiation Site (DIS).[29] Disruption of any of these interactions results in the reduction of viral replication.[23] In bacteria, a majority of current antibiotics bind the prokaryotic ribosome, an RNA-based cellular machine, and interfere with the fidelity of protein synthesis.[30–32]

Numerous therapeutic benefits can be derived from targeting RNA, and while the field is young, the past ten years has provided greater structural understanding of this nucleic acid and its role in biological processes; however, the seemingly limitless number of RNA secondary structures increases the difficulty of ligand design. Recognition principles derived from one RNA–ligand complex are difficult to apply to another RNA structure. DNA, in comparison, has a canonical double helical structure, and has proven to be sequence selectively targeted by synthetic ligands and oligonucleotides.[33–35] RNA lacks general principles for targeting, and few synthetic ligands, for example, have been able to surpass their natural antibiotic counterparts in selectivity and affinity.

To understand the fundamental principles of RNA recognition, the structure of DNA and RNA are first presented. The structures of these two nucleic acids will be discussed in depth to gain insight into their ligand recognition characteristics. Second, the effects of these negatively charged polymers on the surrounding aqueous and ionic environment will be presented. Next, selected examples in the development of targeting DNA and RNA with model compounds will be discussed thoroughly. From these compounds, the basis on which future ligand design can be directed will be discussed. Finally, two RNA–ligand recognition models will be presented in the context of select SAR studies.

1.2 THE CHEMICAL STRUCTURE OF NUCLEIC ACIDS

DNA is a polymer composed of four heterocyclic bases built upon a pentose sugar-phosphate backbone. The pyrimidine (cytosine and thymine) and purine (adenine and guanine) nitrogenous bases are each connected to a 2-deoxyribose sugar through a β-*N*-glycosidic bond forming a nucleoside (Figure 1.2). Nucleosides are connected through phosphodiester bonds at the 5′-OH and 3′-OH position of the sugar to phosphate units which alternate per nucleoside. A polymer of nucleoside-phosphate units form a DNA strand; however, DNA is normally found as a right-handed spiraling duplex of anti-parallel strands whose double helical structure was elucidated by Watson and Crick in 1953.[6] Watson and Crick determined that the bases in DNA interact in a specific hydrogen bonding pattern; normally, guanosine pairs with cytosine and adenosine pairs with thymidine (Figure 1.3). A guanine-cytosine base pair (G•C) is mediated by three hydrogen bonds, where guanine nitrogens N1 and N2 donate hydrogens to electronegative hydrogen bond acceptors O2 and N3 of cytosine.

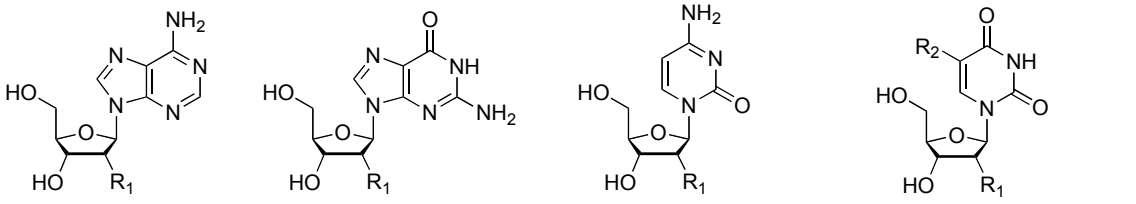


Figure 1.2. Structures of the common nucleosides found in nucleic acids. *Left to right.* Purines, adenosine (A) and guanosine (G), and pyrimidines cytidine (C), thymidine (T)/uridine (U). In DNA, nucleosides are composed of deoxyribose sugars (R_1 = H), and in RNA, sugars are ribose (R_1 = OH). In RNA, a uracil nucleoside is commonly found (R_2 = H) instead of thymine (R_1 = CH_3),

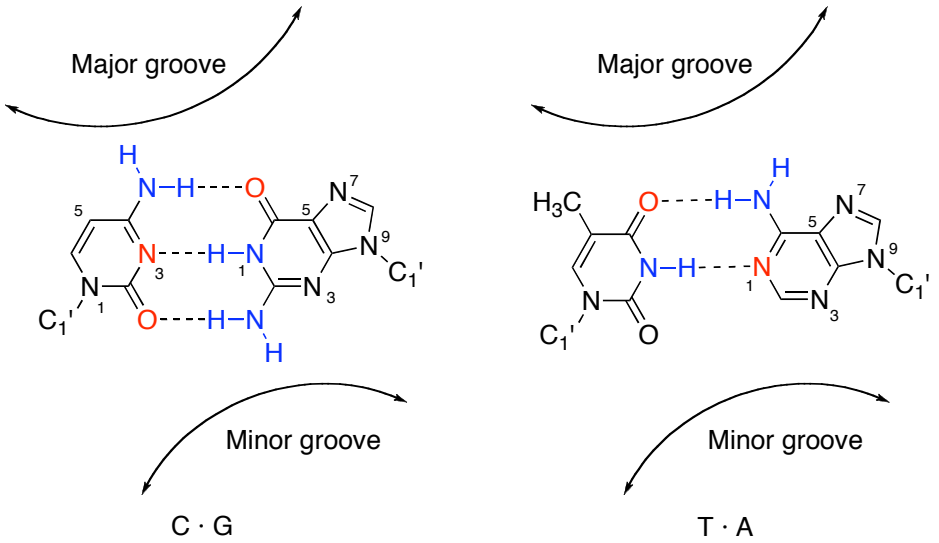


Figure 1.3. Canonical Watson-Crick base pairs. Purine•pyrimidine hydrogen-bonding partners with the major and minor grooves are indicated. Hydrogen bond donating (blue) and accepting (red) functional groups are labeled.

The exocyclic N4 amine of cytosine donates a hydrogen to the O6 carbonyl of guanine to complete the hydrogen-bonding pair. An adenine-thymidine base-pair (A•T) is characterized by hydrogen-bonding pairs between N3 of thymine and the N1 of adenine, while the N6 exocyclic amine of adenine donates a hydrogen to the thymine C4 carbonyl. In RNA, the structure differs slightly. Instead of 2′-deoxyribose, the sugar moiety is a ribose and thymine is replaced with uracil. The A•U base pair forms two hydrogen bonds in the same manner as an A•T base pair. The geometry of Watson-Crick base pairing is conserved in DNA and RNA helices.[36,37] The hydrogen-bonding distance between nucleotides varies slightly, from 2.8 to 2.95 Å, and the average strength of base pairing is 6–10 kJ/mol.[37]

Although Watson-Crick base pairing is the dominant hydrogen-bonding pattern in nucleic acids, other base-pairing interactions have been identified (Figure 1.4) in higher ordered structures such as triple helical nucleic acids and quadruplexes

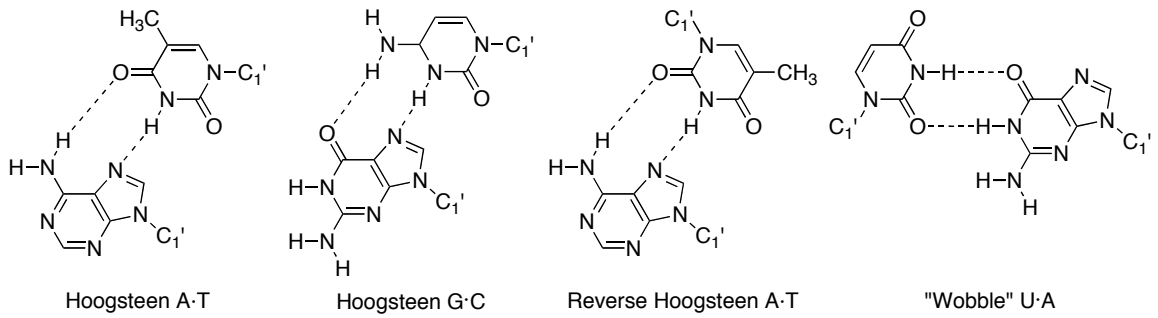


Figure 1.4. Examples of non Watson-Crick base pairing. Both Hoogsteen and reverse Hoogsteen base pairing interactions occur along the major groove face and involve interaction with N7 of purines. "Wobble" base pairing is a "shifting" of normal Watson-Crick interactions.

(see Chapter 1.8 and Chapter 3). Hoogsteen base-pairing occurs along the N7 face of purines. In an A•T Hoogsteen base pair, N6 and N7 of adenine hydrogen bonds with O4 and N3 of thymine, respectively. In a Hoogsteen G•C base pair, O6 and N7 of guanine hydrogen bonds with N4 and N3 of cytosine, respectively. A reverse Hoogsteen base pair can also occur where the orientation of the third base is reversed. Other non Watson-Crick interactions include wobble base pairs, where the normal hydrogen bonding interaction is shifted from pyrimidine C4 to C2 (Figure 1.4). Non-canonical purine-purine and pyrimidine-pyrimidine base mismatches can also occur and are partially responsible for the diverse secondary structure of RNA.

In DNA, the 2-deoxyribose sugar moiety adopts an out-of-plane conformation to limit steric clashes with neighboring substituents and maximize hydrogen-bonding interactions for structure stability.[8] The twisted conformation is known as a sugar pucker, where nomenclature identifies the atom that is above or below the plane of the ribose (Figure 1.5). Normally, C2′ or C3′ are observed out of the plane created by atoms C1′-O4′-C4′. A C2′-endo pucker is normally observed in DNA and purine

ribonucleosides while a C3′-endo puckering is seen with pyrimidine ribonucleosides.[8,37] An intrastrand hydrogen bond between O2′ of ribose to O4′ in an adjacent nucleoside assists in RNA stability and favors C3′-endo puckering. The barrier of interconversion between these two conformations is approximately 2–5 kJ/mol.[36]

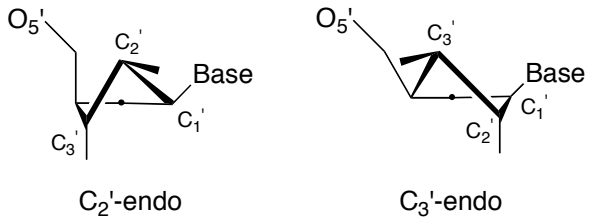


Figure 1.5. Schematic representation of the major sugar puckering modes found in DNA and RNA.

A majority of nucleic acids are found in two different conformations, or polymorphs, as determined by fibre diffraction studies.[36] Each is formed depending on relative humidity and salt concentrations. The right-handed B-form DNA (the predominant conformation found in vivo) is favored at high humidity and low salt conditions while A-form DNA is favored at low humidity and high salt concentrations. Although a number of polymorphs exist, including A′, C, C′, D, E, T, and Z, the B- and A- forms are the predominant secondary structure conformations of DNA and RNA, respectively.

For B-form DNA, the helical twist per base pair is 36°, with 10 residues required to make a full helical turn. The distance between base pairs, or helical rise per base pair, is 3.3–3.4 Å, and the pitch height, or distance for a full helical turn is 34 Å (data summarized in Table 1.1). Consequently, the distance between C1′-C1′ of

nucleosides in a base pair is 10.60 ± 0.15Å with an angle of 68 ± 2° between the two glycosidic bonds. The angles between glycosidic bonds form the major and minor grooves of DNA and RNA, as successive base pairs are stacked along the phosphate backbone (Figure 1.3). The width of the major groove, i.e., the distance between phosphates of opposite strands, is 11.7 Å, and the depth of the groove is 8.8 Å (Figure 1.6). For the minor groove, the width is nearly two-fold smaller at 5.7 Å, with a similar groove depth of 7.5 Å.

Table 1.1. Table of Physical Characteristics between B-form DNA and A-form RNA.

	DNA (B-form)	**RNA (A-form)**
Base pairs per helical turn	10	11
Helical Twist per base pair	36°	32.7°
Distance between base pair (helical rise) (Å)	3.4	2.8
Pitch height (full helical turn) (Å)	34	32.7
Helix Diameter (Å)	20	26
Major Groove Width (Å)	11.7	4.7
Major Groove Depth (Å)	8.8	12.9
Minor Groove Width (Å)	5.7	10.8
Minor Groove Depth (Å)	7.5	3.3

RNA is predominantly in the A-form polymorph which is characterized by a larger helical diameter and deeper, smaller major groove (Figure 1.7). The helical twist per base pair is slightly smaller at 32.7°, requiring 11 base pairs instead of 10 to complete one helical turn. The rise per base pair is also smaller at 2.8 Å with a helical pitch of 30 Å. The width of the major groove is nearly 2.5-fold smaller at 4.7 Å while extremely deep at 12.9 Å. The RNA minor groove width is nearly twice the size of DNA at 10.8 Å, while the depth is extremely shallow at 3.3 Å. As will be discussed later, the dimensions and physical characteristics of the nucleic acid grooves play a large part in directing ligand binding.

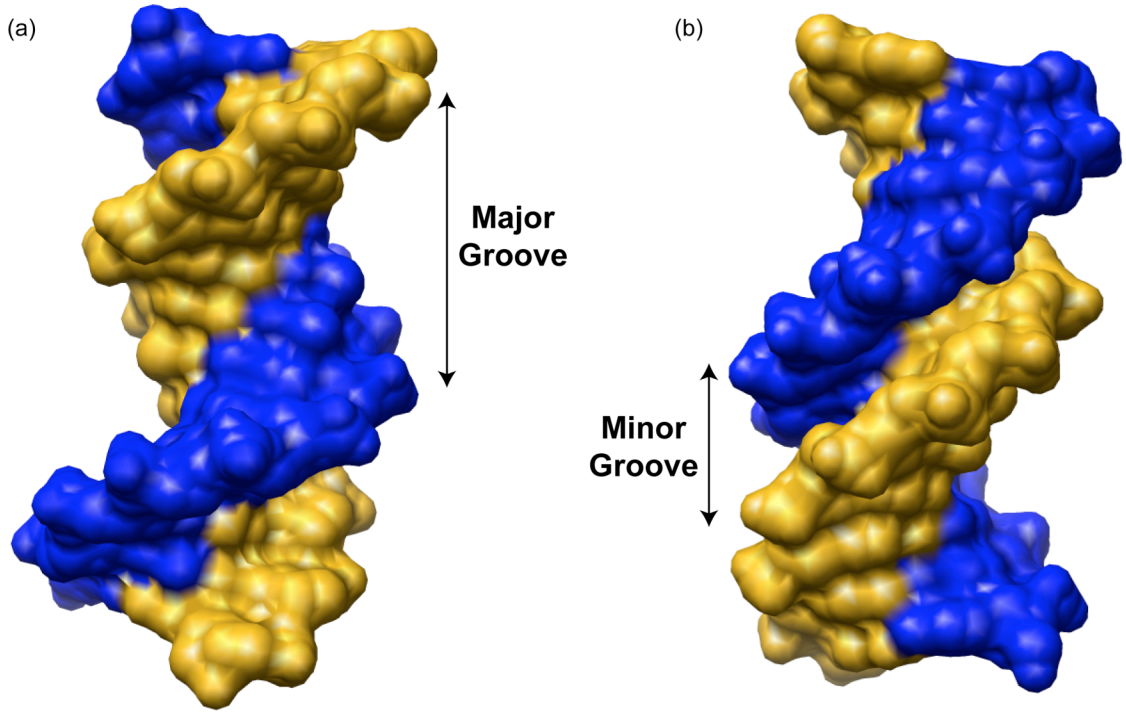


Figure 1.6. Views of the (a) major groove and (b) minor groove of a representative B-form DNA duplex (known as the Dickerson–Drew dodecamer). DNA duplex is an X-ray crystal structure of the self complementary sequence 5′–d(CGCGAATTCGCG)–3′. Each strand is colored separately (blue and goldenrod) while each view is approximately rotated 180° from one another. Major and minor groove widths are the distances between inter-strand phosphates. PDB ID: 2BNA. (Drew, H. R., Samson, S., Dickerson, R. E. Structure of a B-DNA dodecamer at 16K. *Proc. Natl. Acad. Sci. U. S. A.* **1982,** *79*, 4040–4044).

1.3 MODES OF BINDING

Ligand–nucleic acid interactions can be categorized into three groups: a) non-specific electrostatic binding, b) intercalation, and c) groove binding. Preference for one binding mode is largely influenced by ligand shape and nucleic acid conformation, electrostatics, and sequence composition. Depending on structure, ligands may bind to nucleic acids utilizing more than one mode.

Non-specific electrostatic binding. This binding mode governs the behavior of electrolytes and water within the nucleic acid's vicinity, and has little effect on organic

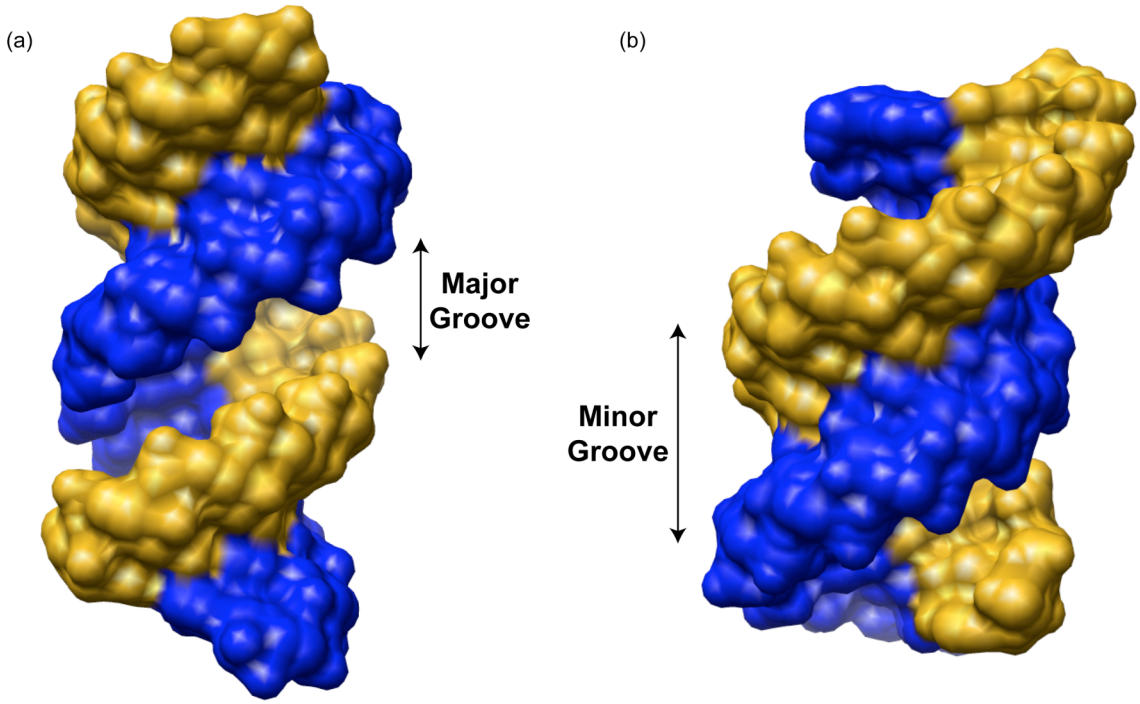


Figure 1.7. Views of the (a) major groove and (b) minor groove of a representative A-form RNA duplex The RNA duplex is an X-ray crystal structure of the self complementary 14mer sequence 5′–U(UA)$_6$A–3′. Each strand is colored separately (blue and goldenrod) while each view is approximately rotated 180° from one another. Major and minor groove widths are the distances between interstand phosphates. PDB ID: 1RNA. (Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J.S., Gough, G.R., Gilham, P.T., Moras, D. Crystallographic structure of an RNA helix: [U(UA)$_6$A]$_2$. *J. Mol. Biol.* **1989**, *209*, 459–474.)

ligands. Indiscriminate interactions between positively charged ions (Na^+, K^+ or Mg^{2+}) or polyamines (i.e., spermine or spermidine) that cluster around the negatively charged sugar-phosphate backbone are characteristic of non-specific electrostatic binding.

Intercalation. The term "intercalation" was first coined by Lerman in 1961 to describe the insertion of polyaromatic, planar molecules between consecutive base pairs of DNA perpendicular to the helical axis.[38] Stability of an intercalation complex results from overlap of the planar aromatic system with Watson-Crick base pairs to

maximize van der Waals contacts.[39] In addition, polarization of the stacked π–π systems by electron-rich or deficient substituents or heteroatoms minimizes repulsion within the complex.

Groove binding. Ligands that bind in the major or minor groove of nucleic acids make direct hydrogen-bonding contacts with functional groups along the edges of base pairs. Generally, small molecule ligands bind in the minor groove of DNA and major groove of RNA due to the high concentration of electrostatic potential (Chapter 1.4).

In contrast, proteins bind to the RNA minor groove and DNA major groove because of the complementary hydrophobic and neutral environment. In addition to electrostatics, binding is highly dependent on ligand shape complementarity to maximize van der Waals contacts.[40] Hydration and polyelectrolyte shells that stabilize nucleic acid structure also affect the mode of ligand binding.

1.4 NUCLEIC ACID POLYELECTROLYTE THEORY AND ELECTROSTATICS

Nucleic acids are negatively charged polymers that dramatically affect the behavior of ions, water and ligands in the surrounding bulk solvent. Studies in the 1970's on the electrostatics and interactions of ions with nucleic acids were mainly centered on DNA, due to its availability and stable double-helical structure. Within the major and minor grooves, the electrostatic potential is determined primarily by the orientation of individual base pairs (Figure 1.8).[8,41,42] Qualitatively, in an A•T base

pair, N3 of adenine and O2 of thymine both point towards the minor groove, contributing overall negative potential from each atoms' relatively high electronegativity. In comparison, the minor groove face of a G•C base pair is relatively neutral because of the functional groups that occupy this base pair edge. Although the analogous N3 in guanine adds negative potential as well as the O2 carbonyl of cytosine, the exocyclic N2 amino functionality adds positive potential to the minor groove neutralizing the negative potential contributed from the electronegative atoms.[8,36] Computationally, Pullman determined that the A•T base pairing edge (specifically near N3 of adenine) that composes the minor groove has more electronegative potential than any other location.[43] For G•C base pairs, the electropositive 2-amino functionality from guanine lowers the overall negative electrostatic potential at this location in the minor groove.[41] The partial electrostatic charges of bases were also confirmed using computational force field simulations (Figure 1.8).[44]

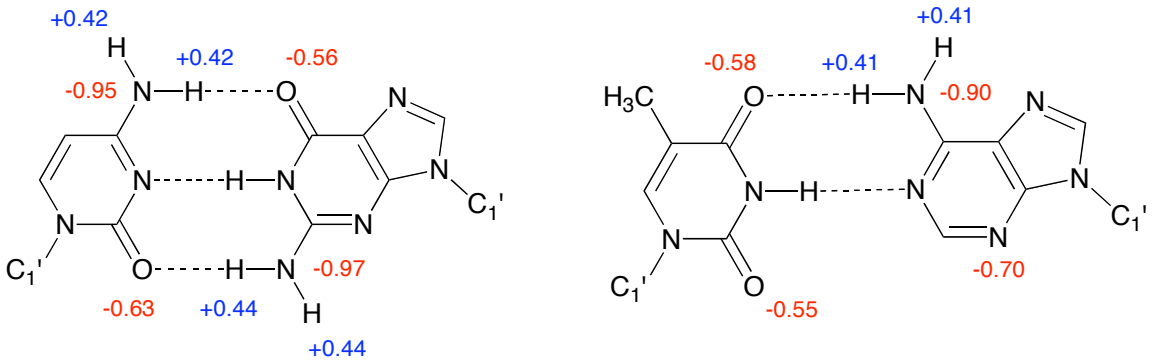


Figure 1.8. Partial electrostatic charges of heteroatoms in Watson-Crick base pairs. Electropositive (blue) and electronegative (red) atoms bordering the major and minor grooves as determined by computational simulations are labeled.[44]

For the major groove, similar qualitative and computational analysis can be performed. In the A•T base pair, a negative potential contribution is made from O4 of thymine, along with N7 of adenine; however, positive electrostatic potential is provided by the exocyclic 6-amino functionality of adenine. Similarly, in G•C base pairs, positive potential from the exocyclic amine at N4 of cytosine partially diminishes the negative potential contribution from N7 and C6 of guanine. Taken together, the major groove face of A•T and G•C bases pairs contain a total of two exocyclic amine functionalities, which lowers the overall negative potential, while in comparison the minor groove contains one electropositive amine.[8] These results were computationally confirmed by Pullman,[41,43] and more recent molecular dynamics simulations have also confirmed the parity in groove electrostatics for DNA.[40]

The location of electronegative atoms along base pair edges primarily determines the density of electrostatic potential in the major and minor groove. Although carbonyls contribute the most electronegative potential to a groove, a complementary electropositive amine neutralizes its contribution. Along the major and minor groove face of G•C base pairs, this interaction is observed. In contrast, the minor groove face of an A•T base pair is only occupied by the O2 carbonyl of thymidine creating localized electronegative potential. This directionality of groove electrostatics greatly impacts the interaction of nucleic acids and ligands.

The difference in electrostatics between A•T and G•C base pairs can also affect sequence selectivity for ligand–nucleic acid binding. For instance, G•C base pairs are more electron-rich, providing a preferable environment for intercalation of electron

deficient aromatic ligands. Additionally, the exocyclic 2-amino functionality of guanine in the minor groove can deter binding to G•C tracts due to sterics and instead favor A•T base pair-ligand interactions.[36] Similarly, along the major groove, the bulky methyl group of thymine can direct ligands to G•C base pairs instead. At the same time, the hydrophobic methyl group can also facilitate van der Waals interactions with other hydrophobic residues such as side chains leucine, isoleucine, and alanine to promote favorable energetic interactions for protein binding.[36]

Due to the wider helical diameter and slower rise of the A-polymorph form of RNA, the shallow and wide minor groove of RNA has a low electrostatic potential while the deep major groove concentrates a large amount of electronegative potential. This results in the major groove being the main binding site for small molecules due to Columbic attraction and hydrogen-bonding interactions, while proteins bind the shallow major groove due to van der Waals contacts between the groove floor and hydrophobic amino acid side chains.

The distribution of nucleic acid electrostatics has dramatic effects on molecules in the surrounding environment. This has a specific impact on metal ions as the negative phosphate backbone attracts cations to neutralize the negative potential. In the early 1970's, Manning proposed a quantitative model describing the interactions of cations with nucleic acid.[45] Specifically, to reduce the effective negative charge of DNA, cations associate within a few angstroms of the helix; however, cations are not locked within a certain orientation, but instead cluster along the negative backbone. Cation association is fully dependent on axial charge density. This has been called the

counter-ion condensation theory.[45] Later, Record incorporated Debye-Hückel activity coefficients into Manning's model suggesting the polarizable electron density of ions must be accounted for full quantification of cation association.[46] This association neutralizes the negatively charged phosphate backbone, thereby diminishing repulsive forces and stabilizes the structure of the nucleic acid.[47]

Since condensed ions occupy the same space as potential nucleic acid ligands, they energetically affect ligand binding.[48] Normally, release of condensed ions upon ligand binding provides a positive entropic contribution to the overall free energy; however, since counterions stabilize DNA structures, releasing condensed ions causes an energetic enthalpic penalty from destabilization.[45,46,49] Experimentally, a connection between salt concentration and ligand binding exists, and ligand affinity normally increases as bulk salt ion concentration decreases due to lowering the entropic penalty of ion displacement.

In summary, nucleic acids are negatively charged polymers that greatly affect their surrounding environment, and are an interesting target motif for positively charged or polarizable small molecules. Molecular simulations have confirmed the large negative electrostatic potential in the RNA major groove as well as DNA minor groove,[40] suggesting electronic as well as structure complementarity are necessary to exploit the charged character of these biomolecules.

1.5 HYDRATION OF NUCLEIC ACIDS

In addition to ions, water plays a pivotal role in structural stabilization and ligand binding.[48] Extensive studies by X-ray crystallography[50] and NMR[51] for the Dickerson-Drew dodecamer[52] reveal how integral water is for structural integrity. Experimentally, the water directly surrounding nucleic acids has been shown to have different properties than the bulk solvent,[53,54] and is necessary for stabilizing the secondary and tertiary structure of nucleic acids. Water, a highly polarizable molecule, clusters around the negatively charged phosphate backbone to neutralize charge and diminish electrostatic repulsion between strands. Additionally, water makes contacts with base pair edges by hydrogen bonding with heteroatoms and simulataneously acting as both a hydrogen bond donor and acceptor. The water involved in these interactions forms the first hydration shell around DNA double helices. Water–DNA interactions studied by infrared spectroscopy (IR) estimate that the first hydration shell surrounding each base pair consists of three water molecules per phosphate and approximately 19–20 water molecules throughout both grooves. A second, less structured hydration shell exists but is short-lived and dependent on the orientation of the first. Third and fourth layer hydration shells were once though to be stable but have subsequently been determined to be extremely short-lived, thus possessing characteristics similar to the bulk solvent.[55]

X-ray crystal structures of the Dickerson-Drew dodecamer provide insight into the hydration shell around one helical turn of B-form DNA.[56] As observed at 1.9 Å resolution, the first hydration shell around the phosphate backbone is not well-ordered.

Instead, an ordered water network is found in the major and minor grooves, with water molecules associating with the polar nitrogen and oxygen atoms along base-pairing edges. In the major groove, there is uniform solvation at both base pairs (A•T and G•C) with water contacts at N7 and N6 for adenine, O4 for thymine, N7 and O6 for guanine, and N4 for cytosine. No ordered secondary hydration shell is observed. In contrast, along the minor groove of the A•T base pairs of the dodecamer, a highly ordered water network is observed (Figure 1.9). The water network is arranged in a zig-zag pattern with water-bridging contacts between O2 of thymine and N3 of adenine. The bridge is constructed between a base on one strand and then the same base on the opposite strand in the next consecutive base pair. This ordered network provides structure for a second solvation shell.[57] Third and fourth layers were too disordered to be detected in the crystal structure. Guanine causes dehydration of the minor groove within the G•C base pairing tracks due to the 2-position exocyclic amine partially occupying the groove. The N2 of guanine disrupts the spine of hydration and expels water. Interestingly, the distance between the oligonucleotide and hydration spine increases from A•T to G•C base pairing tracts. The dehydrating effects of G•C base pairs also seem to affect transitions between polymorphs. Recall, low humidity conditions normally favor the A or Z polymorphs of DNA over the conventional B-form. Transition from B-DNA to A-DNA is easiest with oligonucleotides containing G•C rich sequences. A higher resolution structure of the Dickerson-Drew dodecamer was reported years later at 1.4 Å resolution, confirming the existence of a spine of hydration, but sodium was also observed to occupy the groove.[55]

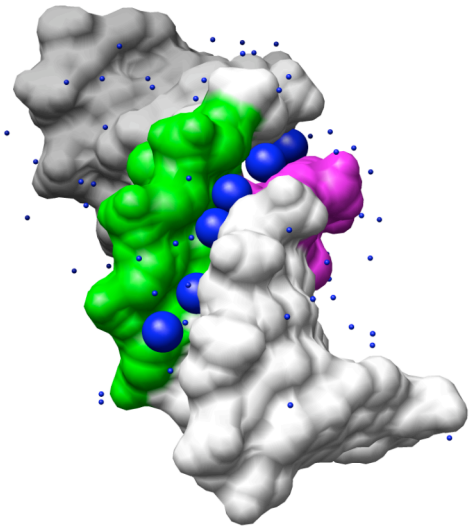


Figure 1.9. Spine of hydration in the minor groove of the Dickerson-Drew dodecamer. Water molecules (small blue balls) form the first and second hydration sphere. Water molecules (large blue balls) that interact with the 5′-AATT-3′ tract (green and magenta, each strand, respectively) form the spine of hydration in the DNA duplex minor groove.

For RNA, a high resolution duplex was visualized by X-ray crystallography in 1996 due to advances in RNA synthesis and purification. The RNA duplex $r(C_4G_4)_2$ at 1.46 Å resolution provided insight into the role structured water plays in secondary structure stabilization and recognition.[58] Comparison of hydration to a high resolution A-DNA duplex[59] $[d(GG^{Br}UA^{Br}UACC)]_2$ revealed a higher degree of hydration surrounding the sugar-phosphate backbone of the RNA duplex. This is due to the additional 2′-OH group on the ribose moiety. As has been observed before in A-form duplexes, the phosphate backbone is more compactly spaced, allowing water molecules to form bridging interactions between intrastrand phosphates. These water bridges enthalpically stabilize the RNA helix. This excess amount of hydration in the first solvation shell has been demonstrated to impart thermodynamic stability relative

to the B-form deoxyribose counterpart $d(C_4G_4)_2$. Under similar low ionic buffer conditions, a 25.5 °C difference in thermal melting points was observed between $d(C_4G_4)_2$ and $r(C_4G_4)_2$ resulting in a free energy difference ($\Delta\Delta G$) of –7.1 kcal/mol (at 37 °C). This observation was determined to be enthalpically driven, which is in contrast to the entropically stabilized DNA duplex.[60] Due to the higher ordered hydration shell, ligand binding to RNA pays a higher entropic cost compared to DNA.[58]

Experimentally, the displacement of water molecules upon ligand binding is thermodynamically difficult to quantify. Normally, the hydration shell and ions around DNA are released into the bulk solvent upon binding, causing a positive change in entropy.[61,62] Recent studies have shown this hypothesis to be more complicated as osmotic stress studies have determined an uptake of water upon ligand binding.[62] Other examples have shown water to mediate ligand binding, as is the case with paromomycin binding in the prokaryotic A-site. Instead of displacing water, solvent molecules are locally rearranged and not lost from the groove.[63,64] Despite the difficulty in identifying the precise location and behavior of water molecules, experimental results demonstrate its importance in the interaction of nucleic acids with small organic ligands.

1.6 THERMODYNAMICS OF BINDING

Partitioning free energies involved with nucleic acid-ligand recognition can provide insight into how water and ions mediate binding. Chaires proposes

partitioning individual free energies associated with different molecular motions and interactions into a linear relationship for the overall free energy of ligand binding (ΔG_{obs}):[49]

$$\Delta G_{obs} = \Delta G_{DNA} + \Delta G_{r+t} + \Delta G_{hyd} + \Delta G_{pe} + \Delta G_{mol}$$

ΔG_{DNA} is the energetic cost of distorting the DNA structure upon ligand binding. This is observed mainly for intercalators due to the 3.4 Å separation and unwinding between base pairs at the site of binding. In contrast, groove binders minimally perturb nucleic acid structure. Upon binding, limiting the degrees of freedom in translation or rotation normally result in an entropic loss upon complex formation (ΔG_{r+t}). One driving force for complex formation is the hydrophobic effect. The binding a hydrophobic ligand to a hydrophobic macromolecule (ΔG_{hyd}) is a favorable contribution to the overall free energy. The energy from counter ion and water release is accounted for in the polyelectrolyte term (ΔG_{pe}). Finally, weak forces (ΔG_{mol}) such as van der Waals, hydrogen bonding, and dipole-dipole interactions are also included.

Although separate molecular interactions are ligand and sequence dependent, approximate quantities can be derived from well-known nucleic acid binders. A comparison of an intercalator and groove binder show how each term influences binding. A parsing of the binding free energies for groove binder Hoechst 33258 (Figure 1.16) reveals no energetic penalty from structural DNA distortion (ΔG_{DNA}) upon binding but a penalty of approximately +15 kcal/mol for loss of molecular rotation and translation from binding (ΔG_{r+t}) was calculated.[49,65] The hydrophobic effect is the major driving force of binding (ΔG_{hyd} = –15 kcal/mol) while there are

minimal contributions from release of condense ions (ΔG_{pe} = –2 kcal/mol) and weak intermolecular forces (ΔG_{mol} = +1). The overall free energy (ΔG_{obs}) for binding is approximately ~ –12 kcal/mol. In comparison, analysis of intercalator ethidium bromide binding results in a small energetic penalty due to distortion of base pairs surrounding the binding site (ΔG_{DNA} = +4 kcal/mol).[66] A similar energetic penalty compared to Hoechest 33258 is observed due to loss of rotational freedom (ΔG_{r+t} = +15 kcal/mol). The release of condensed polyelectrolytes is a minimum driving force (ΔG_{pe} = –2 kcal/mol). The free energy from hydrophobic interactions (ΔG_{hyd} = –12 kcal/mol) is half of the amount for Hoechst 33258 while weak intermolecular forces play an extremely large role in ethidium binding (ΔG_{mol} = –12 kcal/mol). Most likely the π–π stacking and van der Waals interactions of intercalation account for the large terms. The overall observed free energy of binding for ethidium (ΔG_{obs} = –7 kcal/mol) is smaller than Hoechest 33258, but the contributions from ΔG_{mol} may be limited due to ethidium's smaller size.

Experimentally, measuring the separate free energy contributions from ligand–nucleic acid interactions is difficult, since all thermodynamic quantities are interrelated, but approximate estimations are possible. Analysis of ligands that represent distinctive binding modes provides insight into the dominant factors that influence these modes, and can potentially influence the design of future ligands. Since intercalators are influenced by van der Waals forces, those possessing large polyaromatic surfaces increase contact with nucleic acid bases, thereby increasing

affinity and specificity. This has been observed with triple helix selective binders that exploit the larger surfaces of base triplets.

1.7 SMALL MOLECULE STRATEGIES TO TARGET NUCLEIC ACIDS

Understanding of small molecule–nucleic acid interactions was initially based on experimental data with natural dyes. Originally, modified acridines such as proflavine were thought to only interact with the exterior of the DNA double helix,[67] but increased solution viscosity of acridine-DNA duplexes prompted Leonard Lerman to suggest a different theory. In 1961, Lerman proposed his "intercalation" hypothesis describing the insertion of planar aromatic molecules, in this case, proflavine, between the bases of DNA (Figure 1.10).[38] Although fluorescent dyes such as acridine orange were already known to stain the nuclei of cells and were utilized as antiseptics and antimalarials by Paul Ehrlich, the publication of Lerman's hypothesis is considered the beginning of selective targeting of nucleic acids with small organic molecules.[68]

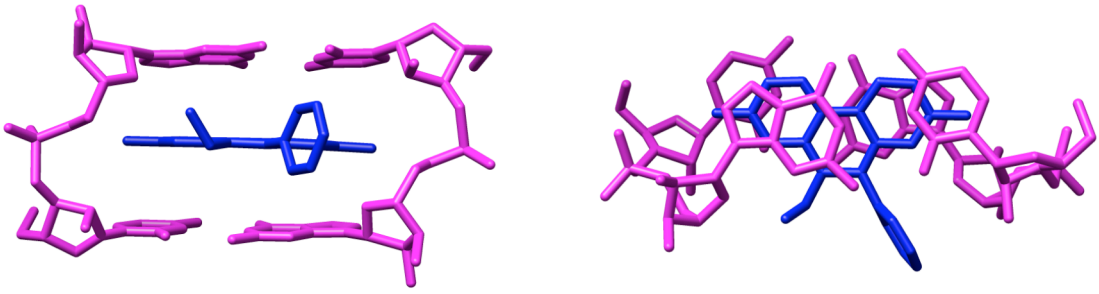


Figure 1.10. Intercalation of ethidium bromide (blue) between dimeric C•G dinucleotides. The side view (left) illustrates the van der Waals radius of ethidium, while the top view (right) demonstrates the degree of base stacking. NDB Code: DBR007. (Jain, S.C., Sobell, H.M. Visualization of drug–nucleic acid interactions at atomic resolution. VIII. Structures of two ethidium dinucleoside monophosphate crystalline complexes containing ethidium: cytidylyl (3'–5') guanosine. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1179–1194.)

Lerman proposed two perturbations affect DNA structure upon ligand intercalation: a) DNA helical length would increase proportionally to acridine concentration, and b) the supercoiling of DNA would be diminished due to helix extension. Upon intercalation, the neighboring base pairs separate approximately 3.4 Å to provide space for the incoming planar molecule. This distortion lengthens the double helix and increases the overall solution viscosity. Helical unwinding is also necessary for accommodation of the intercalator and can be independently observed to verify intercalation as a mode of binding (normally, helix unwinding is not observed with groove binders). Insertion of ethidium reduces the normal DNA helical twist from 36° to 10°, unwinding the helix by 26°. Acridines like proflavine have been shown to unwind the double helix by 17° while anthracyclines daunorubicin and adriamycin unwind the intercalation site by 11°. The distortion of adjacent base pairs by intercalation prevents subsequent binding between the adjacent base pairs. This is known as the nearest-neighbor principle.[37]

Lerman's hypothesis was followed by subsequent biophysical investigations. In 1965, Michael Waring observed the formation of a drug–DNA complex using UV-visible spectroscopy by monitoring absorbance changes of ethidium bromide intercalating into DNA.[69] In 1967, Le Pecq and Paoletti monitored ethidium binding by fluorescence spectroscopy.[70] These early biophysical experiments provided insights into the interaction of low molecular weight molecules and nucleic acids.

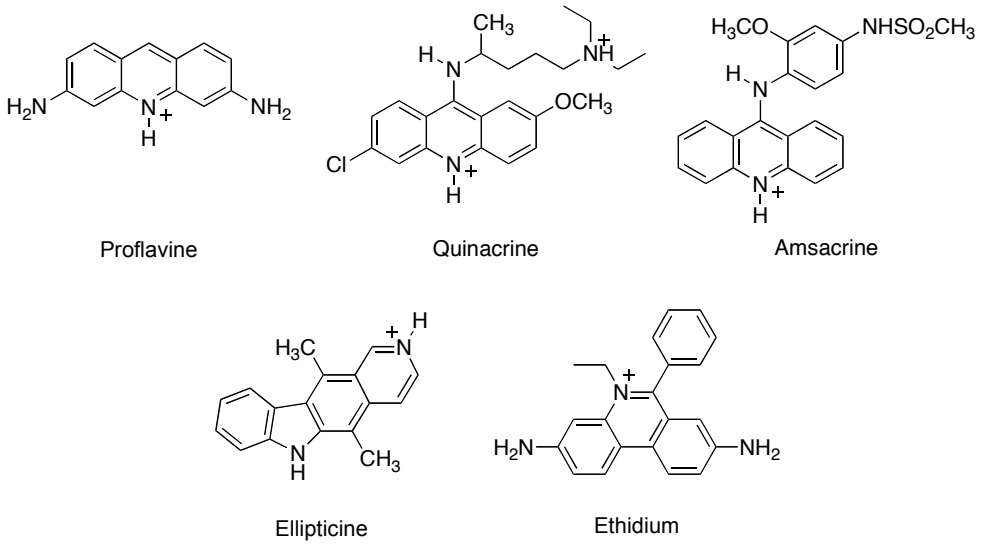


Figure 1.11. Examples of simple intercalators.

Many dyes such as proflavine and ethidium are simple intercalators, consisting mostly of an aromatic core that mediates binding by insertion between base pairs. Other simple intercalators include quinacrine, amsacrine, and ellipticine (Figure 1.11). Structurally more complex intercalators contain positively charged side chains, which allow for interaction with the major and minor grooves (Figure 1.12). In addition to the enthalpic gain from base stacking interactions in intercalation, groove binding side chains like sugars and peptides increase van der Waals interactions and hydrogen bonding.[37] One example is actinomycin D which is used clinically as an anti-tumor agent. Other complex intercalators include daunomycin and nogalamycin, potent anti-tumor antibiotics that can bind both the major and minor grooves in addition to intercalation (Figure 1.13). Porphyrins are considered the quintessential complex intercalators, since they contain four pendent chains where two simultaneously bind the major and minor grooves.

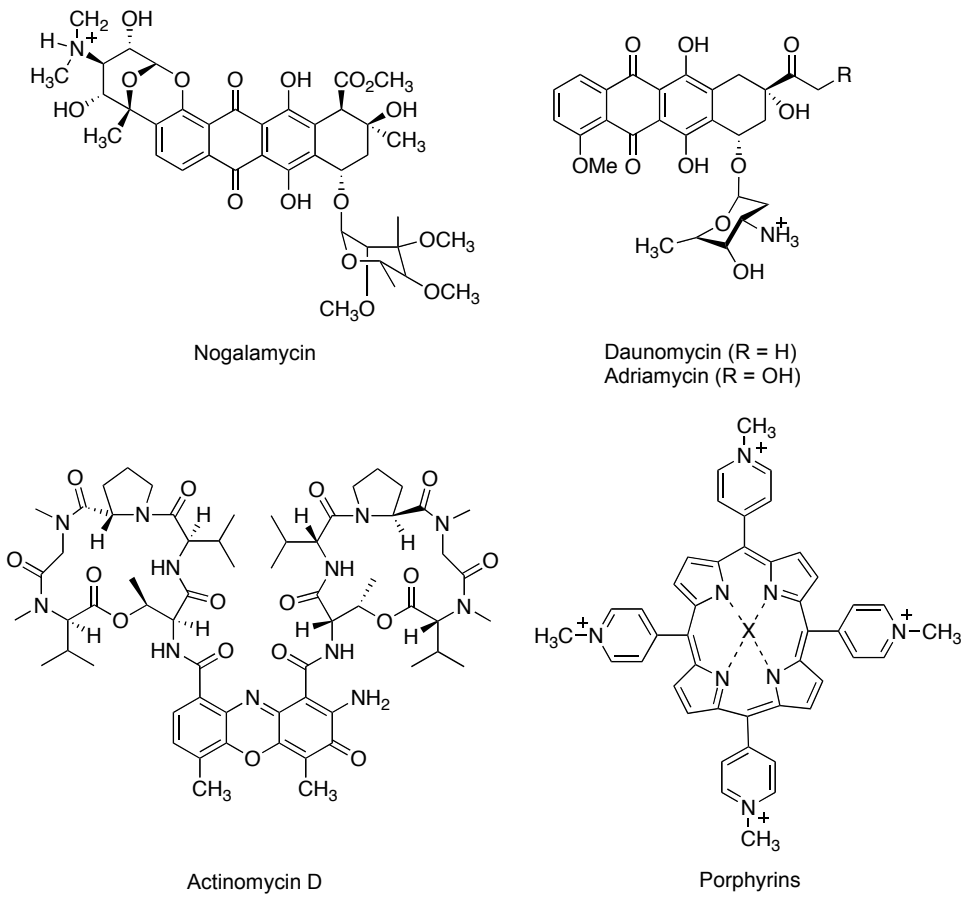


Figure 1.12. Examples of complex intercalators.

One strategy to increase the binding affinity of small molecules to DNA is to utilize polyintercalating scaffolds. In the 1970's, Waring observed echinomycin bound DNA in a "bifunctional mode of intercalation".[71] Additional efforts have followed with the design of synthetic bisintercalators to further exploit π–π stacking interactions and hydrophobic effects seen with simple intercalators such as acridine (Figure 1.14).[72] A daunorubicin-based bisintercalator, WP631, has been reported to have extremely high affinity for herring sperm DNA with an approximate binding constant of 3.1×10^{11} M^{-1} (Figure 1.15).[73] In another example, Dervan and coworkers

synthesized a dimeric phenanthradinium derivative, bis(methidium)spermine.[74] The linker, spermine, is a polyamine known to already bind nucleic acids with high affinity and lies within a groove of DNA.[74] In comparison to the monomer ethidium bromide, bis(methidium)spermine has over a 1000-fold higher affinity for calf thymus DNA, poly d(G•C)•d(G•C), poly d(A)•poly d(T), and poly r(A)•poly d(T).[75] Moreover, selectivity was observed for A-form poly r(A)•poly d(T) over B-form poly d(A)•poly d(T), suggesting specificity for different nucleic acid conformations, most likely due to a difference in polymorph conformation.[75] Sequence-specific bisintercalators have also been reported, combining the principles of intercalation and sequence recognition from groove binders (discussed below).[76]

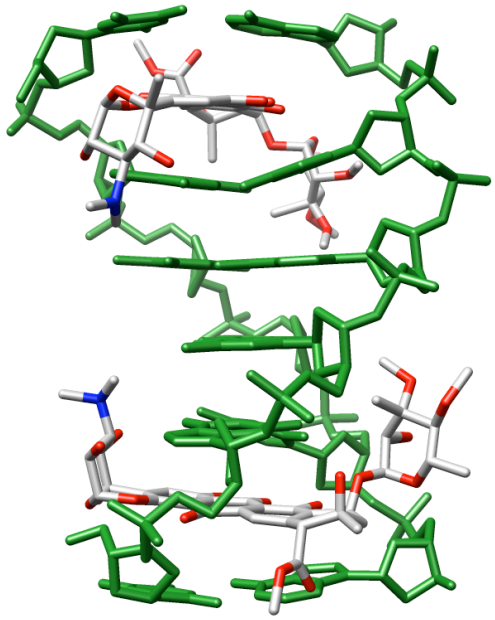


Figure 1.13. X-ray crystal structure of a 2:1 nogalamycin:DNA complex. Intercalation of nogalamycin (carbons, grey; oxygens, red; nitrogen; blue) distorts the DNA duplex (green) upon binding; depicted by the bowing of the Watson–Crick base pairs. Note, the protonated amine of nogalomycin lies in the electrostatically more positive major groove. PDB ID: 258D. (Smith, C.K., Brannigan, J.A., Moore, M.H. Factors affecting DNA sequence selectivity of nogalamycin intercalation: the crystal structure of d(TGTACA)$_2$-nogalamycin$_2$. *J. Mol. Biol.* **1996,** 263, 237–258.)

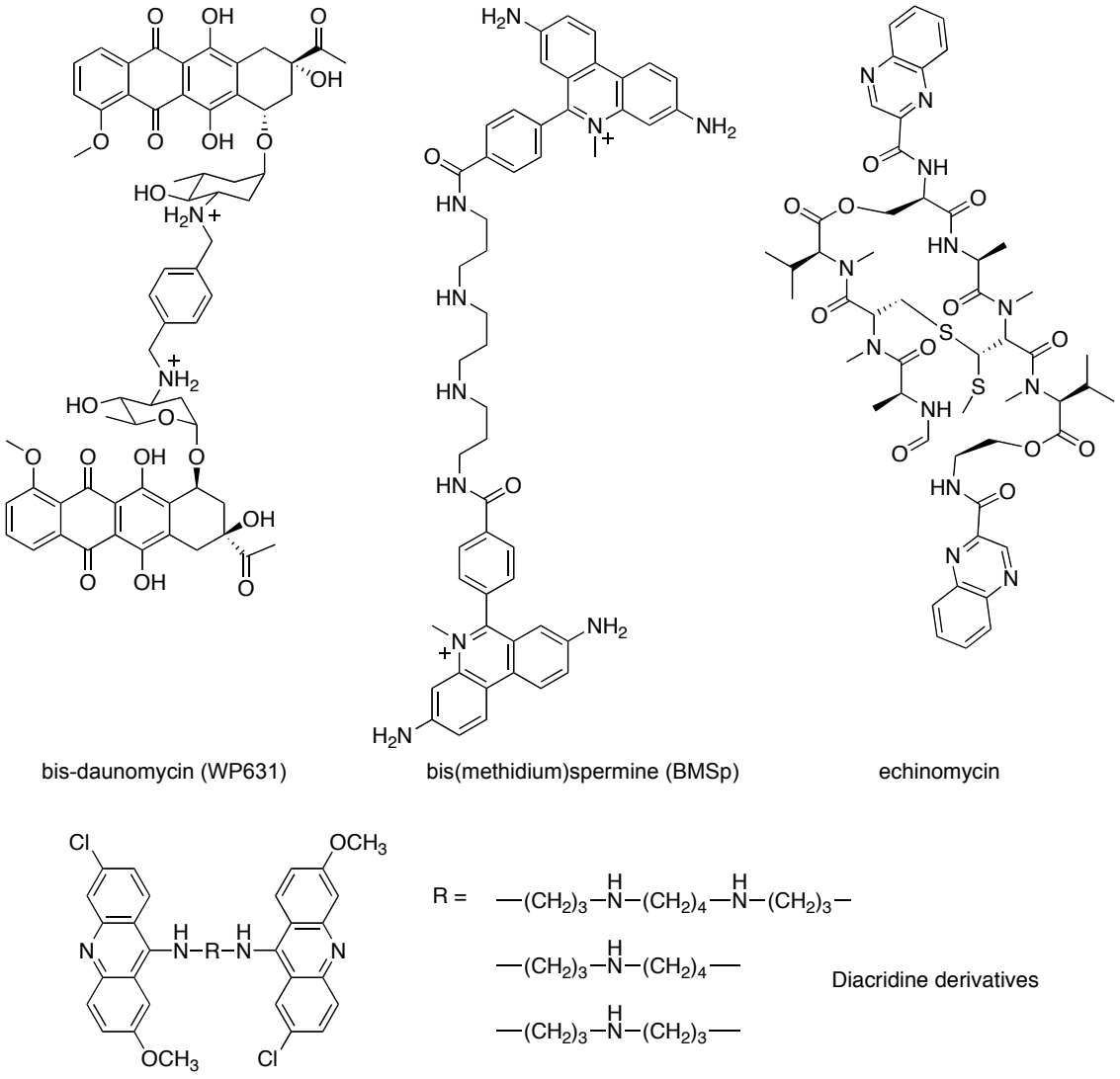


Figure 1.14. Examples of bisintercalators.

Numerous dyes have been shown to bind nucleic acid grooves, exploiting van der Waals and hydrogen-bonding interactions. Fluorescent stains Hoechst 33258, berenil, pentamidine and DAPI are all groove binders (Figure 1.16). Instead of the fused aromatic ring architecture seen in intercalators, groove binders are characterized by linked heterocycles that are usually crescent-shaped. At lower pH's, the heterocyclic nitrogens become protonated which complements the overall negative

charge of DNA and assists in binding. In addition, van der Waals interactions and hydrophobic contacts are driving forces for groove binding (Figure 1.17). These thermodynamic forces have to overcome the energetic penalty of nucleic acid destabilization from displacement of the spine of hydration in the minor groove. Many groove binders are known to bind A•T rich duplexes and have selectivity for B-form DNA over A-form RNA or DNA.[33,36] The absence of the sterically bulky 2-amino guanine moiety and large, deep minor groove of DNA are possible explanations.[77]

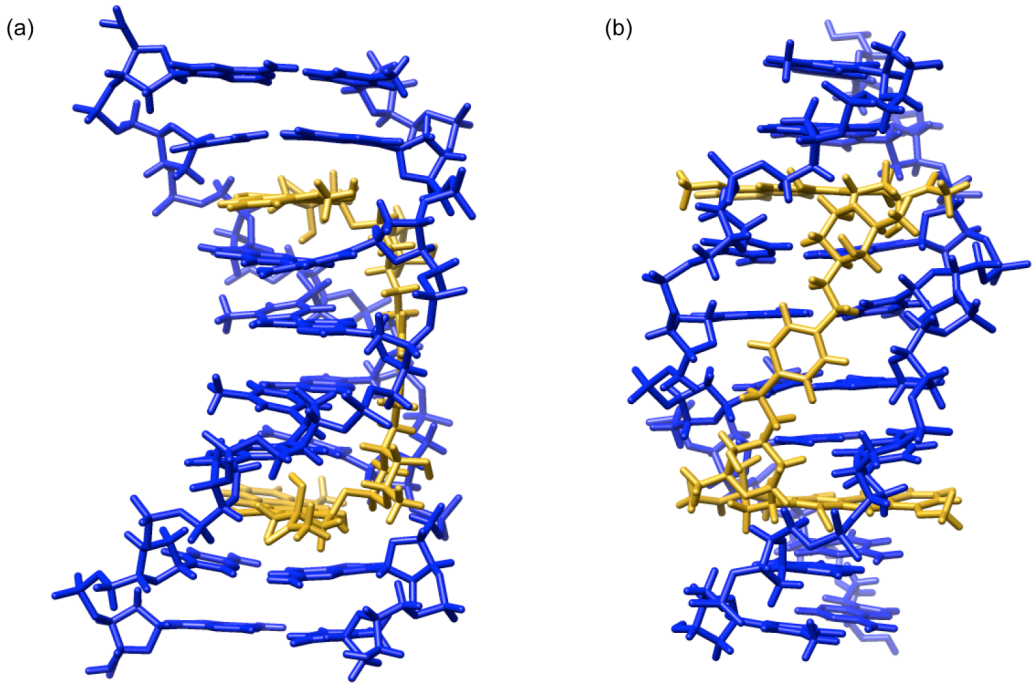


Figure 1.15. NMR structure of bisintercalator bis-dauonrubicin (goldenrod) binding into a DNA duplex (blue). (a) Side view of the spermine linker along the major groove. (b) View of the major groove looking onto the spermine linker in the major groove; rotated approximately +90° from view (a). PDB ID: 1AL9. (Robinson, H., Priebe, W., Chaires, J.B., Wang, A.H. Binding of two novel bisdaunorubicins to DNA studied by NMR spectroscopy. *Biochemistry* **1997**, *36*, 8663–8670.)

Two well-known groove binders, neotropsin and distamycin, are naturally occurring antibiotics that have been used extensively in the clinic.[78] High resolution X-ray crystal structures have revealed the intricate interactions between these groove binders and short DNA oligonucleotides (Figure 1.17).[61,79] For netropsin, the DNA double helix is not extensively elongated or perturbed upon binding (in contrast to intercalation). A slight perturbation is observed in the widening of the minor groove (0.5–2.0 Å), and an 8° helical bend toward the binding site per molecule is also seen. Molecular simulations support such flexibility in the minor groove of AT-rich tracts.[80] The netropsin amide protons hydrogen bond with adenine N3 and thymine O2 in the minor groove, the same functional groups that are bridged by the spine of hydration,

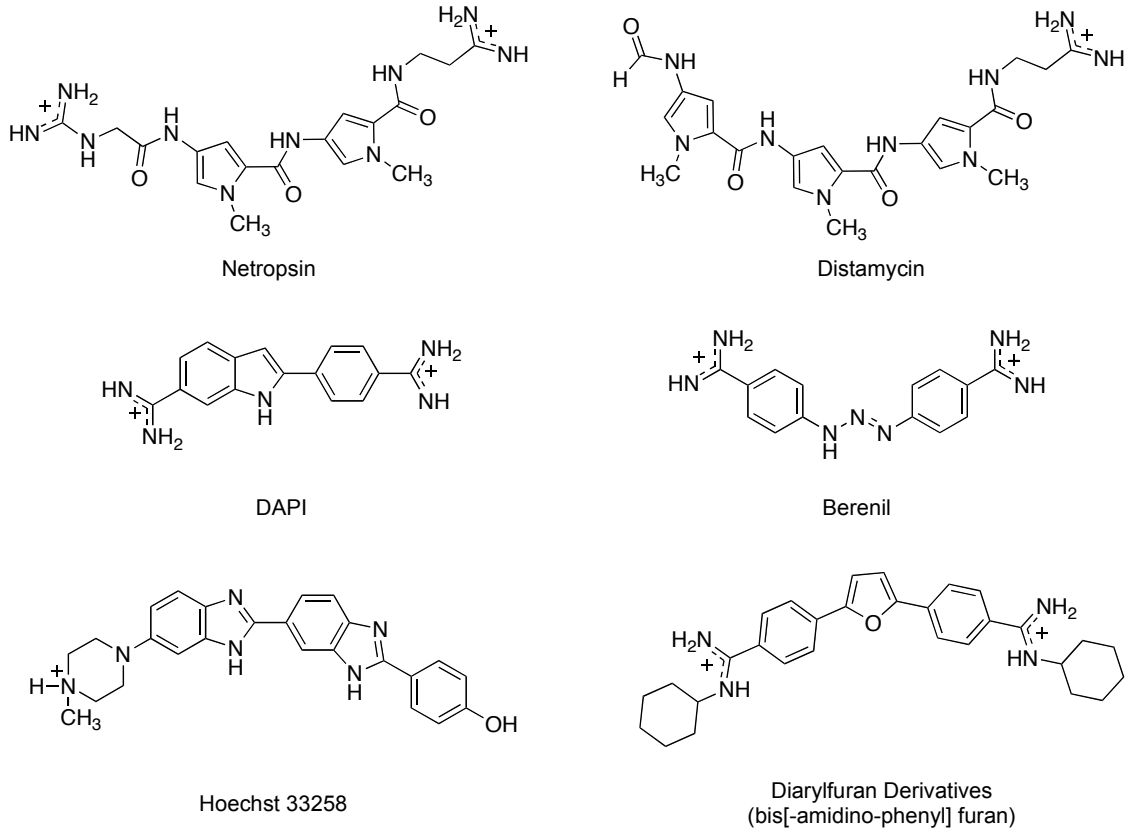


Figure 1.16. Examples of groove binders.

which is expelled upon binding (Figure 1.18). In addition, the 2-amino group of guanine would interrupts binding by sterically clashing with a netropsin-like molecule. To avoid the steric clash from the exocyclic amino group, replacing methylpyrrole with methylimidazole allows for hydrogen-bonding interactions between G and the imidazole nitrogen (Figure 1.19).[81] These *lexitropsins* conceivably could recognize A•T or G•C tracts by replacement of the pyrrole building blocks with imidazoles;[82] however, the G•C selectivity of lexitropsins has been limited at best.[36]

The holy grail in modulating gene expression is selective targeting of long, specific sequences of DNA. Netropsin and distamycin have provided insight into how

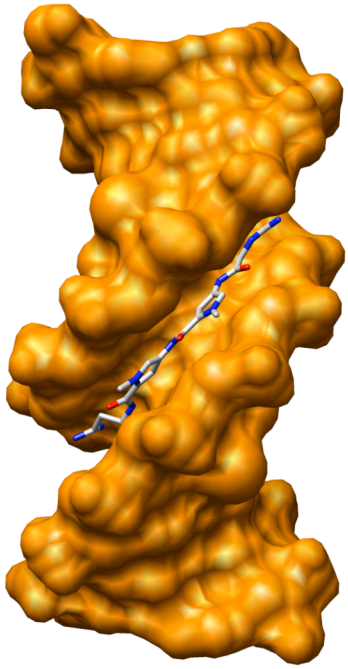


Figure 1.17. X-ray crystal structure of a 1:1 netropsin:DNA complex. Netropsin (carbons, grey; amines, blue; oxygens, red) resides in the minor groove of the DNA helix (orange). PDB ID: 6BNA. (Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R.E. Binding of an antitumor drug to DNA, netropsin and CGCGAATT^{Br}CGCG. *J. Mol. Biol.* **1985**, *183*, 553–563.)

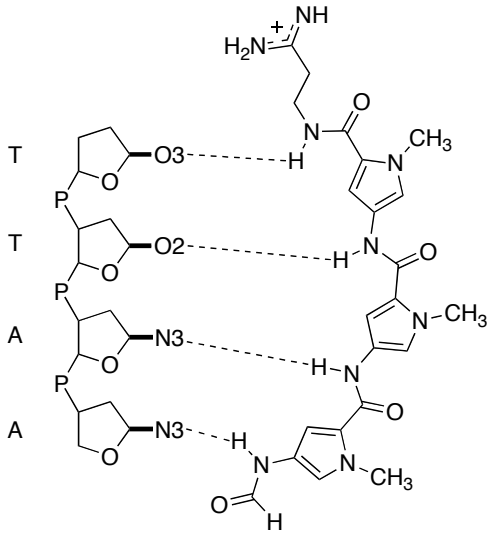


Figure 1.18. Schematic representation of the hydrogen-bonding contacts between distamycin and AATT bases in the minor groove of DNA. The nucleic acid structure has been minimized with base identities indicated along the left column. The DNA sugar-phosphate backbone is represented by a reduced ribose and "P" (for phosphorous). Only the purine or pyrimidine atoms involved in hydrogen bonding are indicated. Hydrogen-bonding is represented by dashed lines. (Adapted from Ref. 36).

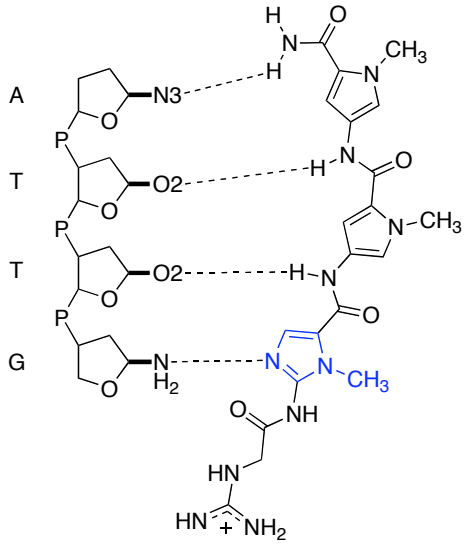


Figure 1.19. Schematic representation of the hydrogen-bonding contacts between a PyPyIm "lexitropsin" and ATTG bases in the minor groove of DNA. Note, the additional imidazole (blue) nitrogen allows for hydrogen-bonding with the N2 of guanosine which normally clashes with a pyrrole hydrogen. The nucleic acid structure has been minimized with base identities indicated along the left column. The DNA sugar-phosphate backbone is represented by a reduced ribose and "P" (for phosphorous). Only the purine or pyrimidine atoms involved in hydrogen bonding are indicated. Hydrogen-bonding is represented by dashed lines.

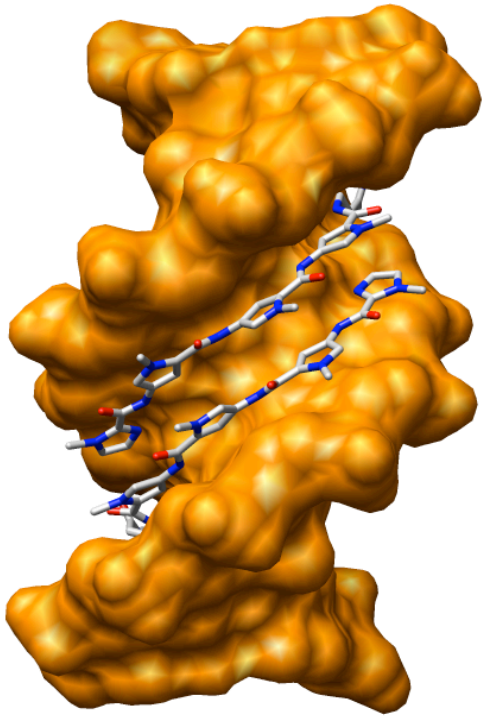


Figure 1.20. A 2:1 complex of ImPyPy with a DNA duplex in the minor groove. The synthetic polyamide ImPyPy (carbons, grey; oxygens, red; nitrogens, blue) molecules stack on top of one another in a head-to-tail orientation in an expanded minor groove of DNA (orange). PBD ID: 408D. (Kielkopf, C.L., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B., Rees, D.C. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **1998**, *282,* 111–115).

natural polyamide molecules are able to selectively target certain base sequences. However, up to this point, no broad principles governing nucleic acid recognition existed. A chance observation that the natural antibiotic distamycin binds the minor groove of DNA in a 2:1 ratio[83] set into motion the development of the first sequence-specific binding molecules (Figure 1.20).[33] Realizing polyamides selectively targeted certain bases through NMR studies,[84] Dervan was able to develop a pairing system of heterocycles that specifically bind short DNA sequence using hairpin polyamides (Table 1.2).[33] Sequence affinities have been reported in the nanomolar range. An

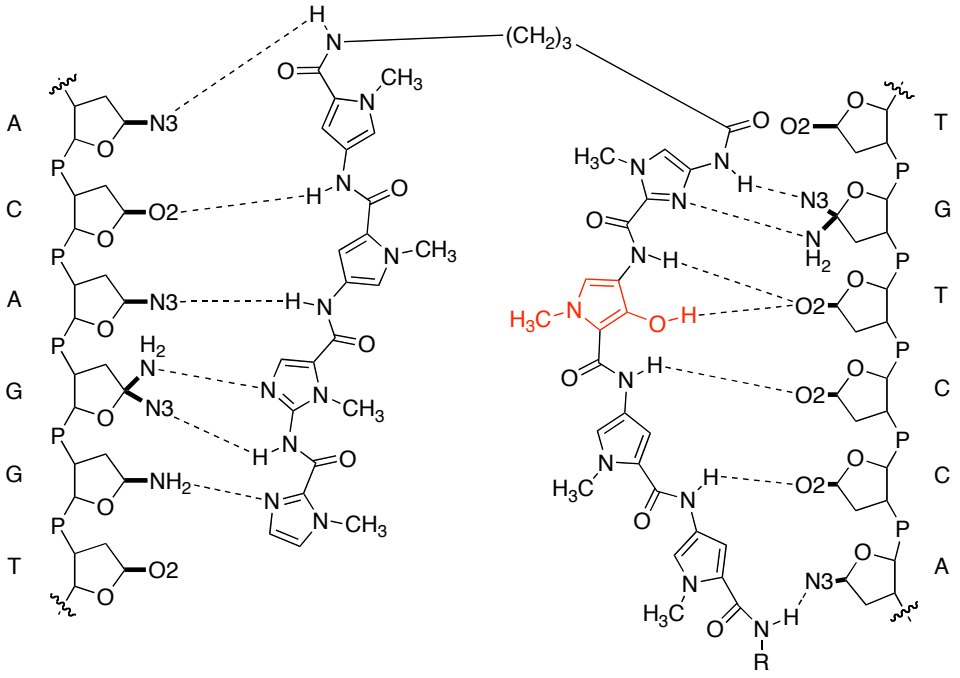


Figure 1.21. Schematic representation of the hydrogen-bonding contacts between a sequence-selective hairpin polyamide in the minor groove of DNA. Note, the hydroxypyrrole (Hp, red) allows for discrimination of an T•A base pair over A•T due to additional hydrogen-bonding interactions. The nucleic acid structure has been minimized with base identities indicated along the left and right columns. The DNA sugar-phosphate backbone is represented by a reduced ribose and "P" (for phosphorous). Only the purine or pyrimidine atoms involved in hydrogen bonding are indicated. Hydrogen-bonding is represented by dashed lines. (Adapted from Ref. 36).

imidazole/pyrrole (Im/Py) pair can selectively target G•C base pairs while a pyrrole/imidazole (Py/Im) pairing can instead target C•G base pairs selectively. At first, pyrrole/pyrrole (Py/Py) pairs indiscriminately targeted A•T or T•A base pairs, but the development of hydroxypyrrole (Hp) building blocks allowed for targeting T•A base pairs with hydroxypyrrole/pyrrole (Hp/Py) pairings and A•T base pairs with pyrrole/hydroxypyrrole (Py/Hp) pairings (Figure 1.21).[77] As verified through NMR and x-ray crystal structures, the amide protons in Dervan's modular polyamides make direct hydrogen-bonding contacts with the O2 of thymine and N3 of adenine in

DNA.[84] Moreover, guanine is selectively recognized with methylimidazole due to the hydrogen-bonding interaction between N2 of guanine and the nitrogen heterocycle (Table 1.2).

Table 1.2. Table of Dervan's Pairing Rules for Sequence Selective Targeting with Polyamides.

Pair	G·C	C·G	T·A	A·T
Im/Py	√	×	×	×
Py/Im	×	√	×	×
Hp/Py	×	×	√	×
Py/Hp	×	×	×	√

Im = imidazole; Py = pyrrole; Hy = hydroxypyrrole

1.8 TARGETING TRIPLE HELICES AND QUADRUPLEXES

Beyond targeting DNA double helices, other DNA structures have been the focus of therapeutic developments. Triple helices can be either intra- or intermolecular and are formed when a third DNA strand binds to the Hoogsteen face of base pairs in existing double helices (Figure 1.22). In 1957, the first triple helix was discovered from a 2:1 mixture of synthetic polynucleotides poly r(U) and poly r(A), respectively.[85,86] Since then, synthetic oligonucleotides have been successfully demonstrated to target sequence-specific stretches of DNA duplexes (Figure 1.23).[87] Targeting specific sequences has been investigated as an anti-gene strategy where designed oligonucleotides bind to DNA sequences that code for a disease gene and inhibit its expression.[88] Oligonucleotides have been demonstrated to chemically modify and oxidatively cleave their DNA targets, as well as photocrosslink nucleic acid substrates.[89,90] In addition, the formation of triple helices can interfere with the binding of transcription factors and other DNA maintenance enzymes, potentially affecting gene expression.[91] In the 1980s, H-DNA, a new polymorph of triple helical

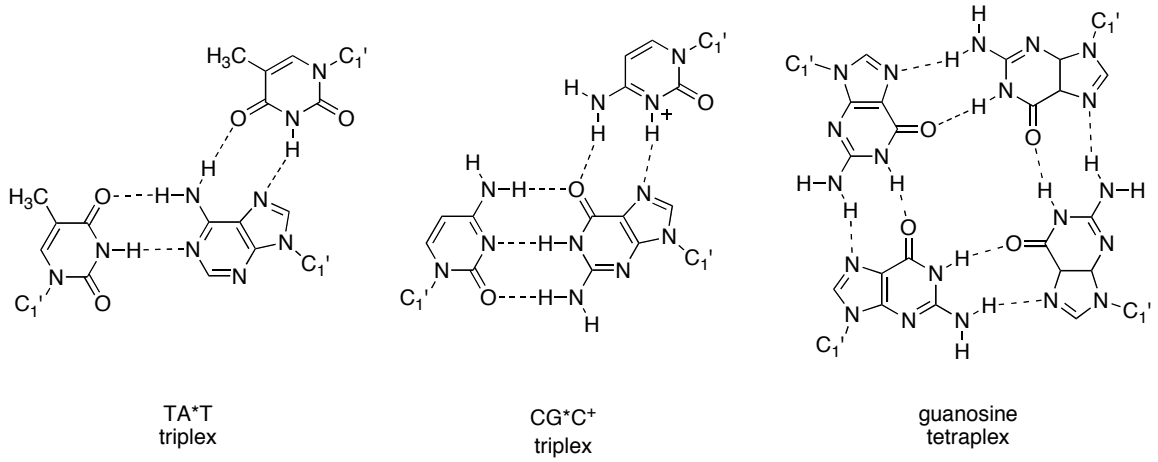


Figure 1.22. Hydrogen bonding pattern of PyPu×Py triplet base pairs and G-quadruplex base-pairs.

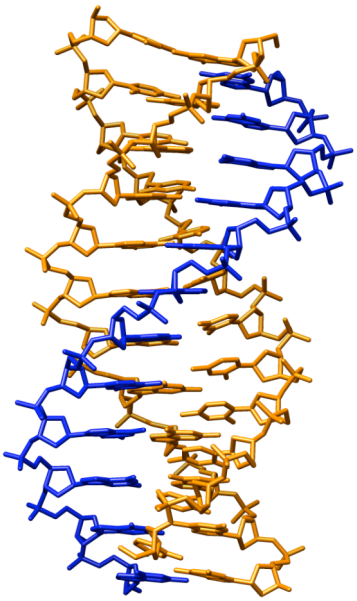


Figure 1.23. Schematic structure of a CG×G antiparallel triple helix. The poly(G)-oligonucleotide (blue) binds in the major groove of the C·G double helix (orange). NADB ID M1UDJ049. (Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B., Précigous, G., Kennard, O. *Science* **1996**, *273*, 1702–1705.)

DNA which forms along long tracts of repeating homopurine and homopyrimidine base pairs, was discovered to exist in vivo.[92] These mirror repeat sequences are

located near the promoter region of many genes, suggesting H-DNA plays an important role in modulating gene expression.

Triple helical structures formed naturally or by the addition of designed oligonucleotides, however, are unstable at physiological conditions. Short oligonucleotides that bind specific double helical sequences known to code for a particular gene can also be stabilized by small molecule ligands as another method for modulating gene expression (Figure 1.24). Selective binding of small molecules to triple helical DNA enhances triplex stability.[93] Stabilization of H-DNA in vivo may be a potential strategy to control gene expression (see Chapter 2).

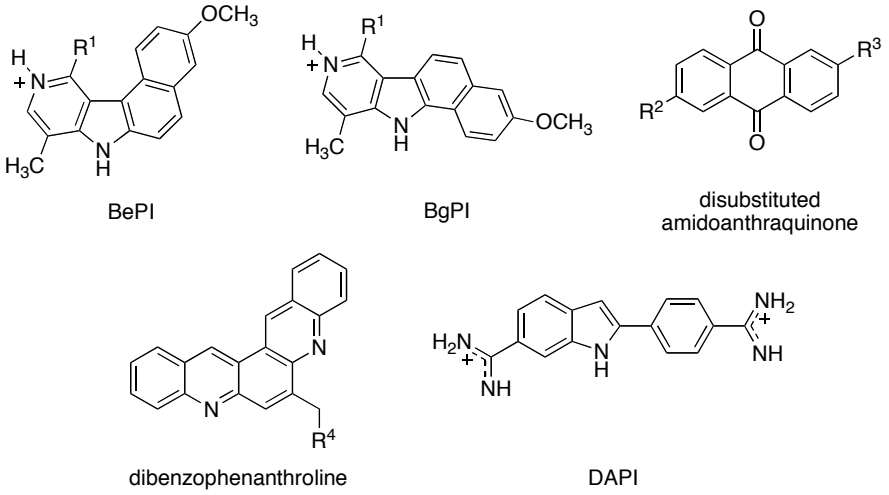


Figure 1.24. Representative structures of triple helix selective binders.

G-quadruplex DNA, a larger DNA structure, has also been the focus of therapeutic efforts recently.[94,95] Single-stranded DNA with long tracts of guanines form quadruplexes through Watson-Crick and Hoogsteen base pairing within all four bases (Figure 1.22).[96] Telomeres, specialized structures at the end of eukaryotic chromosomes, are thought to be rich in G-quadruplexes and suspected to play a role in

cell longevity.[97] Telomerase, an enzyme found in mostly cancer cells, normally extends telomeres, giving cancer cells the ability of indefinite proliferation.[98] G-quadruplexes, however, inhibit telomerase activity since the enzyme only recognizes single-stranded regions of DNA.[99] The development of quadruplex-specific ligands is currently an active field of research.[100]

1.9 THE NEED FOR RNA GENERAL RECOGNITION PRINCIPLES

As stated previously, until the mid-1990's, understanding small molecule–nucleic acid interactions was mostly centered around DNA and not RNA. The reasons for this were empirical. DNA is more stable than RNA and more easily synthesized. The lack of large amounts of high purity RNA oligonucleotides inhibited NMR and X-ray crystallography studies until the last 15 years where understanding RNA structure and its interactions with small molecules has seen the most growth.[18]

Within this time, targeting RNA for the development of therapeutics against viral and bacterial pathogens has been an area of active investigation.[17] In retroviruses, RNA encodes all the genetic information necessary for replication and function. HIV, arguably, is the most understood and well-studied retrovirus. Due to the high mortality rates and social stigma surrounding AIDS, efforts to discover drugs that inhibit the retroviral lifecycle have increased dramatically in the past decade.[101] Active anti-retroviral therapeutics programs also exist for other pathogens such as hepatitis, West Nile and Sudden Acute Respiratory Syndrome (SARS).[102–104] RNA-specific molecules such as aminoglycosides have been successful in inhibiting key

steps in the viral lifecycle. High-resolution X-ray crystal structures have revealed contacts between clinically prescribed antibiotics and the bacterial ribosome are solely with RNA and not protein components.[31] The dramatic increase in drug-resistant bacteria has placed extreme importance on understanding drug–RNA interactions.

Although RNA is viewed as a promising target in antiretroviral and antibacterial therapy, very little is understood about small molecule interactions with this biomolecule. Unlike DNA, with Dervan's pairing rules, few principles exist for sequence-specific targeting in RNA–small molecule recognition. Instead, RNA recognition has proven **not** to be sequence-selective at all but rather structure-selective.[105] Cellular RNA is primarily found in secondary structures such as internal and hairpin loops, bulges and stem junctions. The secondary and tertiary structure of RNA makes it more akin to proteins than to DNA; however, no recognition pattern has emerged for specific RNA–ligand pairing with these structural motifs, despite the numerous RNA–small molecule structures that have been visualized by NMR[106] or X-ray crystallography.[31,63,107,108] Although this variation in secondary structure may aid in RNA's function, it greatly hinders formulation of a general principle for RNA recognition. The complex and diverse structure of RNA in vivo is prohibitive to any application of broad recognition rules at this point. Ligand interactions observed in one distinct RNA structure cannot be applied to an RNA sequence with different secondary structure. Additionally, the differences in RNA's electrostatics and hydration pattern compared to DNA affect ligand recognition. Particularly, the

inaccessible RNA duplex major groove, with the highest density of electronegative potential, plays a key role for small molecule recognition.

The interactions between antibiotics and the ribosome have been a paradigm for understanding RNA–ligand interactions. The structural work of Westhof, Ramakrishnan, Yonath and Puglisi has provided insight into the direct contacts made between aminoglycoside antibiotics and the ribosome.[31,32,63,106] Westhof and Hermann have extensively studied the effects of electrostatics and metal ion binding in RNA recognition.[109] Additionally numerous groups including Tor, Hermann, Wong, Chung, and Griffey have all attempted to exploit these contacts in designing more effective aminoglycoside-based antibiotics (see Chapter 3). These efforts have resulted in the identification of a number of high affinity binders, but none has surpassed the clinical efficacy of the natural aminoglycosides. Despite advances in understanding RNA–ligand interactions in the past 15 years, crystal structures of bound and unbound RNA can provide only limited insight into designing new drugs. Since RNA is conformationally flexible or "plastic", X-ray crystal structures do not provide a full picture of the RNA in either state. Nonetheless, investigation of RNA–ligand interactions must at least be minimally based along the same structural lines that have dictated advancements in DNA–ligand interactions.

1.10 RNA STRUCTURE AND GROOVE ACCESSIBILITY

Double helical RNA is closely related to the analogous DNA double helix in structural dimension, but single-stranded RNA can form complex three-dimensional

structures more related to proteins.[5,110] Single-stranded regions of RNA are able to fold upon themselves to form hairpin loops, bulges, internal loops and junctions (Figure 1.25).[5] Hairpin loops can incorporate any number of base pairs, though a tetraloop is most commonly found.[37] Depending on the loop sequence, a variety of base-pairing configurations can be adapted. A GAAA loop has been observed to form a noncanonical G•A base pair while a UUYG loop forms a wobble base pair (Y = pyrimidine). Bulges exist when one or more residues are unpaired in an RNA helix. Although bulged RNA may adapt numerous structural arrangements, geometry is typically restricted to preserve base-stacking interactions, which contributes greatly to RNA stability by minimizing exposure of the hydrophobic bases to water.[105] Four characteristic RNA features accompany bulges: 1) stacking is distorted, 2) the RNA bends at the site of bulging, 3) helix stability is reduced, and 4) major groove accessibility increases.[37] Internal loops occur at mismatched base pairs such as noncanonic purine-purine or pyrimidine-pyrimidine interactions. Finally, junctions are formed at the merging of three or more RNA stems. At these junctures, mismatched or noncanonical base pairing can occur. Junctions have also been observed in tRNA, which gives it its famous cloverleaf structure.

The unique secondary structure RNA allows for recognition by proteins and small molecules. In a normal RNA duplex, the major groove, although narrow, possesses the greatest amount of electrostatic potential but is relatively inaccessible.[111] Disruptions to an RNA duplex expose the major groove allowing for access by small molecules and proteins.[111] Bulges and loops reveal the discriminatory faces of base

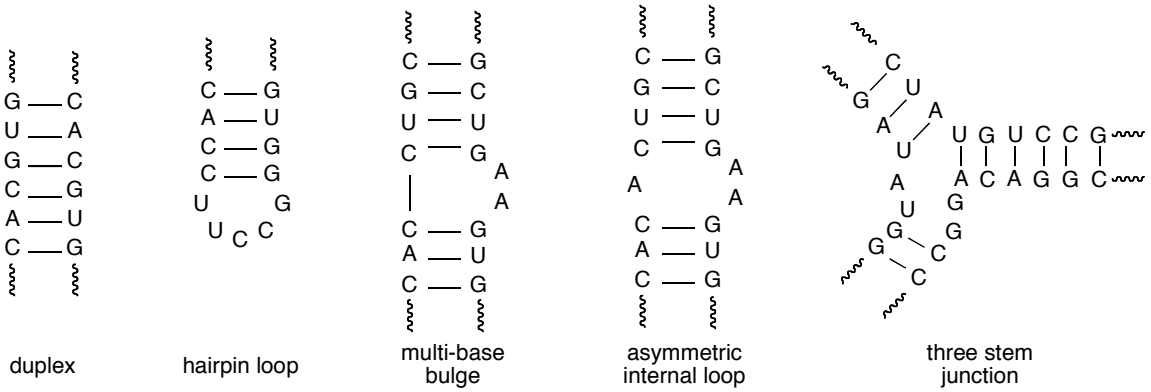


Figure 1.25. Schematic representation of different RNA secondary structures.

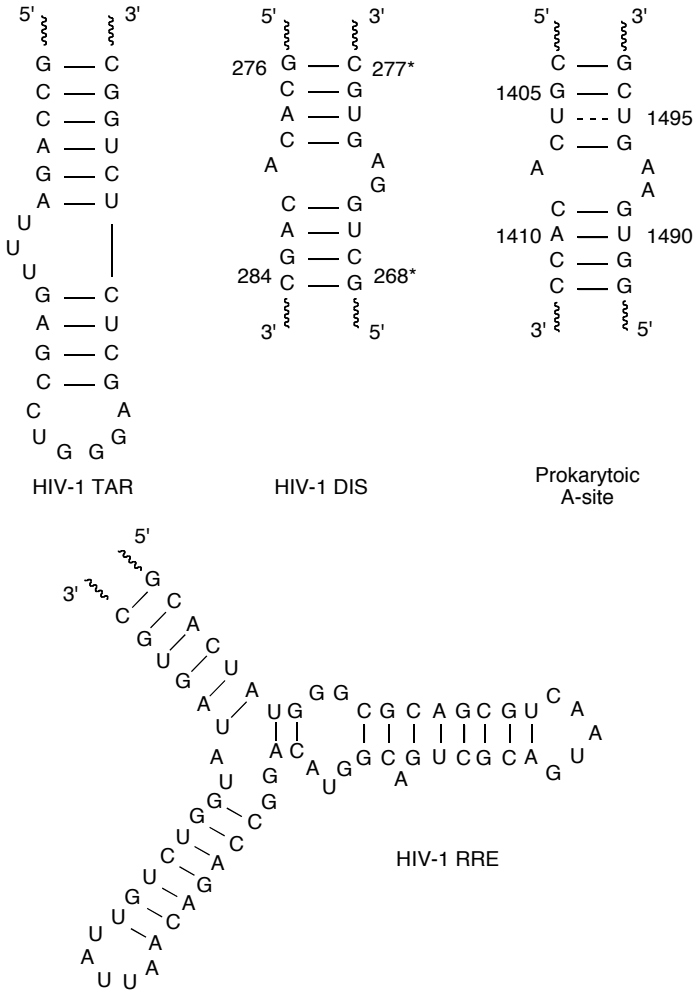


Figure 1.26. RNA secondary structure of therapeutically targeted RNA constructs: the HIV-1 TAR, DIS and RRE as well as the prokaryotic A-site.

pairs as well as distort the RNA backbone to widen the deep major groove.[105,111] For example, NMR studies have shown the Tat peptide binds the widened major groove of the HIV-1 TAR, This would not have been possible with the normal narrow groove of duplex RNA (Figure 1.26).[28] The asymmetric bulge of the ribosomal prokaryotic A-site, as well as the kissing hairpin loops of the HIV-1 DIS and the junction of the HIV-1 RRE, have been shown to accommodate certain small molecule drugs.[23,29,108,112] Current efforts of RNA targeting have all been focused on noncanonical secondary structures, as few drugs have been found to bind a RNA duplex.

1.11 RNA–SMALL MOLECULE INTERACTIONS: AMINOGLYCOSIDES AND THE RIBOSOME

The ribosome is a large riboprotein responsible for the assembly of cellular polypeptides from RNA templates. Composed of both RNA and protein, the large 70S ribosome is divided into two subunits — the larger 50S and smaller 30S subunit (Chapter 3). The ribosome and mRNA template orchestrate the synthesis of proteins by pairing a cognate tRNA to the appropriate mRNA codon. Numerous clinically prescribed antibiotics interfere with either binding of aminoacyl-tRNA, decoding mechanisms, translocating docked tRNA complexes, or releasing the growing polypeptide chain. Aminoglycosides are a class of antibiotics that interfere with the fidelity of protein synthesis by binding to the A-site (aminoacyl) and disrupt RNA conformational changes necessary for cognate aminoacyl-tRNA recognition (see Chapter 3 for discussion on the effects of aminoglycoside binding). Interactions of the

prokaryotic ribosome and aminoglycosides provide an immense amount of structural data for understanding small molecule–RNA interactions.

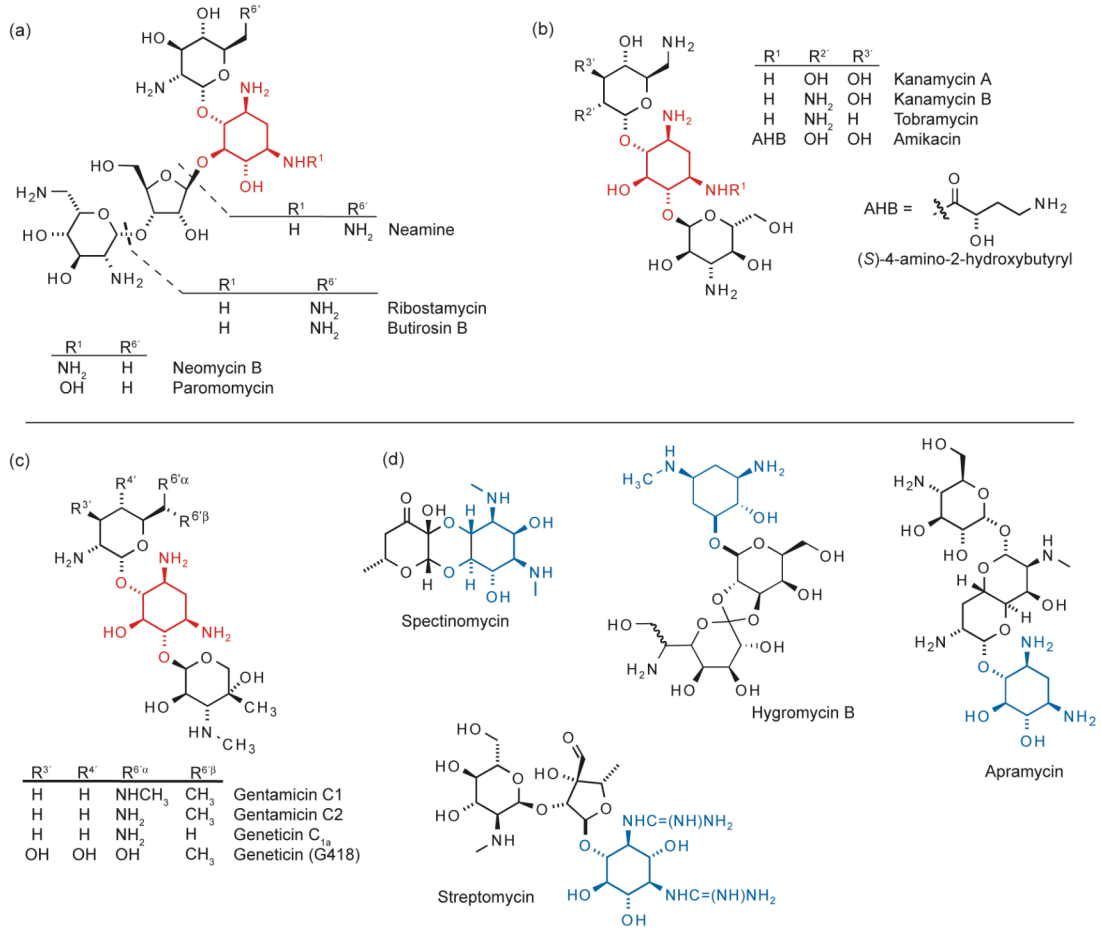


Figure 1.27. Structures of different classes of aminoglycosides. (a) Neomycin family. (b) Kanamycin and tobramycin family. (c) Gentamicin family. (d) Atypical aminoglycosides. Typical disubstituted 2-deoxystreptamine core (red) is depicted along with related streptamine cores (blue) found in atypical aminoglycosides.

Aminoglycosides have been clinically used since the 1950's (Figure 1.27). Streptomycin, the first aminoglycoside antibiotic, was the second antibacterial clinically administered after penicillin.[113] The isolation of other aminoglycoside antibiotics followed including neomycin B in 1949,[114] the kanamycin class in 1957,[115]

and the gentamicins in 1963.[116] Many aminoglycoside antibiotics are still used today but users can suffer from nephro- and ototoxicity.[117] These pseudo-oligosacharrides are characterized by a 2-deoxystreptamine (2-DOS) or streptamine core. The neomycin family consists of a 4,5-disubstituted 2-DOS core while the tobramycin/kanamycin class consists of a 4,6-disubstituted 2-DOS core. Additional families of aminoglycosides are characterized by fused rings or streptamine cores such as streptomycin, hygromycin, apramycin and spectinomycin (Figure 1.27).

Despite their heavy clinical usage, a direct structure-function correlation had not been determined for ribosomal antibiotics until two decades ago.[31,106,118] It is now known that aminoglycoside antibiotics which bind to the 30S small ribosomal subunit cause misreading during codon decoding and inhibit translocation of tRNA from the A-site.[30] In 1987, Moazed and Noller performed chemical footprinting studies to identify the exact locations of ribosomal binding of several antibiotics.[118] Streptomycin, tetracycline, spectinomycin, hygromycin and the neomycin class of aminoglycosides were all shown to interact directly with the 16S rRNA. Understanding the complexity of the largest cellular organelle was simplied when the prokaryotic A-site was shown to be effectively mimicked by smaller RNA oligonucleotides.[119,120] Even when docked with antibiotics, the minimized RNA resembled normal subunit function. Two years later, NMR spectroscopy studies by Puglisi and coworkers revealed the direct interactions made between aminoglycosides and a minimized A-site construct.[106] The internal loop created by mismatched base pairs U1406-U1495 and the unpaired adenines A1408, A1492, and A1493 created a

binding pocket for aminoglycosides. In 2000, these contacts were verified by X-ray crystallography in a seminal paper by Ramakrishnan.[31] More NMR and x-ray crystal structures of relative aminoglycosides including paromomycin,[63] apramycin,[121] tobramycin,[108] geneticin[107] have all revealed the intricate contacts involved in small molecule–RNA recognition, many of which require water and magnesium mediated hydrogen bonds. Aminoglycosides have also been found to bind other RNA structures such as the TAR, DIS, RRE, hammerhead ribozyme, and group I intron which are biologically relevant with diverse secondary structure. A combined effort of NMR and crystallography studies has been enacted to understand aminoglycoside interactions with these constructs as well. Interestingly, all the above RNA structures are not structurally related (with the exception of the prokaryotic A-site and HIV-1 DIS), but can be recognized with an aminoglycoside scaffold. The inherent flexibility of aminoglycosides has been suggested to be responsible for recognition of diverse secondary structures, and provides further evidence of structure selective recognition in RNA targeting. Again, this view greatly differs from the sequence selective binding of DNA.

1.12 RNA RECOGNITION MODELS

RNA recognition is a very sophisticated process and may soon achieve the same level of understanding enjoyed by sequence-selective targeting of DNA. Unfortunately, no extensive structure-activity relationship exists for RNA–ligand interactions, and there exists no RNA-selective ligands beyond the aminoglycosides.

The interaction between RNA structures and ligands can be described by two prevailing complementary models. The first theory is termed *conformational capture*, a modified version of the induced-fit theory.[122,123] The conformational capture theory hypothesizes that since RNA is a dynamic structure, it assumes numerous conformations that are both biologically functional and non-functional. Upon ligand binding, the RNA is locked into a conformation that is non-functional, preventing recognition by enzymes or other effector molecules. The second theory, *electrostatic complementarity*, states that since RNA structure dictates electrostatic potential in pockets and folds, selective binding ligands should have a complementary spatial electrostatic orientation to recognize these concentrated negative domains.[109,124]

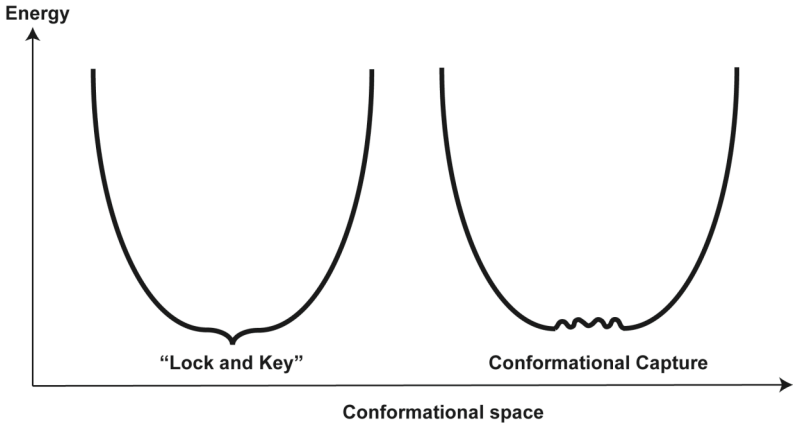


Figure 1.28. Energy landscape of the lock and key model and conformational capture model hypothesized to be associated with RNA-ligand recognition.

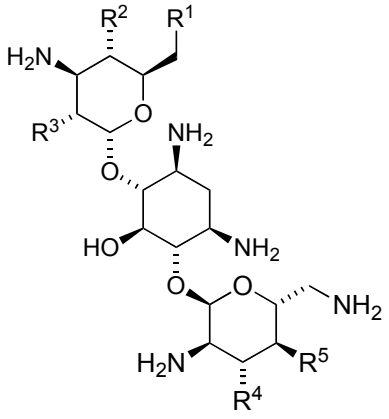
It is thought that conformational capture may mediate the initial phase of nucleic acid recognition.[123] Conformational capture has previously been applied to protein–RNA intermolecular recognition where proteins possess more than one folding pathway to reach a biomolecular complex energy minima.[125] The interaction

can be described in terms of an energy landscape (See Figure 1.28). In the canonical lock-and-key model, complex formation can be represented with an energy well with only one minimum signifying a thermodynamically stable complex. In a conformational capture model, the bottom of the energy well is characterized with numerous minima representing the potentially infinite stable conformations RNA can assume. The ligand can then "select" an RNA state, even if populated by the RNA for a short amount of time, and "capture" the RNA in this conformation. Although this model is similar to induced fit, if the free RNA already pre-exists in the appropriate complex conformation, then the ligand is locking this RNA structure and not inducing a new RNA conformation.[122] This is believed to be energetically favorable because no forced structural perturbations are required. The locking of this RNA structure may render it biologically non-functional as well as prevent the binding of proteins or cofactors. Additionally, by ligands targeting less populated conformations, the ligand does not need to compete with proteins associated with certain RNA structures that are more highly populated.[122,123]

A prototypical example that is consistent with conformational capture is the Tat protein binding to the HIV-1 TAR. Aminoglycosides that bind the TAR lock the asymmetrically bulged RNA into a conformation that is not recognized by the Tat peptide, interfering with normal viral replication. The prokaryotic A-site, with its asymmetric bulge is also locked into a conformation upon aminoglycoside binding that prevents cognate recognition of tRNA during protein synthesis.[126] Another

example of conformational capture is the binding of tobramycin to $tRNA^{Asp}$ which prevents further recognition by tRNA synthetase.[127]

Solely quantifying the energetics of RNA structural changes induced by ligand binding is difficult. Isothermal calorimetry is often used for quantifying binding thermodynamics, but overall calculated free energies (ΔG_{obs}) normally include contributions from hydrophobic effects, intermolecular forces such as hydrogen bonding and van der Waals interactions, release of condensed polyelectrolytes and reduction in rotational and translational degrees of freedom. Pilch and co-workers suggest that the energetic penalty from RNA base destacking caused by binding aminoglycosides to the prokaryotic A-site is less than 1 kcal/mol.[64] This estimation is based upon stacking studies by Turner[128] and Biltonen[129] with polymeric adenosine strands. Since such a small energetic penalty is paid for small distortions in RNA structure, the theory of conformational capture as a recognition pathway seems thermodynamically favored.



Aminoglycoside	R^1	R^2	R^3	R^4	R^5
1 Kanamycin B	OH	OH	OH	OH	OH
2 Tobramycin	OH	OH	OH	**H**	OH
3 4′-deoxytobramycin	OH	OH	OH	**H**	**H**
4 6″-deoxytobramycin	**H**	OH	OH	**H**	OH
5 4″-deoxytobramycin	OH	**H**	OH	**H**	OH
6 2″-deoxytobramycin	OH	OH	**H**	**H**	OH

Figure 1.29. Structures of different deoxytobramycin analogs synthesized by Tor and co-workers. Hydroxyl functional groups adjacent to amino groups were removed to determine their influence on amin pK_a and electrostatics.

Since RNA is a negatively charged polymer, electrostatics play a key role in recognition. A systematic structure-activity relationship (SAR) of electrostatics was initiated by Tor and co-workers using a series of deoxygenated tobramycin derivatives (Figure 1.29).[130,131] A comparison of pK_a values for ethylamine (pK_a = 10.7) and ethanolamine (pK_a = 9.5) suggests the presence of hydroxyl groups adjacent to an amine lowers the basicity of the amine. Deoxygenated tobramycin derivates were synthesized and their inhibitory effects on hammerhead ribozyme catalysis were compared (Figure 1.29).[130] Most deoxygenated derivatives had higher inhibitory activity compared to the parent tobramycin aminoglycoside. A derivative with the 2″-OH hydroxyl removed had the highest inhibitory activity. Despite the fact that 4″-deoxytobramycin is an isomer of 2″-deoxytobramycin, the 2″-position is adjacent to one amine and interacts through hydrogen bonds to another, affecting the basicity of two amines simulataneously. By modulating the basicity and, thereby, the electrostatics of amines, the charge density of the aminoglycoside ligands are changed so that binding is comparable if not better than the parent molecules.

In another study by the same group, amino-aminoglycosides were synthesized to address the underlying importance of electrostatics in RNA recognition (Figure 1.30).[131] Derivatives of tobramycin and kanamycin were synthesized with additional amino functionalities and tested for inhibition of hammerhead ribozyme cleavage. All "amino-aminoglycoside" derivatives showed higher cleavage inhibition activity compared to the parent molecule (Table 1.3). At 100 μM, kanamycin A diminishes cleavage of the ribozyme by 75%, while at 10 μM, no inhibition is observed. In

comparison, the analog 6″-amino-6″-deoxykanamycin A at 10 μM has inhibitory activity and a rate constant comparable to that of kanamycin A at 100 μM. This provides evidence that a single functional group modification can dramatically increase RNA binding. Even a potent RNA binder, neomycin B, can have enhanced affinity by addition of an amino functionality. At 1μM, the rate constant of inhibition is two-fold higher for the 5″-amino-5″-deoxyneomycin B analog compared to parent neomycin B.

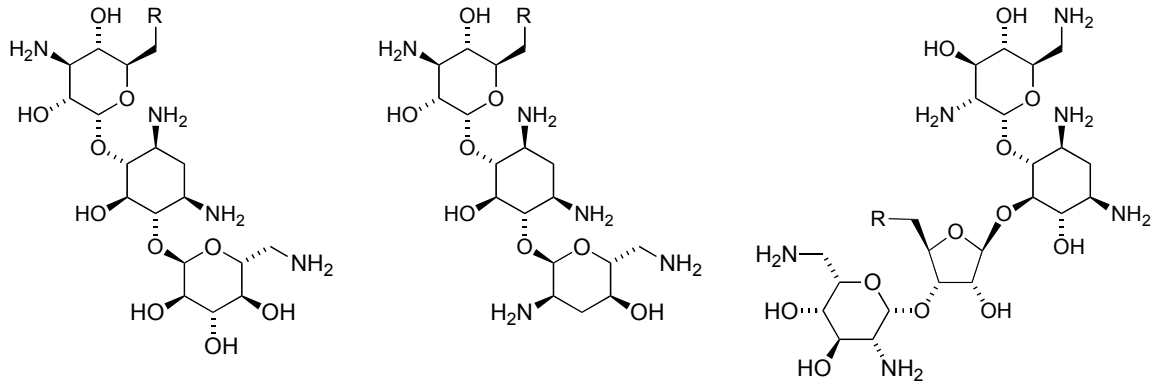


Figure 1.30. Structures of natural aminoglycosides and the corresponding "amino-aminoglycosides".

The orientation and three-dimensional charge distribution of ammonium groups in aminoglycosides are equal in importance to the total number of amine groups. A crystal structure of the hammerhead ribozyme revealed five magnesium ions intimately stabilized the RNA's tertiary structure (Figure 1.31).[132] Displacement of these magnesium ions by aminoglycosides is believed to mediate binding. Molecular dynamics simulations[124] suggest the interionic distances between Mg^{2+}–

Mg^{2+} are similar to the three dimensional arrangement and distances of neamine's ammonium groups.[133, 134] In addition to orientation, the electostatics of the positively charged ammonium groups are believed to be comparable to the charge of dicationic magnesium and complement the negative RNA binding pocket of the hammerhead ribozyme.[124] X-ray crystal structures of the ribosome have also shown binding of chloramphenicol to the 50S ribosomal subunit displaces magnesium ions.[135]

Table 1.3. Table of Rate Constants for Hammerhead Ribozyme Cleavage.

	k_2[min^{-1}]		
Derivatives	**100 µM**	**10 µM**	**1 µM**
none (control)		0.075	
kanamycin A	0.058	–[a]	–[a]
6″-amino-6″-deoxykanamycin A	0.018	0.050	–[a]
tobramycin	0.012	0.051	–[a]
6″-amino-6″-deoxytobramycin	–[b]	0.011	–[a]
neomycin B	–[b]	0.018	0.062
5″-amino-5″-deoxyneomycin B	–[b]	0.003	0.030

Insight into the thermodynamics of RNA binding through the electrostatics of charged ligands has been aided through studying metal ions and protonated aminoglycosides with isothermal calorimetry.[64] Recall, paromomycin and neomycin differ at the 6′-position by a hydroxyl or amino functionality, respectively; however, neomycin binds the hammhead ribozyme more tightly by one order of magnitude. This is due to the positively charged ammonium groups, along with additional hydrogen-bonding and van der Waals interactions due to its presence. Molecular modeling has shown eleven hydrogen-bonding contacts for neomycin with the prokaryotic A-site versus eight for paromomycin.[64] Following Chaires' proposed partition of free energies of molecular processes involved in nucleic acid binding, Pilch and co-workers quantified these terms through ITC studies.[64] The free energy

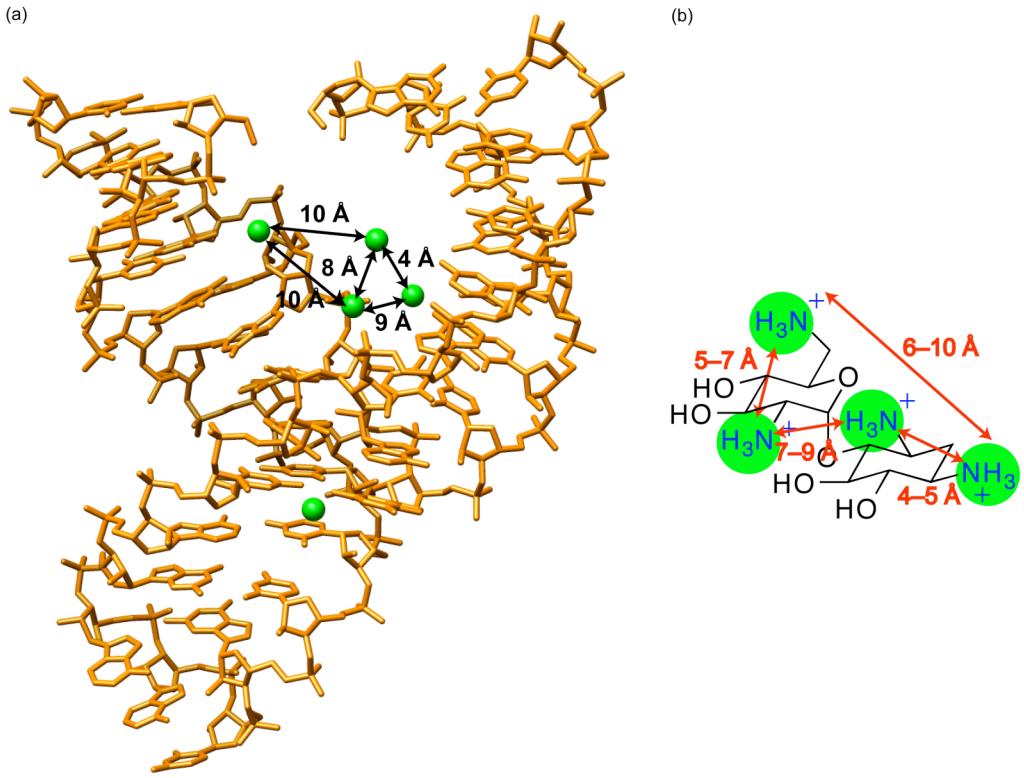


Figure 1.31. Comparison of distances between magnesium cations in the hammerhead ribozyme and ammonium functional groups in neamine. (a) X-ray crystal structure of the hammerhead ribozyme (orange) soaked with Mg^{+2} cations (green). The distances between magnesium ions are indicated. (b) Approximate distances between the ammonium groups (highlighted in green) of neamine. (Adapted from Tor, Y., Hermann, T., Westhoff, E. *Chem Biol*. **1998**, *5*, R277–R283).

from weak molecular interactions (ΔG_{mol}) includes van der Waals and hydrogen-bonding contacts which are mediated by the charged ammonium groups. The free energy of these interactions is the largest driving force for paromomycin binding to the A-site (–22.5 kcal/mol). Pilch's studies show over 86% of RNA–paromomycin contacts are characterized as weak intermolecular interactions. In contrast, Chaires and Wilson have analyzed a number of intercalators and groove binders to DNA, suggesting that hydrophobic effects (ΔG_{hyd}) play a larger role than weak molecular

interactions (ΔG_{mol}). The disparity compared to DNA is hypothesized to be caused by the highly polar character of aminoglycosides overall which exploit hydrogen-bonding and water mediated interactions versus DNA binding ligands which are typically more nonpolar and contain more neutral functional groups.

The role electrostatics plays in RNA recognition is two-fold. One, electrostatics enables ligands to form high-affinity interactions with RNA. Second, electrostatics allows for recognition of numerous "plastic" therapeutically-interesting RNA scaffolds by flexible, positively charged ligands. The ability to discriminate with a certain degree of selectivity demonstrates that spatial complementarity and electrostatics are essential in recognition.

1.13 CONCLUDING REMARKS

The therapeutic potential of targeting nucleic acids is limitless. Selectively modulating gene expression in vivo with small molecules is a distinct possibility in the near future, while the development of new generations of antibiotics has been effective against new resistant bacterial strains. Although selective targeting of DNA has progressed further than RNA, both nucleic acids share similar factors that must be taken into account for the design of high affinity ligands. Electrostatics, hydration, hydrophobic effects and weak intermolecular forces contribute greatly to the affinity of any particular binder. Efforts to computationally model and quantify these factors have mostly correlated with experimental results. One challenge still facing ligand design for DNA-binding is the ability to selectively and specifically target moderately

long nucleic acid sequences (over 14 base pairs).[136] Targeting RNA faces greater challenges due to its variable secondary structure that limits the broad application of recognition principles.

Despite these challenges, in 2000, Pfizer introduced the first fully synthetic antibiotic in over 50 years. *Linezolid*™, a last resort antibiotic, part of the new oxazolidinone class,[137] demonstrates the virtue of structure-based design and success of constructing molecular scaffolds drastically different from natural RNA binders. Future nucleic acid binders will require an immense amount of SAR studies to understand the localized effects caused by different ligand functional groups on electrostatic and hydrophobic interactions. Rigorous biophysical experiments will be necessary to fully characterize the target structure, its effects on the surrounding environment, and the thermodynamics of ligand binding. In addition to experimental investigations, molecular dynamics simulations will be an important component of drug design.

The beginning of molecular biology started over 50 years ago, and science has made large advances in understanding the subcellular underpinnings of life. Building on the discoveries of pioneering scientists, we now understand the importance and role of nucleic acids in dictating all aspects of life. The future of targeting nucleic acids is divided into three disciplines: a) an antigene strategy targeting specific DNA sequences with synthetic oligonucleotides; b) an RNAi (RNA interference) mechanism to suppress gene expression; and c) development of small molecules to affect nucleic acid activity.

To effectively silence a gene or treat any antimicrobial infection would be a success contributed to past advances in molecular biology. Achieving such a large goal requires small steps and systematic elucidation of principles. To this end, our efforts have been focused on understanding the interaction between small organic molecules with nucleic acids. The distinct nucleic acid structures that form triple helices, the parokaryotic A-site, and the HIV-1 DIS allow for selective interaction and targeting by small effector ligands to possibly solve current biomedical problems.

1.14 REFERENCES

1. Temin, H. M., The DNA provirus hypothesis: the establishment and implications of RNA-directed DNA synthesis. *Science* **1976**, *192*, 1075–1080.

2. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., Cech, T. R., Self-splicing RNA: autoexcision and autocyclication of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **1982**, *31*, 147–157.

3. Cech, T. R., Zaug, A. J., Grabowski, P. J., In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981**, *27*, 487–496.

4. Stark, B. C., Kole, R., Bowman, E. J., Altman, S., Ribonuclease P: an enzyme with an essential RNA component. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 3717–3721.

5. Chow, C. S.; Bogdan, F. M., A structural basis for RNA-ligand interactions. *Chem. Rev.* **1997**, *97*, 1489–1514.

6. Watson, J. D., Crick, F. H. C., Molecular structure of nucleic acids. *Nature* **1953**, *171*, 737–738.

7. Wilkins, M. H. F., Stokes, A. R., Wilson, H. R., Molecular structure of deoxypentose nucleic acids. *Nature* **1953**, *171*, 738–740.

8. Saenger, W., *Principles of nucleic acid structure*. Springer-Verlag: New York, 1984; p 556.

9. Wyatt, J. R., Tinoco, I., RNA structural elements and RNA function. In *The RNA world*, Gesteland, R. F., Atkins, J. F., Ed. Cold Spring Harbor Laboratory Press: Plainview, NY, 1993; pp 465–496.

10. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Maruissee, M., Merrill, S. H., Penswick, H. R., Zamir, A., Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462–1465.

11. Hermann, T., Strategies for the design of drugs targeting RNA and RNA-Protein complexes. *Angew. Chem. Int. Ed.* **2000**, *39*, 1890–1905.

12. Sucheck, S. J.; Wong, C. H., RNA as a target for small molecules. *Curr. Opin. Chem. Biol.* **2000**, *4*, 678–686.

13. Foloppe, N.; Matassova, N.; Aboul-Ela, F., Towards the discovery of drug-like RNA ligands? *Drug Discov. Today* **2006**, *11*, 1019–1027.

14. Hermann, T., Westhof, E., Rational drug design and high-throughput techniques for RNA targets. *Comb. Chem. High Through. Screen.* **2000**, *3*, 219–234.

15. Wilson, W. D., Ke, L., Targeting RNA with small molecules. *Curr. Med. Chem.* **2000**, *7*, 73–98.

16. Knowles, D., Foloppe, N., Matassova, N., Murchie, A., The bacterial ribosome, a promising focus for strcture-based drug design. *Curr. Opin. Pharm.* **2002**, *2*, 501–506.

17. Pearson, N. D., Prescott, C. D., RNA as a drug target. *Chem. Biol.* **1997**, *4*, 409–414.

18. Tor, Y., Targeting RNA with small molecules. *ChemBioChem* **2003**, *4*, 998–1007.

19. Hermann, T., Tor, Y., RNA as a target for small-molecule therapeutics. *Exper. Opin. Therap. Patents* **2005**, *15*, 49–62.

20. Vicens, Q., Westhof, E., RNA as a drug target: the case of aminoglycosides. *ChemBioChem* **2003**, *4*, 1018–1023.

21. Hermann, T., Drugs targeting the ribosome. *Curr. Opin Struct. Biol.* **2005**, *15*, 355–366.

22. Kirk, S., Luedtke, N. W., Tor, Y., Neomycin–acridine conjugate: a potent inhibitor of Rev–RRE binding. *J. Am. Chem. Soc.* **2000**, *122*, 980–981.

23. Zapp, M. L., Stern, S., Green, M. R., Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **1993**, *74*, 969–978.

24. Faber, C., Sticht, H., Schweimer, K., Rösch, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **2000**, *275*, 20660–20666.

25. Pollard, V. W., Malim, M. H., The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **1998**, *52*, 491–532.

26. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., Williamson, J. R., α Helix–RNA major groove recognition in an HIV-1 Rev peptide–RRE RNA complex. *Science* **1996**, *273*, 1547–1551.

27. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., Crothers, D. M., Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **1990**, *249*, 1281–1285.

28. Weeks, K. M., Crothers, D. M., RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell* **1991**, *66*, 577–588.

29. Ennifar, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C.; Dumas, P.; Walter, P., HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. *J. Biol. Chem.* **2003**, *278*, 2723–2730.

30. Davies, J., Davis, B. D., Misreading of ribonucleic acid code words induced by aminoglycoside antibiotics. *J. Biol. Chem.* **1968**, *243*, 3312–3316.

31. Carter, A. P., Clemons, W. M., Broderson, D. E., Morgan-Warren, R. J., Wimberly, B. T., Ramakrishnan, V., Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340–348.

32. Yonath, A., Structural insight into functional aspects of ribosomal RNA targeting. *ChemBioChem* **2003**, *4*, 1008–1017.

33. Dervan, P. B., Design of sequence-specific DNA-binding molecules. *Science* **1986**, *232*, 464–471.

34. Dervan, P. B., Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.

35. Strekowski, L., Wilson, B., Noncovalent interactions with DNA: an overview. *Mutat. Res.: Fundam. mol. Mech. Mutagen.* **2007**, *in press*.

36. Neidle, S., *Nucleic acid structure and recognition*. Oxford University Press: New York, 2002; p 188.

37. Blackburn, G. M.; Gait, M. J.; Loakes, D.; Williams, D. M., *Nucleic acids in chemistry and biology*. 3rd ed.; Oxford University Press: Oxford, England ; New York, 2006; p 470.

38. Lerman, L. S., Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **1961**, *3*, 18–30.

39. Hunter, C. A., Lawson, K. R., Perkins, J., Urch, C. J., Aromatic interactions. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651–669.

40. Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M., Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* **1999**, *6*, 1055–1061.

41. Larvey, R., Pullman, B., The molecular electrostatic potential of the B-DNA helix. *Theoret. Chim. Acta.* **1979**, *53*, 175–181.

42. Larvey, R., Pullman, B., The electrostatic field of DNA: the role of the nucleic acid conformation. *Nuc. Acids Res.* **1982**, *10*, 4383–4395.

43. Perahia, D., Pullman, A, The molecular electrostatic potential of the B-DNA helix. *Theoret. Chim. Acta. (Berl.)* **1979**, *50*, 351–354.

44. Cornell, W. D., Cieplank, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

45. Manning, G. S., Counterion binding in polyelectrolyte theory. *Acc. Chem. Res.* **1979**, *12*, 443–449.

46. Record, M. T., Anderson, C. F., Lohman, T. M., 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. *Quart. Rev. Biophs.* **1978**, *11*, 103–178.

47. Manning, G. S., Comments on selected aspects of nucleic acid electrostatics. *Biopolymers* **2003**, *69*, 137–143.

48. Alden, C. J.; Kim, S. H., Solvent-accessible surfaces of nucleic acids. *J. Mol. Biol.* **1979**, *132*, 411–434.

49. Chaires, J. B., Energetics of drug-DNA interactions. *Biopolymers* **1997**, *44*, 201–215.

50. Kopa, M. L., Fratini, A. V., Drew, H. R., Dickerson, R. E., Ordered Water Structure around a B-DNA Dodecamer. *J. Mol. Biol.* **1983**, *163*, 129-146.

51. Kubinec, M., Wemmer, D., NMR evidence for DNA bound water in solution. *J. Am. Chem. Soc.* **1992**, *114*, 8739–8740.

52. Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980**, *287*, 755–758.

53. Chalikian, T. V.; Sarvazyan, A. P.; Plum, G. E.; Breslauer, K. J., Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes and apparent molar adiabatic compressibilities of synthetic and natural DNA duplexes at 25 °C. *Biochemistry* **1994**, *33*, 2394–2401.

54. Chalikian, T. V., Sarvazyan, A. P., Breslauer, K. J., Hydration and partial compressibility of biological compounds. *Biophys. Chem.* **1994**, *51*, 89–109.

55. Shui, X., McFail-Isom, L., Hu, G., Williams, L. D., The B-DNA Dodecamer at High Resoltuion Reveals a Spine of Water on Sodium. *Biochemistry* **1998**, *37*, 8341-8355.

56. Drew, H. R.; Dickerson, R. E., Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **1981**, *151*, 535–556.

57. Kopka, M. L., Fratini, A. V., Drew, H. R., Dickerson, R. E., Ordered water structure around a B-DNA dodecamer. *J. Mol. Biol.* **1983**, *163*, 129–146.

58. Egli, M.; Portmann, S.; Usman, N., RNA hydration: a detailed look. *Biochemistry* **1996**, *35*, 8489–8494.

59. Kennard, O.; Cruse, W. B.; Nachman, J.; Prange, T.; Shakked, Z.; Rabinovich, D., Ordered water structure in an A-DNA octamer at 1.7 Å resolution. *J. Biomol. Struct. Dyn.* **1986**, *3*, 623–647.

60. Egli, M., Structural Aspects of Nucleic Acid Analogs and Antisense Oligonucleotides. *Angew Chem Int Ed Engl* **1996**, *35*, 1894-1909.

61. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R. E., The molecular origin of DNA–drug specificty in netropsin and distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 1376–1380.

62. Kiser, J. R., Monk, R. W., Smalls, R. L., Petty, J. T., Hydration changes in the association of Hoechst 33258 with DNA. *Biochemistry* **2005**, *44*, 16988–16997.

63. Vicens, Q., Westhof, E., Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **2001**, *9*, 647–658.

64. Pilch, D. S., Kaul, M., Barbieri, C. M., Kerrigan, J. E., Thermoydnamics of aminoglycoside-rRNA recognition. *Biopolymers* **2003**, *70*, 58–79.

65. Chaires, J. B., A thermodynamic signature for drug-DNA binding mode. *Arch. Biochem. Biophys.* **2006**, *453*, 26–31.

66. Ren, J., Jenkins, T. C., Chaires, J. B., Energetics of DNA intercalation reactions. *Biochemistry* **2000**, *39*, 8439–8447.

67. Peacocke, A. R., Skerrett, J. N. H., The interaction of aminoacridines with nucleic acid. *Trans. Farad. Soc.* **1956**, *52*, 261–279.

68. Lerman, L. S., The structure of the DNA–acridine complex. *Proc. Natl. Acad. Sci. U. S. A.* **1963**, *49*, 94–102.

69. Waring, M., Complex formation between ethidium bromide and nucleic acids. *J. Mol. Biol.* **1965**, *13*, 269–282.

70. LePecq, J. B., Paoletti, C., A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization. *J. Mol. Biol.* **1967**, *27*, 87–106.

71. Waring, M. J., Wakelin, L. P. G., Echinomycin: a bifunctional intercalating antibiotic. *Nature* **1974**, *252*, 653–657.

72. Le Pecq, J.-B., Le Bret, M., Barbet, J., Roques, B., DNA polyintercalating drugs: DNA binding of diacridine derivatives. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 2915–2919.

73. Leng, F., Priebe, W., Chaires, J. B., Ultratight DNA binding of a new bisintercalating anthracycline antibiotic. *Biochemistry* **1998**, *37*, 1743–1753.

74. Dervan, P. B., Becker, M. M., Molecular recognition of DNA by small molcules. Synthesis of bis(methidium)psermine, a DNA polyintercalating molecule. *J. Am. Chem. Soc.* **1978**, *100*, 1968–1970.

75. Becker, M. M., Dervan, P. B., Molecular-recognition of nucleic acid by small molecules. Binding affinity and structural specificty of bis(methidium)spermine. *J. Am. Chem. Soc.* **1979**, *101*, 3664–3666.

76. Fetcher, E. J., Olenyuk, B., Dervan, P. B., Design of a sequence-specific DNA bisintercalator. *Angew. Chem. Int. Ed.* **2004**, *43*, 3591–3594.

77. Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E., Dervan, P. B., Sequence selectivity of 3-hydroxypyrrole/pyrrole ring pairings in the DNA minor groove. *J. Am. Chem. Soc.* **1999**, *121*, 11621–11629.

78. Arcamone, F., Penco, S., Orezzi, Nicolella, V., Pirelli, A., Structure and synthesis of distamycin A. *Nature* **1964**, *203*, 1064–1065.

79. Coll, M., Frederick, C. A., Wang, A., Rich, A., A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer

d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 8385–8389.

80. Bostock-Smith, C. E., Laughton, C. A., Searle, M. S., Solution structure and dynamics of the A–T tract DNA decamer duplex d(GGTAATTACC)2: implications for recognition by minor groove binding drugs. *Biochem. J.* **1999**, *342*, 125–132.

81. Kopa, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., Dickerson, R. E., Defining GC-specificity in the minor grooove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* **1997**, *5*, 1033–1046.

82. Lown, J. W., DNA recognition by lexitropsins, minor groove binding agents. *J. Mol. Recognit.* **1994**, *7*, 79–88.

83. Pelton, J. F., Wemmer, D. E., Structural characterization of a 2:1 distamycin A·d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 5723–5727.

84. Mrkish, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., Dervan, P. B., Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7586–7590.

85. Felsenfeld, G., Davies, D. R., Rich, A., Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.

86. Rich, A., Formation of a two- and three- stranded helical molecules by polyinosinic acid and polyadenylic acid. *Nature* **1958**, *181*, 521–525.

87. Thuong, N. T., Helene, C., Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666–690.

88. Sarkar, P. S., Brahmachari, S. K., Intramolecular triplex potential sequence within a gene down regulates its expression in vivo. *Nucleic Acids Res.* **1992**, *20*.

89. Moser, H. E., Dervan, P. B., Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **1987**, *238*, 645–650.

90. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.-L., Thoung, N. T., Lhomme, J., Hélène, C., Sequence-specific recognition, photocrossllinking and cleavage of the DNA double helix by an oligo-[α]-

thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **1987**, *15*, 7749–7760.

91. Duval-Valentin, G., Thuong, N. T., Hélène, C., Specific inhibition of transcription by triple helix-forming oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504–508.

92. Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*. 3rd ed.; Worth Publishers: New York, 2000; p 1 v. (various pagings).

93. Escude, C., Garestier, T., Sun, J.-S., Drug interaction with triple-helical nucleic acids. *Methods Enzymol.* **2001**, *340*, 340–357.

94. Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Kayakawa, Y., Seto, H., Telomestatin, a novel telomerase inhibitor from *Stretomyces anulatus*. *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.

95. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., Hurley, L. H., Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593–11598.

96. Gellert, M., Lipsett, M. N., Davies, D. R., Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*.

97. Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. G., Bacchetti, S., Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **1992**, *11*, 1921–1929.

98. Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L. C., Coviello, G. M., Wright, W. E., Weinrich, S. L., Shay, J. W., Specific association of human telomerase activity with immortal cells and cancer. *Science* **1994**, *266*, 2011–2015.

99. Zahler, A. M., Williamson, J. R., Cech, T. R., Prescott, D. M., Inhibition of telomerase by G-quartet DNA structures. *Nature* **1991**, *350*, 718–720.

100. Sun, D., Thompson, B., Cathers, B. E., Salazar, M., Kerwin, S. M., Trent, J. O., Jenkins, T. C., Neidle, S., Hurley, L. H., Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **1997**, *40*, 2113–2116.

101. Baliant, G. A., Antiretroviral therapeutic possibilities for human immunodeficiency virus/acquired immunodeficiency syndrome. *Pharm. Therap.* **2001**, *89*, 17–27.

102. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., Hilgenfeld, R., Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767.

103. Enserink, M., West Nile drugs, vaccine still years away. *Science* **2000**, *290*, 1483.

104. Weinberg, M. S., Ely, A., Passman, M., Mufamadi, S. M., Arbuthnot, P., Effective anti-hepatitis B virus hammerhead ribozymes derived from multimeric precursors. *Oligonucleotides* **2007**, *17*, 104–112.

105. Hermann, T.; Patel, D. J., RNA bulges as architectural and recognition motifs. *Structure* **2000**, *8*, R47–54.

106. Fourmy, D., Recht, M., Blanchard, S., Puglisi, J., Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996**, *274*, 1367–1371.

107. Vicens, Q., Westhof, E., Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.* **2003**, *326*, 1175–1188.

108. Vicens, Q., Westhof, E., Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding A site. *Chem. Biol.* **2002**, *9*, 747–755.

109. Hermann, T., Westhof, E., Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure* **1998**, *6*, 1303–1314.

110. Batey, R. T.; Rambo, R. P.; Doudna, J. A., Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 2326–2343.

111. Weeks, K. M.; Crothers, D. M., Major groove accessibility of RNA. *Science* **1993**, *261*, 1574–1577.

112. Mujeeb, A.; Clever, J. L.; Billeci, T. M.; James, T. L.; Parslow, T. G., Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.* **1998**, *5*, 432–436.

113. Schatz, A., Bugie, E., Waksman, S. A., Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **1944**, *55*, 66–69.

114. Waksman, S. A., Lichevalier, H. A., Neomycin, a new antibiotic active against streptomycin resistant bacteria, including tuberculosis organisms. *Science* **1949**, *109*, 305–307.

115. Umezawa, H., Ueda, K., Maeda, K., Production and isolation of a new antibiotic, kanamycin. *J. Antibiot.* **1957**, *A10*, 187–189.

116. Weinstein, M. J., Leudemann, G. M., Oden, E. M., Gentamicin, a new broad spectrum antibiotic complex. *Antimicrob. Agents Chemother.* **1964**, *4*, 1–7.

117. Hutchin, A., Cortopassi, G., Proposed molecular and cellular mechanism for aminoglycoside ototoxicity. *Antimicrob. Agents Chemother.* **1994**, *38*, 2517–2520.

118. Moazed, D., Noller, H. F., Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **1987**, *327*, 389–394.

119. Purohit, P., Stern, S., Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. *Nature* **1994**, *370*, 659–662.

120. Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D., Puglisi, J. D., RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J. Mol. Biol.* **1996**, *262*, 421–436.

121. Han, Q.; Zhao, Q.; Fish, S.; Simonsen, K. B.; Vourloumis, D.; Froelich, J. M.; Wall, D.; Hermann, T., Molecular recognition by glycoside pseudo base pairs and triples in an apramycin-RNA complex. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 2694–2700.

122. Leulliot, N., Varani, G., Current topics in RNA–protein recognition: control of specificty and biological function through induted fit and conformational capture. *Biochemistry* **2001**, *40*, 7947–7956.

123. Hermann, T., Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie* **2002**, *84*, 869–875.

124. Hermann, T., Auffinder, P., Westhof, E., Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* **1998**, *27*, 153–165.

125. Ma, B., Kumar, S., Tsai, C.-J., Nussinov, R., Folding funnels and binding mechanisms. *Protein. Eng.* **1999**, *12*, 713–720.

126. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Tarry, M. J., Carter, A. P., Ramakrishnan, C. V., Recognition of cognate transfer RNA by the 30S subunit. *Science* **2001**, *292*, 897–902.

127. Walter, F., Putz, J., Giege, R., Westhof, E., Binding of tobramycin leads to conformational changes in yeast rTNA(Asp) and inhibition of aminoacylation. *EMBO J.* **2002**, *21*, 760–768.

128. Dewey, T. G., Turner, D. H., Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry* **1979**, *18*, 5757–5762.

129. Suurkuusk, J., Alvarez, J., Freire, E., Biltonen, R., Calorimetric determination of the heat capacity changes associated with the conformational transitions of polyriboadenylic acid and polyribouridylic acid. *Biopolymers* **1977**, *16*, 2641–2652.

130. Wang, H., Tor, Y., Electrostatic Interactions in RNA Aminoglycosides Binding. *J. Am. Chem. Soc.* **1997**, *119*, 8734–8735.

131. Wang, H., Tor, Y., RNA–Aminoglycoside Interactions: Design, Synthesis, and Binding of "Amino-aminoglycosides" to RNA. *Angew. Chem. Int. Ed.* **1998**, *37*, 109–111.

132. Scott, W. G., Murray, J.B., Arnold, J. R. P., Stoddard, B. L., Klug, A., Capturing the structure of a catalytic RNA interdieate: the hammerhead ribozyme. *Science* **1996**, *274*, 2065–2069.

133. Hermann, T., Westhof, E., Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J. Mol. Biol.* **1998**, *276*, 903–912.

134. Tor, Y., Hermann, T., Westhof, E., Deciphering RNA recognition: aminoglycoside binding to the hammerhead ribozyme. *Chem. Biol.* **1998**, *5*, R277–R283.

135. Hermann, T., Chemical and functional diversity of small molecule ligands for RNA. *Biopolymers* **2003**, *70*, 4–18.

136. Opalinska, J. B., Gewirtz, A. M. Nucleic-acid therapeutics: basic principles and recent applications. *Nat. Rev. Drug Disc.* **2002**, *1*, 503–514.

137. Barbachyn, M. R., Ford, C. W., Oxazolidinone structure–activity relationships leading to linezolid. *Angew. Chem. Int. Ed.* **2003**, *42*, 2010–2023.

Chapter 2

"Extended" Ethidium Bromide Analog as a Triple Helix Intercalator

---

2.1 THE IMPORTANCE OF TRIPLE HELICES

Higher ordered nucleic acid structures other than the canonical DNA double helix are also considered a target of small effector ligands. In recent years, triple helical nucleic acids have attracted attention because of their potential to modulate gene expression. Two discoveries in the 1980's resulted in renewed interest in this atypical nucleic acid. First, short oligonucleotides were demonstrated to sequence-specifically target double helical DNA to form short intermolecular triple helices.[1–3] These synthetic oligonucleotides both modify their DNA target and interfere with DNA-related processes such as transcription.[3,4] Second, a new polymorph of DNA, H-DNA, consisting of oligopurine-oligopyrimidine mirror repeat sequences that fold to form intramolecular triple helices, was discovered in vivo.[5] H-DNA is thought to play an important role in gene regulation since oligopurine-oligopyrimidine mirror repeat sequences are abundantly found near gene regulatory regions and are thought to make up over 1% of eukaryotic DNA.[6] Understanding the formation and structure of this interesting nucleic acid motif could potentially result in therapeutics that selectively modulate the expression of genes that code for a particular disease.

Independently, Dervan and Hélène were the first to demonstrate DNA duplexes could be sequence-specifically targeted and chemically modified with a third

nucleic acid strand. Dervan and colleagues designed homopyrimidine oligodeoxyribonucleotide sequences up to 15 bases in length conjugated to EDTA•Fe(II) to sequence-specifically bind and oxidatively cleave their target DNA.[1] Previous studies have shown oligonucleotides covalently attached to EDTA•Fe(II) to similarly bind and modify single-stranded DNA.[7] Upon forming an intermolecular triple helix, Dervan was able to visualize cleavage of the target duplex by footprinting studies, as well as determine the binding orientation of the third strand in the major groove as parallel or anti-parallel. At the same time, Hélène and colleagues targeted oligodeoxyadenylate sequences with an azidoproflavine-conjugated oligo-[α]-thymidylate strand.[2] The synthetic α-oligonucleotide prevents cleavage by nucleases making this oligonucleotide potentially viable in an in vivo environment. Azidoproflavine was used as a photocrosslinking fluorophore, which after treatment with base, cleaves strands in the target DNA double helix. Similar to Dervan, Hélène was able to determine the orientation of the third strand and location of chemical cleavage by footprinting studies.

Disruption of functions associated with the target DNA duplex does not necessarily need to occur through chemical modification. The formation of an intermolecular triple helix near a specific gene's promoter region could potentially inhibit sliding of RNA polymerase, therefore, inhibiting transcription. Additionally, the binding of a short oligonucleotide could interfere with protein binding that is necessary for normal cellular function. For H-DNA, its strategic location near many gene promoter regions suggests an important role in regulating gene expression.[8] Its

formation under superhelical stress conditions can potentially switch on and off translation of the nearby gene. Interestingly, the formation of H-DNA from mirror repeats of oligopyrimidine and oligopurines also forms a single stranded DNA region. Conceivably, this single-stranded DNA could be the target of an anti-sense targeting strategy but for a DNA substrate in contrast to mRNA.

Despite the therapeutic virtues of triple helical nucleic acids, its low stability limits its utility in medicinal applications. One strategy to stabilize triple helical structures is with low molecular weight ligands.[9] Binding of small molecules with intramolecular H-DNA structures located near specific gene sequences may assist in stabilizing its structure under superhelical stress conditions. This strategy may affect gene expression detrimentally and disrupt normal biological function. Intermolecular triplex-forming oligonucleotides (TFOs) can also be stabilized with small molecule ligands and affect transcription, recombination or protein binding. Despite the structural similarity between DNA triple helices to the DNA double helix, the triplet base pairing formed upon binding of the third strand has specific properties and characteristics that can be exploited for the design of selective ligands.

2.2 THE STRUCTURE OF TRIPLE HELICAL NUCLEIC ACIDS

Triple helices are formed when short oligonucleotides occupy the major groove of pre-existing double helices, resulting in a Hoogsteen or reverse Hoogsten hydrogen bonding pattern with the target duplex (Figure 2.1). Two types of base triplets can form, depending if the third base is a pyrimidine (PyPu×Py) or a purine

(PyPu×Pu). Hoogsteen base pairing orients the third strand in a parallel fashion to the duplex purine strand while reverse Hoogsteen base pairing forces an antiparallel orientation (Figure 2.1).

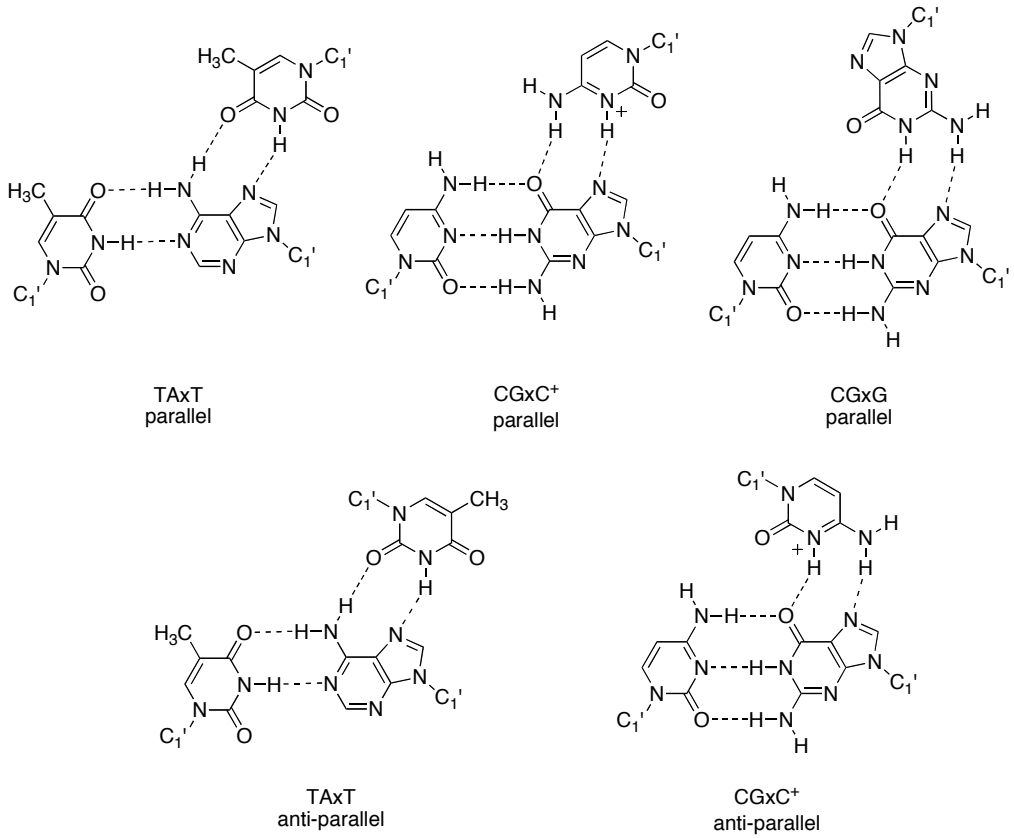


Figure 2.1. Examples of base triplet orientations observed in triple helices. *Top row*. Hoogsteen base pairing in pyrimidine•purine•pyrimidine triplets (TA×T and CG×C^+) as well as a pyrimidine•purine•purine triplet (CG×G) orients the third strand parallel to the purine strand in the Watson-Crick base pair. *Bottom row*. Reverse Hoogsteen base pairing for pyrimidine• purine•pyrimidine triplets orients the third strand anti-parallel to the purine strand in the Watson-Crick base pair.

One of the structural challenges to forming a triple helix is the electrostatic charge repulsion of the phosphate backbones. Favorable conditions for triple helix formation include a low pH or the presence of divalent cations at neutral conditions to

partially screen and neutralize the phosphate backbone. Even in the first reported observation of triple helical DNA by Felsenfeld in 1957 between poly r(A) and poly r(U),[10] the formation of triple helices was dependent on the concentration of divalent cations in solution. One particular base triplet, CG×C^+, requires a lower pH to protonate N3 of cytosine, which is necessary to hydrogen bond with O6 and N7 along the Hoogsteen face of guanine. Water has also been observed through NMR studies to play a key role in diminishing the electrostatic repulsion between the phosphate backbones of strands.[11]

Slight structural perturbations exist for triple helical DNA compared to the Watson-Crick double helix.[11] Interestingly, Watson-Crick base pairs are displaced approximately 2 Å towards the minor groove, most likely to accomodate for the approaching third strand along the major groove. Additionally, slight unwinding of the helix was observed from the normal B-form helical twist of 36° to ca. 31°.

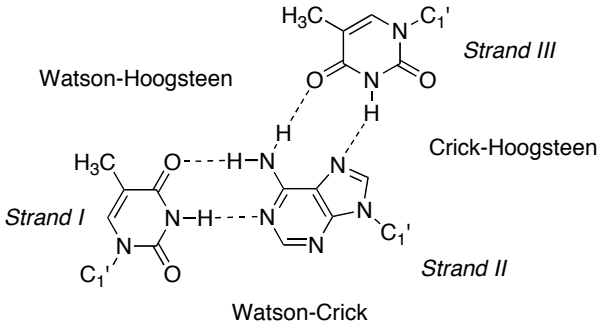


Figure 2.2. Nomenclature of grooves in a triple helix.

Complexation of the third strand results in the formation of three new grooves: the original duplex minor groove becomes the Watson-Crick groove (between strands

I and II), while the Crick-Hoogsteen (strands II and III) and Watson-Hoogsteen (strands I and III) grooves are created (Figure 2.2). The Watson-Hoogsteen groove is the widest (>7 Å) and has a moderate hydrophobic character due to the methyl groups from thymine that line the groove walls. At low temperatures, water molecules have been observed to exist for short periods of time.[12,13] The Watson-Crick groove (~ 6–7 Å) shows similar structural characteristics to the B-form minor groove but the "spine of hydration" normally observed with A•T base pair sequences was not clearly observed by NOE studies.[14] Finally, the Crick-Hoogsteen groove (~ 2–3 Å) is the narrowest and observed to support long-lived water molecules (1 > ns) that are believed to screen the phosphate backbone.

Despite screening of the negatively charged phosphate backbone by cations, the third strand is normally less stable than the Watson-Crick interactions in double helices. The type of base triplet formed may actually determine triplex stability. Calorimetric measurements have shown the formation of Hoogsteen base pairs with a smaller free energy change than Watson-Crick hydrogen bonding for Y×RY triplexes;[11] however, Additional calorimetric studies revealed the free energy change for R×RY triplex formation (–2.6 kcal/mol per base triple) is larger than for duplex formation (–1.3 kcal/mol per base pair).[15] The comparable change of free energies to Watson-Crick base pairs suggests R×RY triplexes are more stable than Y×RY triple helices. Since triple helix formation is highly dependent on pH, temperature, and ion valency and concentration, modulating stability of this fragile nucleic acid structure may have applications in gene suppression strategies.

As discussed before, the structure of nucleic acids play an important role in determining ligand binding. Groove widths and density of electrostatic potential in triple helices can direct ligands to specific intercalation sites and determine affinity. Hydration patterns suggest structural accessibility by small molecules and also areas of neutralized negative electrostatic charge. The larger helical diameter and polarity of triplet base pairs are also factors that have been considered in the design of triple helix selective ligands.

2.3 TRIPLE HELIX SELECTIVE BINDERS

Small molecule ligands that selectively target triple helical DNA over their duplex counterparts have been shown to stabilize its fine structure. The three grooves allow for interaction with base-pairing edges by groove binders such a netropsin and DAPI; however, a majority of triple helix selective ligands are polyaromatic intercalators specifically designed to exploit the larger, flat base-pairing surface of triplexes.

Benzopyridoindole derivatives. In 1992, the first triple helix-specific ligand, a benzo[*e*]pyridoindole derivative (BePI), was reported by Hélène and co-workers to significantly stabilize a 14-mer intermolecular triplex over the corresponding duplex in thermal melting studies (Figure 2.3).[16] Other benzopyridoindole derivatives were subsequently synthesized and successfully demonstrated to efficiently stabilize DNA triple helices including benzo[*g*]pyridoindole (BgPI),[17] benzo[*f*]quino[3,4-b]quinoxaline (BfPQ),[18] and pentacylic, expanded intercalators benzoindoloquinoline

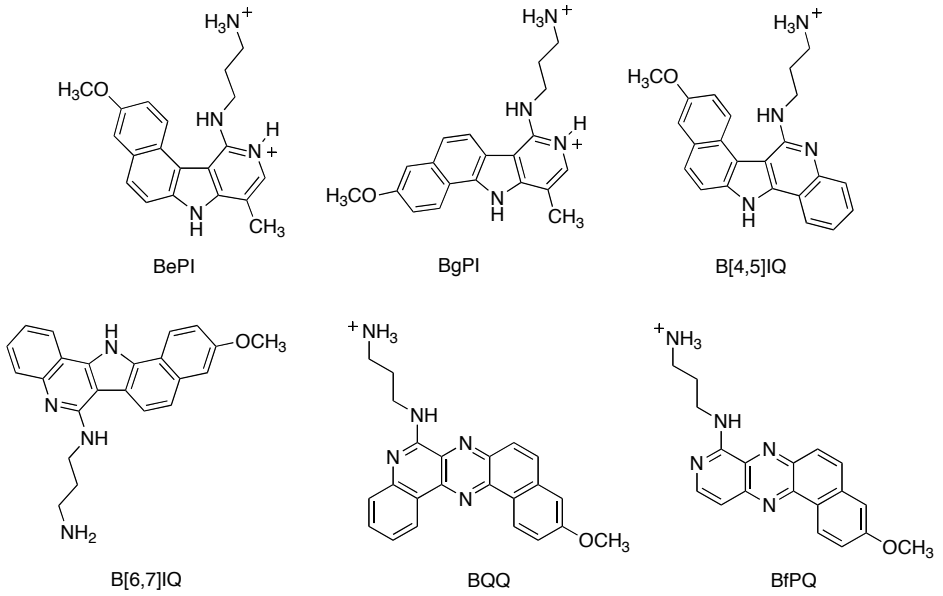


Figure 2.3. Polyaromatic benzopyridoindole derivatives.

derivatives 13*H*-benzo-[6,7]-indolo[3,2-*c*]qunoline (B[6,7]IQ) and 13*H*-benzo[4,5]-indolo[3,2-*c*]quinoline (B[4,5]IQ).[19] Intercalator scaffolds provide insights into ligand geometry. One example from modeling studies shows the structural isomeric framework of BePI and BgPI orients the aminoalkyl side chain into different grooves.[20] Expanded intercalators B[6,7]IQ and B[4,5]IQ were designed to exploit the larger surface provided by base triplets. An additional expanded intercalator with five fused aromatic rings (instead of four), benzo[f]quino[3,4-b]quinoxaline (BQQ), was compared against smaller quinoxaline derivatives BfPQ and BgPI for stabilizing an intermolecular triple helix formed by annealing a 14-mer oligonucleotide to a larger 26-bp double helix. Compared to the absence of ligand, BgPI and BfPQ only increased thermal stability of the triplex by 29 °C while BQQ increased triplex stability by 51 °C.[18] Another class of polyaromatic ligands with five fused rings, the

crescent-shaped dibenzophenantrholines, were determined by Hélène to stabilize triple helices better than earlier generation benzopyridoindole (BPI) ligands.[21] These family of molecules dramatically shows the importance of π–π sacking in stabilizing triple helices.

Naphthylquinoline derivatives. Another family of triple helix selective binders are naphythlquinolines, characterized by an unsubstituted 2-naphthalyl moiety bonded to the 4-substituted quinoline core (Figure 2.4). Modeling studies suggests the quinoline fragment intercalates with Watson-Crick A•T base pairs in a TA×T triple helix, while the naphthalyl moiety overlaps with thymine from the third strand.[22] To investigate this interaction, numerous derivatives were synthesized based on this scaffold. Chaires and Strekowski reported naphthylquinoline derivatives with alkyl amines of varying lengths bonded at the 4-position of quinoline (Figure 2.4a).[23] Through competitive dialysis assays, a majority of the compounds selectively bind poly d(T)•d(A)•d(T) sequences. Dimeric naphthylquinoline derivatives with alkyl amine, ether and amide linkers over 35-atoms in length have also been reported to have high affinity toward poly d(T)•d(A)•d(T) triplexes (Figure 2.4b,c).[24,25] A fluorescent sensor 4-aminonaphthalimide has also been attached to the naphthylquinoline core,[26] in addition to different heterocycles and larger polyaromatics to increase affinity and selectivity (Figure 2.4d).[22,27] Overall, many derivatives containing this scaffold have a high affinity for poly d(T)•d(A)•d(T) triple helices.

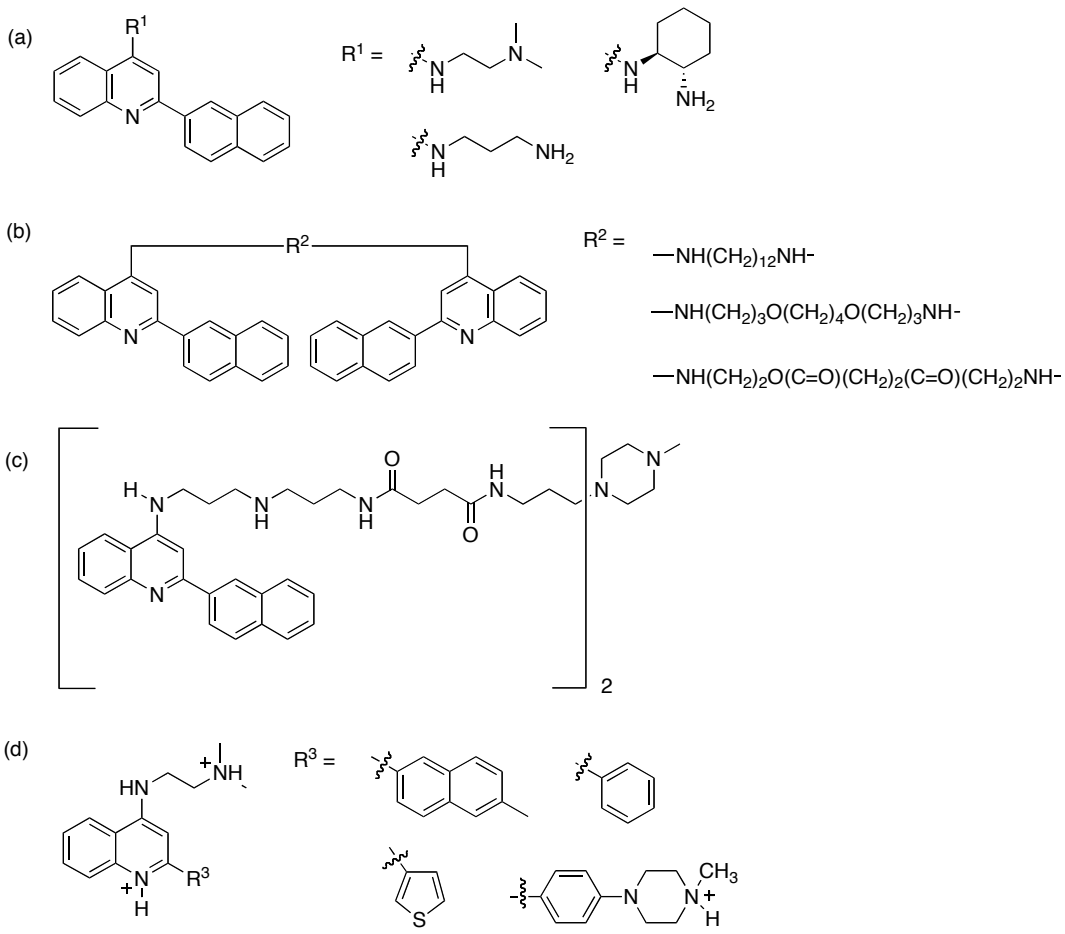


Figure 2.4. Examples of naphthylquinoline-based derivatives. (a) 4-quinolinyl substituted derivatives. (b) Dimeric naphthylquinoline derivatives with various aminoalkane and (c) amide linkers. (d) 2-quinolinyl substituted derivatives where the naphthalyl moiety is substituted with different aromatics and heterocycles.

Substituted amidoanthraquinone derivatives. Amidoanthraquinone derivatives illustrate the effects side chain placement and orientation have on triple helix affinity and selectivity (Figure 2.5). Several groups report numerous disubstituted amidoanthraquinones with high selectivity for triple helical DNA. Jenkins and co-workers report selectivity for a poly d(T)•d(A)•d(T) triplex by the 2,6-isomer over the 1,4-diaminoanthracenedione isomer (Figure 2.5a).[28] A series of isomeric

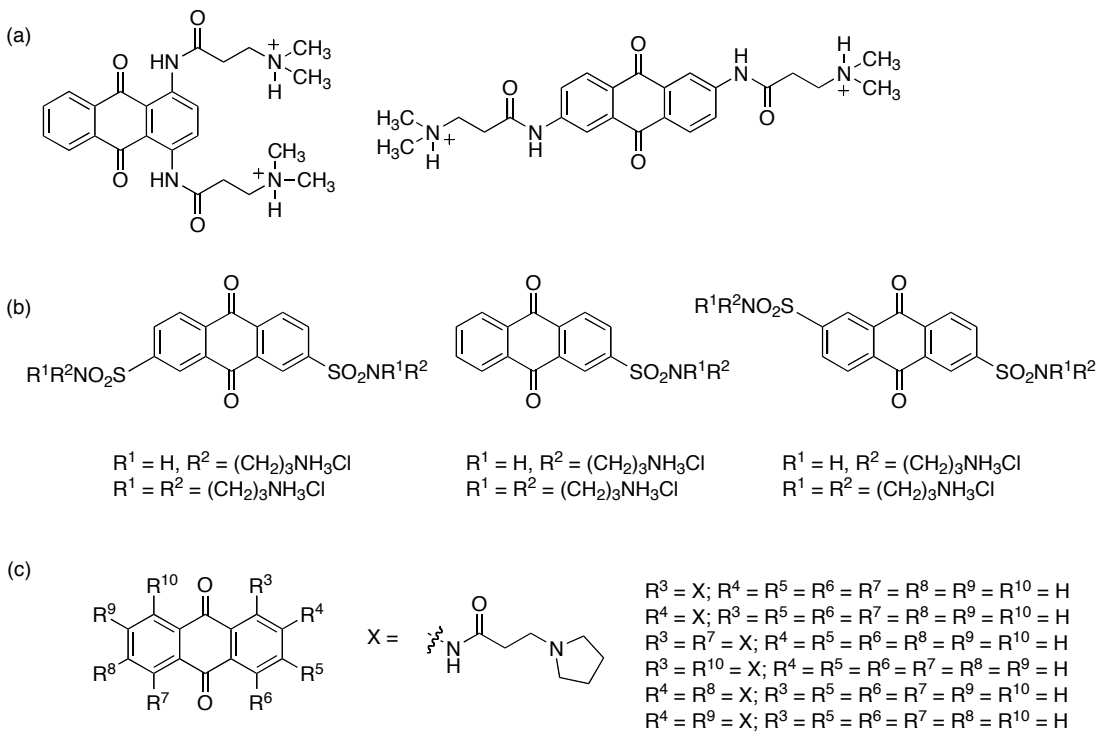


Figure 2.5. Examples of substituted amidoanthroquinones. (a) 1,4- and 2,7- disubstituted amidoanthroquinones. (c) Anthroquinones derivatives with sulfonamide side chains. (c) Additional disubstituted derivatives with alkyl and cyclic amine side chains.

disubstituted sulfonamidoanthraquinones were also demonstrated to selectively stabilize poly d(T)•d(A)•d(T) triplexes by UV-visible absorbance and thermal melting studies (Figure 2.5b). Viscometry was used to confirm intercalation as a binding mode.[29] Neidle and colleagues synthesized a series of mono- and disubstitued anthraquinone derivatives with an amidopropylpyrrazine side chain (Figure 2.5c). Substitution at the 2,7-position imparted the most stability to the triple helices studied.[30,31] Overall, significant stabilization of triple helices is observed when substitutions positioned at opposite ends of the anthraquinone scaffold are employed.

Other triple helix binders. In addition to designed triple helix selective ligands, dyes and antibiotics have been demonstrated to selectively bind triple helices

(Figure 2.6). The polyaromatic antitumor drug coralyne binds both triple helices CG×C$^+$ and TA×T nonspecifically, while another fused intercalator ethidium bromide (discussed below) has shown preference for only TA×T triplexes.[32] In addition to binding triple helices, the presence of certain ligands has been shown to *induce* the formation of hybrid DNA–RNA triple helices as is the case with well-studied duplex binders berenil, ethidium bromide, DAPI and netropsin.[33] The aminoglycoside antibiotic neomycin[34] and neomycin conjugated with polyaromatics pyrene[35] and BBQ[36] have also been reported to have high affinity and selectivity for triple helices. Interestingly, a tetraplex binding 3,3′-diethyloxadicarbocyanine (DODC) has higher selectivity for triplex DNA by competative dialysis studies.[37]

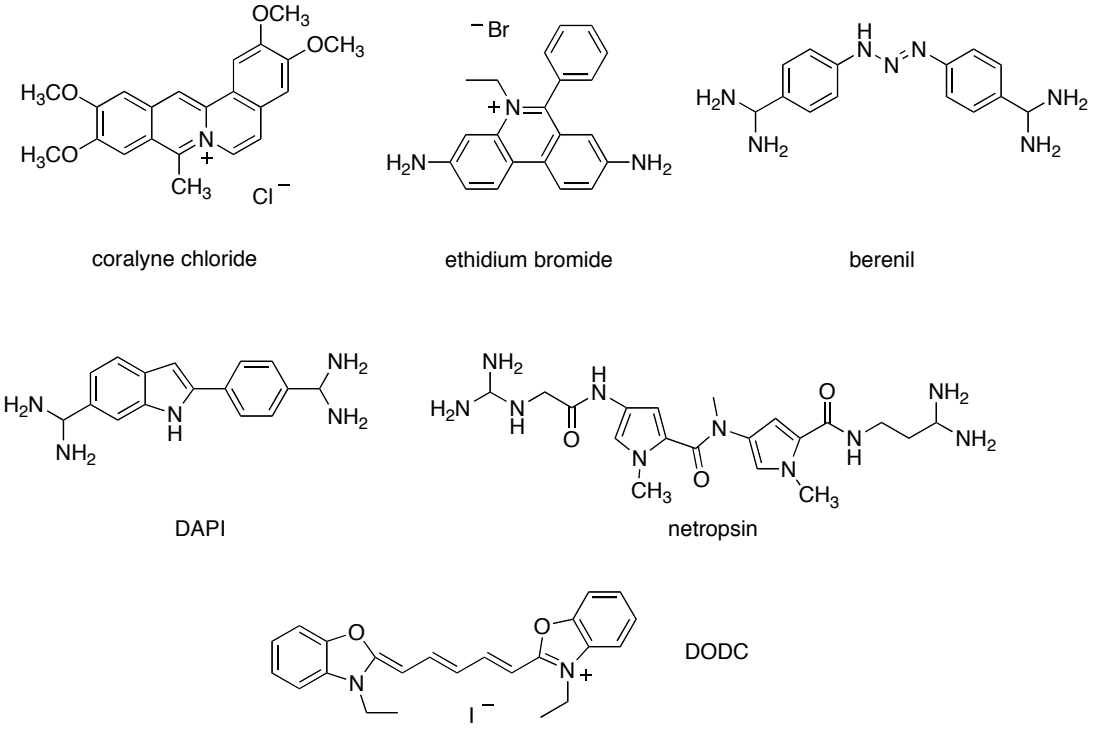


Figure 2.6. Triplex selective dyes and antibiotics. DODC = 3,3′-diethyloxadicarbocyanine.

Targeting triple helices with small molecular weight ligands is a relatively young yield in comparison to double helices; however, two characteristics of high affinity ligands can be observed: a) selectivity and affinity can be increased with large, aromatic surfaces, and b) selectivity can be modulated depending on the placement of side chains around the intercalating scaffold. These characteristics can potentially be applied to numerous duplex binding intercalators to increase their selectivity and affinity to triple helices.

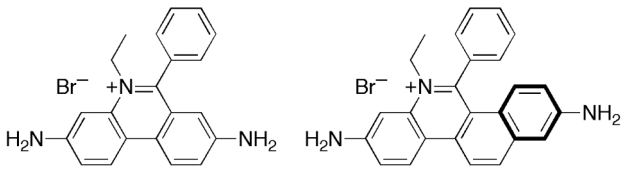
2.4 DESIGN OF AN EXTENDED ETHIDIUM BROMIDE SCAFFOLD

One polyaromatic ligand that binds DNA duplexes with high affinity is ethidium bromide (3,8-diamino-5-ethyl-6-phenylphenanthridinium bromide), and is synonymous with the study of nucleic acids. Based on a phenanthridine core, it has been used extensively as a stain while its trypanocidal effects have been traced to its intercalative properties in mammalian chromosomes[38] (trypanosomes are parasitic protozoan known to cause diseases such as African sleeping sickness and Chagas disease). The binding of ethidium bromide to DNA and RNA double helices has been extensively studied by various biophysical and spectroscopic techniques.[39–41] Since ethidium bromide has been established as a potent nucleic acid binder, several groups have synthesized derivatives to enhance its affinity to specific biologically important DNA and RNA structures. Additionally, the binding of this dye to triple helices has been thoroughly investigated,[42–44] with selectivity for TA×T over CG×C^+ triplexes, due to repulsion between the protonated cytosine and quarternary amine.[43]

Aforementioned, numerous aromatic scaffolds have been extended by fusing additional aromatic rings to increase selectivity and affinity for triple helical DNA. Our hypothesis is to "extend" the phenanthridine core of ethidium bromide, which has not been utilized before, as a core for triplex selective ligands (Figure 2.7).

In designing the "extended" ethidium analog, electronic properties of the parent heterocycle were considered.[45] Through resonance structures and electronic studies, it has been revealed that the exocyclic 3-amino functionality electronically "communicates" with the quaternary nitrogen. Since *ab initio* and DFT studies have shown electrostatic polarizability of intercalators to play a role in nucleic acid selectivity and affinity, this interaction was maintained.[46] Instead, the additional aromatic ring was fused on the opposing edge of ethidium bromide. This allows the exocyclic amine on the new aromatic ring to electronically communicate with the core scaffold in the same manner as the 8-amino functionality in ethidium bromide.

Synthetically, two key steps were employed to form the extended phenanthridinium framework: (a) assembly of the phenylnapthalene core by Suzuki-Miyaura cross-coupling of a functionalized naphthalene precursor and halobenzene building block carrying a benzamide unit, and (b) a Morgan-Walls ring-closure to assemble the bridging heterocyclic ring.

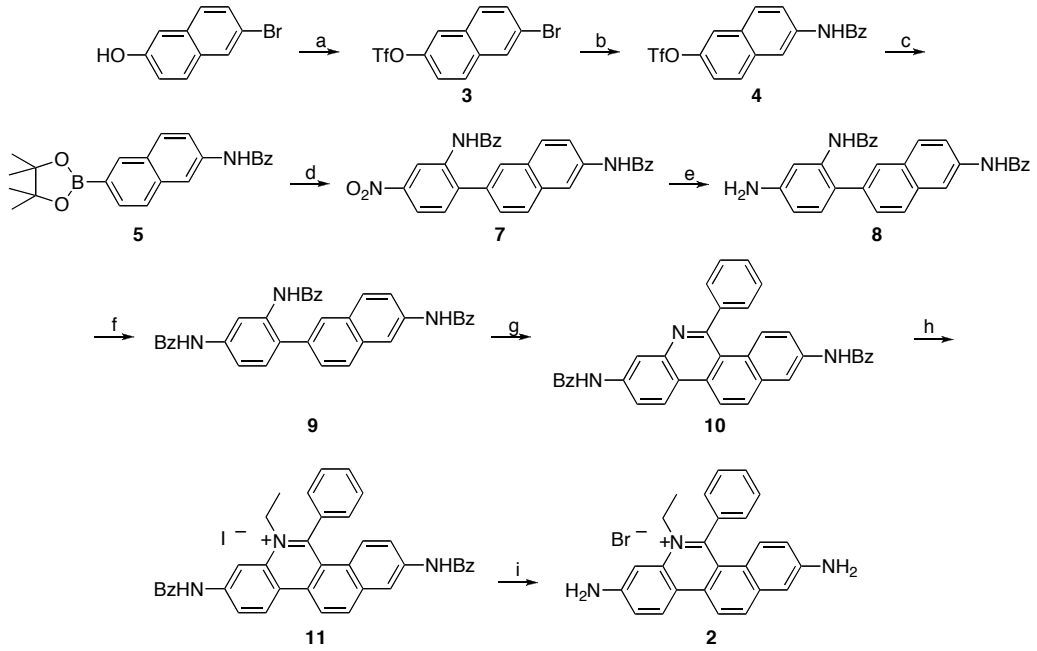


Ethidium Bromide **1** **2**

Figure 2.7. Ethidium bromide **1** and extended ethidium bromide **2**.

2.5 SYNTHESIS OF AN EXTENDED ETHIDIUM BROMIDE ANALOG

The synthesis of the extended ethidium analog **2** commenced with triflation of commercially available 6-bromo-2-naphthol to afford **3** in high yield (Scheme 2.1). Although cross-coupling can occur with triflates and aromatic halogens,[47] selective amide bond formation with benzamide occurred at the 6-bromo position **4**.[48] Benzamide is an excellent coupling partner as a masked nitrogen source for the 3-position exocyclic amine. Palladium-mediated borylation between the triflate **4** and 4,4,5,5-tetramethyl-[1,3,2]dioxaborolane afforded the boronic ester **5** in modest yield.[49,50] Completion of the aromatic scaffold was accomplished through a Miyaura-



Scheme 2.1. Synthesis of extended analogue **2**. *Reagents and conditions*: (a) Tf_2O, pyr., CH_2Cl_2, 0 °C to r.t., 3 h, 98%; (b) *N,N*'-dimethylethylenediamine, benzamide, CuI, K_2CO_3, toluene, 100 °C, 48 h, 96%; (c) 4,4,5,5-tetramethyl-[1,3,2]dioxaborolane, $Pd(dppf)Cl_2 \bullet CH_2Cl_2$, Et_3N, dioxane, 95 °C, 18 h, 65%; (d) **6**, $Pd(dppf)Cl_2 \bullet CH_2Cl_2$, 1M Na_2CO_3, DMF, 80 °C, 16 h, 90%; (e) 80% N_2H_4, Pd/C (10%), H_2 (1 atm), EtOH, 65 °C, 4 h; (f) BzCl, Et_3N, r.t., 12 h, 85% over 2 steps; (g) $POCl_3$, 100 °C, 18 h, 90%; (h) EtI, DMF, reflux, 20 h, 71%; (i) 48% HBr, reflux, 12 h, 80%.

Suzuki coupling of **5** and 2′-bromo-5′-nitrobenzanilide **6** to afford **7**.[51] Reduction of the aromatic nitro functionality was performed under standard catalytic hydrogenation conditions in the presence of hydrazine. To prevent any possible oxidation or degradation of the free aromatic amine, **8** was immediately reacted with benzoyl chloride under basic conditions to form the tribenzamide **9**. Treatment with phosphorous oxychloride to complete the extended phenanthridine core afforded **10** in high yield. Alkylation required reaction with iodoethane in refluxing DMF over 20 h to afford the quaternary amine **11** as a purple solid. Removal of the benzoyl groups in refluxing 48% HBr over 12 h yielded the desired extended ethidium analog **2**.

2.6 PHOTOPHYSICAL PROPERTIES OF EXTENDED ETHIDIUM ANALOG

The photophysical properties of the newly synthesized analog **2** were investigated to understand changes in the electronic structure, as well as its potential as a probe for monitoring interactions with nucleic acids. Similar to ethidium bromide **1**, the extended analog **2** displays intense absorption in the UV range and a weaker charge transfer band around 500 nm (Fig. 2.8a). Solvent polarity influences both molecules in a similar fashion. In methanol, acetonitrile and dichloromethane, the charge transfer band is observed at or above 500 nm. A dramatic hypsochormic shift is observed in water with a decrease in wavelength of approximately 20 nm for both **1** and **2** (Figure 2.8a). Excitation of the major absorption bands leads to emission of both molecules at approximately 600 nm. As with the ground state absorption bands,

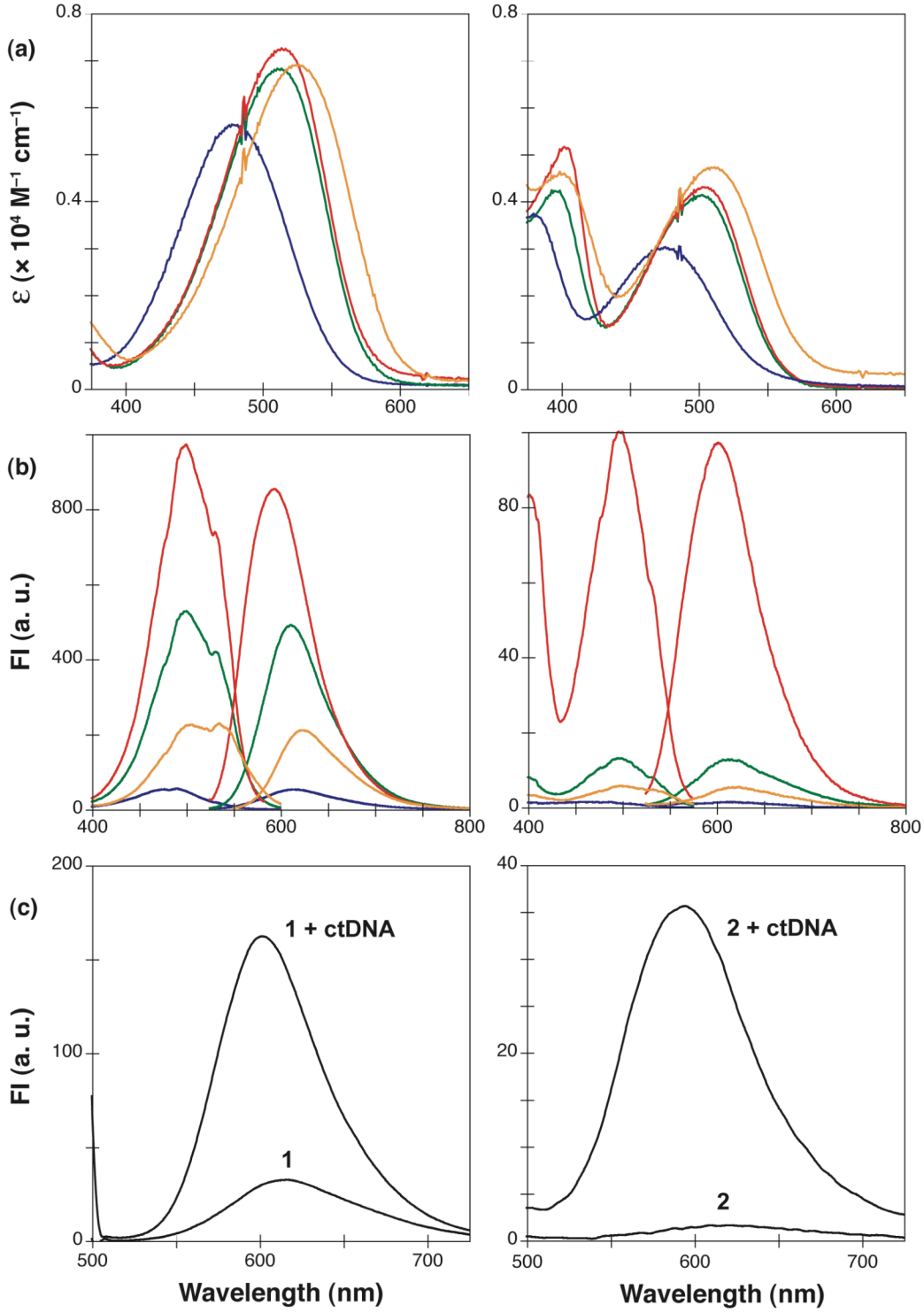


Figure 2.8 (a) UV-Vis absorbance spectra of **1** (left) and **2** (right). (b) Excitation and emission spectra of **1** and **2.** Solvents: H_2O (blue), MeOH (yellow), CH_3CN (green), CH_2Cl_2 (red). (c) **1** and **2** in the presence and absence of ctDNA in pH 7.4 buffer.

the relatively broad fluorescence emission band of both chromophores is sensitive to its microenvironment. Increasing solvent polarity results in a modest red shift of the emission maxima (Figure 2.8b). These observations are consistent with stabilization of the charged ground state upon increasing solvent polarity and concomitant modest destabilization of the more delocalized excited state.

Both **1** and **2** are less emissive in water when compared to their emission in more nonpolar organic solvents. While the extended analog **2** is inherently less emissive than the parent heterocycle **1**, it is almost completely quenched in water. As a result, addition of nucleic acids to an aqueous solution of the intercalators results in significant enhancement of emission, which is accompanied by a slight hypsochromic shift (Figure 2.8c). Clearly shown in Figure 2.8c, a greater enhancement of emission is observed upon nucleic acids binding for **2**, when compared to the emission intensities of free and bound ethidium bromide **1** (Figure 2.8c).

2.7 BINDING TO DUPLEX AND TRIPLEX NUCLEIC ACIDS

Binding studies with diverse nucleic acids were performed using steady-state fluorescence spectroscopy. To compare the nucleic acid affinity of **1** and **2** towards random DNA sequences, calf thymus DNA (ctDNA), which normally ranges from 200 – 6,000 base pairs, was employed. Monitoring the fluorescence emission from both ligands while titrating ctDNA, yielded dissociation constants of 15.1 ± 4.1 μM and 11.2 ± 1.6 μM for **1** and **2**, respectively. Within experimental error, no discrimination between the two heterocycles was observed (Figure 2.9, data summarized in Table

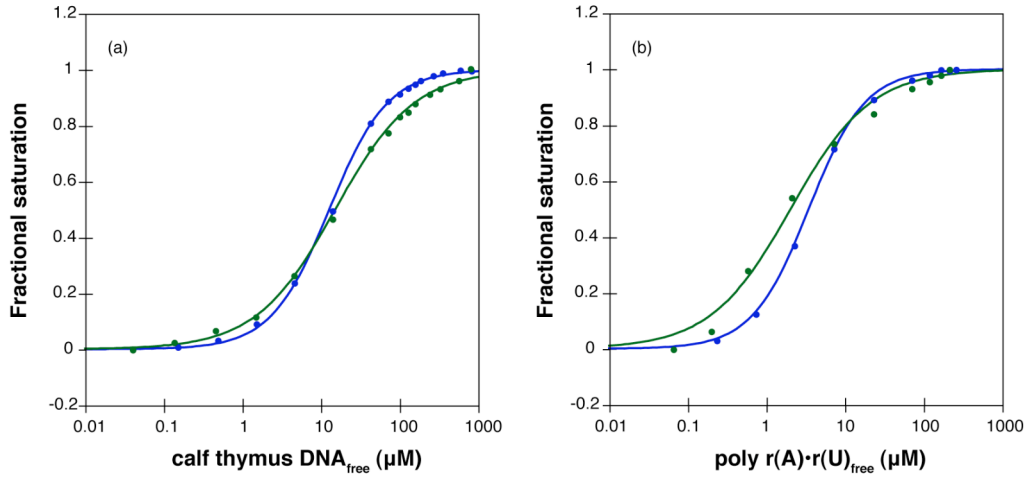


Figure 2.10. Nucleic acid binding isotherms of **1** and **2** with (a) ctDNA and (b) poly r(A)•r(U). The binding isotherms of **1** and **2** are depicted in blue and green, respectively.

2.1). Affinities for A-form RNA were evaluated using the synthetic polymeric duplex poly r(A)•poly r(U). The extended ethidium analog **2** exhibited slightly higher affinity (1.6 ± 0.2 μM) for A-form RNA compared to ethidium **1** (5.0 ± 0.9 μM) (Figure 2.9). Since A-form RNA has a larger helical diameter compared to B-form DNA (approximately 26 to 20 Å, respectively), it provides a larger intercalating surface for the extended ligand, which may explain the enhanced affinity compared to ethidium. To further explore this hypothesis, the ability of **2** to bind triple helical oligonucleotides was investigated. Additional fluorescence binding affinity studies revealed that **2** has over a 9-fold higher affinity for the DNA polymeric triple helix poly d(T)•d(A)•d(T) than **1** (1.0 ± 0.3 μM and 9.5 ± 5.2 μM, respectively) (Figure 2.10). This is likely due to more favourable stacking of the extended analog **2** with the larger intercalating surface provided by the base triplets,[52] as has been observed with other intercalating polyaromatic molecules. Studies with an

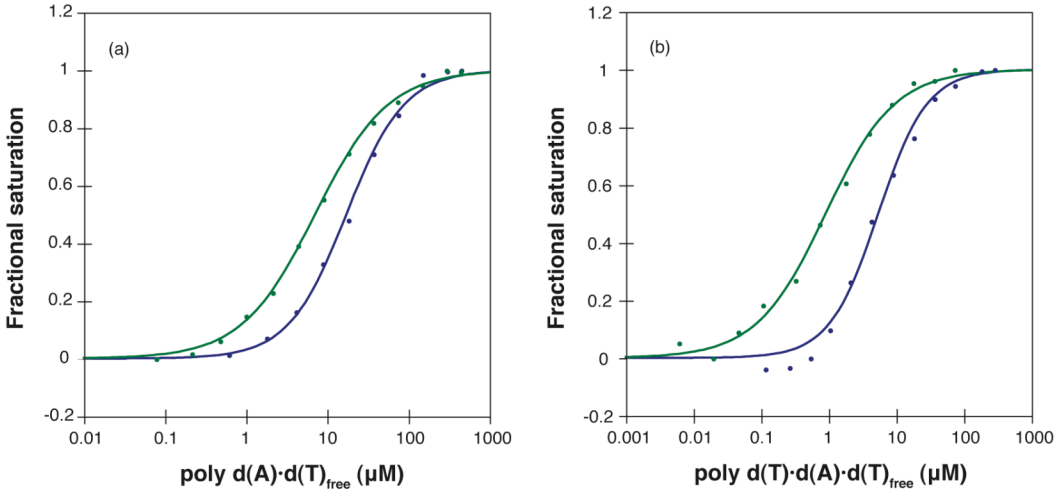


Figure 2.10. Nucleic acid binding isotherms of **1** and **2** with (a) duplex poly d(A)•d(T) and (b) triple helix poly d(T)•d(A)•d(T). The binding isotherms of **1** and **2** are depicted in blue and green, respectively.

RNA triple helix, possessing an even larger helical diameter than triplex DNA, were also employed (Figure 2.11). The affinities of **1** and **2** to the RNA triple helix poly r(U)•r(A)•r(U) were determined (16.1 ± 2.6 μM and 1.7 ± 0.3 μM, respectively), but **2** showed a slightly lower affinity for a RNA triplex than DNA triplex. A thorough search of the literature revealed our findings correlate well with previous studies reporting the destabilizing effect of **1** on triplex formation.[53] Although an equilibrium mixture of duplex and triplex RNA may exist during the fluorescence titration, Lehrman and Crothers reported that in a 1:2 stoichiometric ratio of poly r(A) and poly r(U), respectively, this mixture binds **1** with a lower affinity than an RNA duplex of 1:1 poly r(A) and poly r(U).[39] Waring and Hélène also report similar observations concerning the destabilizing effect ethidium bromide **1** has on polymeric RNA triple helices. A correlation between structural stability and ligand affinity was investigated next.[42,44]

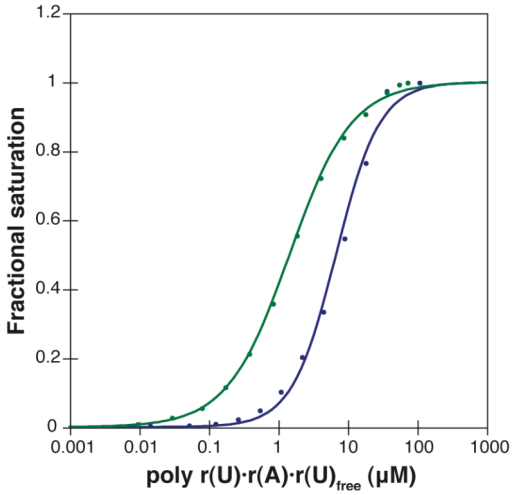


Figure 2.11 Binding isotherm of **1** (blue) and **2** (green) with triple helix poly r(U)•r(A)•r(U).

Table 2.1. Nucleic Acid Affinities of **1** and **2**.

Nucleic Acid Polymers	1 (μM)	2 (μM)
calf thymus DNA	15.1 ± 4.1	11.2 ± 1.6
poly r(A)•r(U)	5.0 ± 0.9	1.6 ± 0.2
poly d(T)•d(A)	30.8 ± 10.8	7.9 ± 1.3
poly d(T)•d(A)•d(T)	9.5 ± 5.2	1.0 ± 0.3
poly r(U)•r(A)•r(U)	16.1 ± 2.6	1.7 ± 0.3

2.8 THERMAL MELTING EXPERIMENTS

Thermal melting experiments were employed to measure the effects of ligands **1** and **2** on duplex and triplex stability (Figure 2.12). Instead of homopolymers used in fluorescence emission affinity studies, 19-nucleotide long homo-oligomers were used in to monitor a set absorbance at 260 nm during melting. A minimal stabilization was observed for the dT_{19}•dA_{19} duplex with both intercalators, increasing the helix melting point by approximately 4 °C degrees (Table 2.2) above that of the 19-mer duplex in the absence of ligand. Dramatic stabilization of the dT_{19}•dA_{19}•T_{19}

triple helix was observed with the extended ethidium analog **2** (Figure 2.12. and Table 2.2). Under low salt conditions, the triple helical dT_{19}•dA_{19}•dT_{19} begins to melt below 10 °C in the absence of ligand. While addition of ethidium bromide **1** slightly enhances the stability of the triple stranded oligonucleotide (T_m = 24 °C), a dramatic stabilization is apparent with the extended analog **2** (T_m = 37 °C). These results correlate well with affinities determined by the fluorescence-based titrations with polymeric oligonucleotides (Table 2.1).

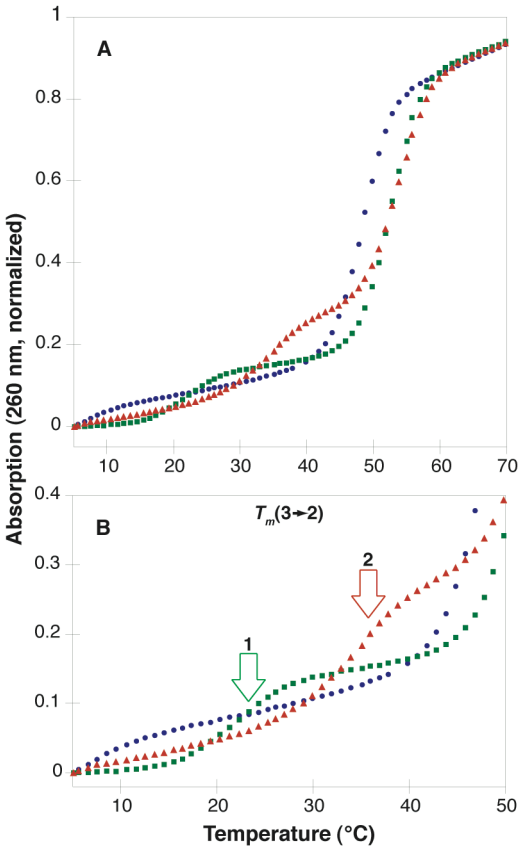


Figure 2.12 (A) Thermal melting profiles for dT_{19}•dA_{19}•dT_{19} with no ligand (●), ethidium bromide (■), and the extended ethidium analog (▲). (B) is an expansion of the lower temperatures in (A) where triple helix melting is observed.

Table 2.2. Thermal melting points of a DNA 19mer duplex and triple helix.

Ligand	Thermal Melting Points			
	$dA_{19}\cdot dT_{19}$		$dT_{19}\cdot dA_{19}\cdot dT_{19}$	
	$T_m(2\rightarrow1)$	$T_m(3\rightarrow2)$	$T_m(2\rightarrow1)$	$T_m(3\rightarrow2)$
no ligand	48 ± 0.7 °C	---	49 ± 0.1 °C	< 10 °C
1	52 ± 1.0 °C	---	53 ± 0.6 °C	24 ± 0.1 °C
2	53 ± 1.0 °C	---	54 ± 0.6 °C	37 ± 1.0 °C

Thermal melting studies also employed the short oligonucleotide RNA triple helix $rU_{19}\bullet rA_{19}\bullet rU_{19}$ (Figure 2.13) In the absence of ligand, the triple helix beings to melt at a temperature slightly higher than its DNA counterpart (T_m = 13 °C). In the presence of **2**, no stabilization of the triple helix is observed (Table 2.3), while melting of the triple helix is not observed in the presence of **1**. This illustrates the destabilizing effect of **1** on the RNA triple helix as also observed in fluorescence studies.

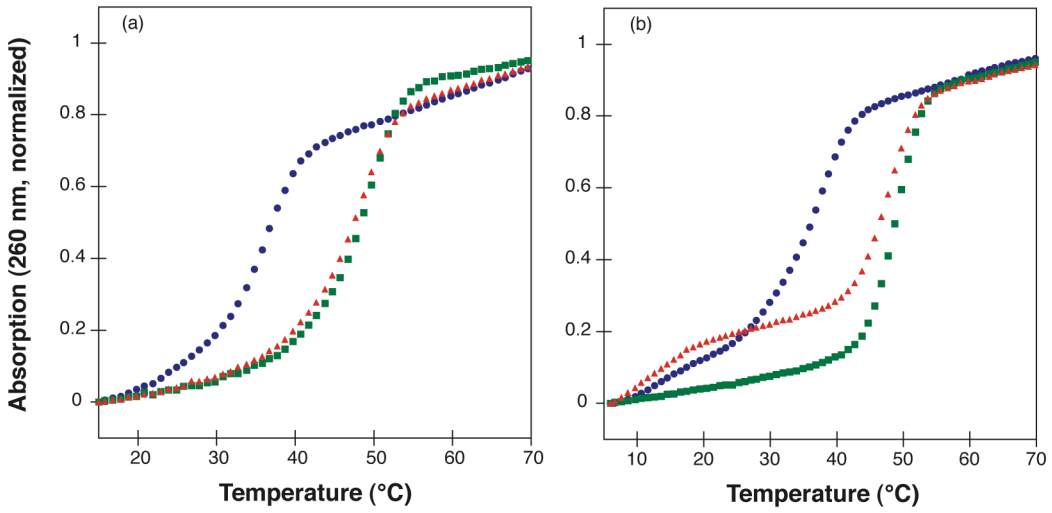


Figure 2.13. Thermal melting profiles for (a) $rA_{19}\cdot rU_{19}$ and (b) $rU_{19}\cdot rA_{19}\cdot rU_{19}$ with no ligand (●), ethidium bromide (■), and the extended ethidium analog (▲). The $T_m(2\rightarrow1)$ transition for both duplex and triple helix studies were similar in the absence and presence of both ligands. The $T_m(3\rightarrow2)$ transition for ethidium bromide is not observed because of its destabilizing effect on the RNA triple helix

Table 2.3. Thermal melting points of a RNA 19mer triple helix rU_{19}•rA_{19}•rU_{19}.

Ligand	T_m(2→1)	T_m(3→2)
no ligand	37 ± 0.6 °C	13 ± 1.0 °C
1	49 ± 0.6 °C	none observed
2	49 ± 1.0 °C	14 ± 1.7 °C

2.9 CONCLUSIONS

The formation of triple helices, both through the binding of a third synthetic oliognucleotide or intramolecularly as in H-DNA, has great potential in modulating gene expression. Altering the stability of these topologically interesting nucleic acids could have potential biotechnological applications. Significant efforts in recent years have yielded a handful of new heterocycles with diverse triple-helical selectivity traits, none, however, were based on ethidium. Our observations suggest that extending ethidium by fusing an additional aromatic ring yields an analog that can be viewed as a new triplex-selective motif. It is thought that these structural modifications exploit the increased π–π interactions between triplex base pairs and planar aromatic systems.[52] Further modification of the ethidium core can potentially fine tune the nucleic acids affinity and selectivity of this extended heterocycle.

2.10 ACKNOWLEDGEMENT

Portions of Chapter 2 have been previously published in the article "Extended Ethidium Bromide Analogue as a Triple Helix Intercalator: Synthesis, Photophysical Properties and Nucleic Acids Binding," *Chem. Commun.* **2006**, 2684–2686 by V. Tam, Q. Liu and Y. Tor. The dissertation author was the primary investigator and author of this paper.

2.11 EXPERIMENTAL

General Considerations. All chemicals were obtained from commercial suppliers and used without further purification unless otherwise specified. Anhydrous solvents were used where noted and obtained using a two-column purification system (Glasscontour Systems, Irvine, CA). Analytical thin-layer chromatography was performed on pre-coated silica gel aluminum-backed plates (Kieselgel 60 F254, E. Merck & Co., Germany). Flash chromatography was performed using silica gel (230-400 mesh) from E.M. Science or Silicycle. NMR solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). All NMR spectra were recorded on a Varian Mercury 400MHz instrument with chemical shifts reported relative to residual deuterated solvent peaks. Chemical shifts (δ) are reported in parts per million (ppm); multiplicities are indicated by s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublet), or m (multiplet); coupling constants (*J*) are reported in Hertz. Mass spectra were recorded at the UCSD Mass Spectrometry Facility, utilizing either a LCQDECA (Finnigan) ESI with a quadrupole ion trap or a MAT900XL (ThermoFinnigan) FAB double focusing mass spectrometer.

UV-Visible Absorbance Studies. UV-Visible experiments were carried out at ambient temperature in a quartz micro cell with a path length of 1.0 cm (Hellma GmbH & Co KG, Müllheim, Germany) on a Hewlett Packard 8452A diode array spectrometer. Both ethidium bromide **1** and the extended analog **2** samples were measured at 1.0×10^{-5} M (1% MeOH) in the appropriate spectroscopic grade solvent.

Fluorescence Spectroscopy Studies. Steady state fluorescence experiments were carried out at ambient temperature (unless otherwise noted) in a micro fluorescence cell with a path length of 1.0 cm (Hellma GmH & Co KG, Mullenheim, Germany) on a Perkin Elmer LS 50B luminescence spectrometer. Both ethidium bromide **1** and extended analog **2** samples were measured at 5.0×10^{-6} M (0.5% MeOH) in the appropriate spectroscopic grade solvent (excitation slit = 7.0 nm, emission slit = 7.0 nm, scan speed = 300 nm/min, spectra averaged over three scans).

Solvent Polarity Measurements. The relationship between emission energy and microenvironment solvent polarity [$E_T(30)$] was determined for the extended analog **2**. All emission spectra were determined at 5.0×10^{-6} M with 0.5% MeOH in four different solvents (water, methanol, acetonitrile and dichloromethane). Emission spectra were converted from wavelength (nm) to wave numbers (cm^{-1}). $E_T(30)$ values for each solvent were determined using Reichardt's salt (Reichardt, C. *Chem Rev.* **1994**, *94*, 2319–2358).

Relationship between Emission Energy and $E_T(30)$ Values.

Solvent	$E_T(30)$	Emission Energy of Extended analog (cm^{-1})
Water	63.1	16,181
Methanol	55.0	16,051
Acetonitrile	48.1	16,340
Dichloromethane	44.0	16,611

Nucleic Acid Binding Fluorescence Studies. A sonicated solution of calf thymus (ctDNA) was purchased from Gibco BRL. Poly r(A)•r(U) was purchased from Sigma-Aldrich and poly d(A), poly d(T), poly r(A), and poly r(U) were

purchased from Amersham Bioscience. Stock solutions were made in 1 × TE and quantified according to the following extinction coefficients in 50 mM sodium phosphate (pH 7.5) buffer at the stated wavelengths:

Nucleic Acid polymer	Extinction coefficient
calf thymus DNA	13,100 cm^{-1} M^{-1} per base pair (260 nm)
poly r(A)•r(U)	14,280 cm^{-1} M^{-1} per base pair (260 nm)
poly d(A)	8,600 cm^{-1} M^{-1} per base pair (257 nm)
poly d(T)	8,700 cm^{-1} M^{-1} per base pair (265 nm)
poly r(A)	9,900 cm^{-1} M^{-1} per base pair (260 nm)
poly r(U)	9,300 cm^{-1} M^{-1} per base pair (260 nm)

Fluorescence Titration Buffer Conditions. All fluorescence titrations were conducted at ambient temperature in a "physiological" buffer containing 30 mM HEPES (pH 7.4), KCl (100 mM), Na_2HPO_4 (10 mM), guanidinium hydrochloride (20 mM), $MgCl_2$ (2 mM), NaCl (20 mM), EDTA (0.5 mM), and Nonidet P-40 (0.001%).

Fluorescence Titrations. In a Perkin Elmer LS-50B luminescence spectrometer, a dilute solution of ethidium bromide **1** or extended analog **2** (1.0 μM) in "physiological" buffer was excited at 479 nm or 389 nm, respectively. The emission of each compound was monitored near 605 nm (excitation slits = 10 nm, emission slits = 20 nm, scan speed = 300 nm/min, averaged over three scans). Small volumes of highly concentrated nucleic acids were titrated, with a small increase in the final volume (≤10%) of the sample. Data analysis accounts for the small change in volume. Titrations with duplex nucleic acid polymers were performed at ambient temperature. Titrations with polymeric triple helices were conducted at 0 °C with a thermally jacketed cuvette holder. All triple helical samples were maintained at 0 °C during these titrations.

Fluorescence Titration Data Analysis. The fractional change in emission intensity was taken as the fraction of ethidium **1** or ethidium analog **2** bound (χ) and was used to calculate the concentration of free nucleic acid [NA] and concentration of bound analog [$EtBr_{bound}$] at each concentration of nucleic acid [NA_{tot}] (total concentration). The calculated fractional saturation was fit to the following equation using Kaleidagraph to derive binding affinity (K_d) of ethidium **1** and extended analog **2** for each respective nucleic acid:

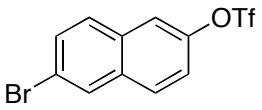
$$\chi_{bound} = \frac{[NA]^n}{[NA]^n + K_d}$$

where *n* is the Hill coefficient or degree of cooperativity associated with binding (in most cases n~1).

Thermal UV-Visible Denaturing Experiments. Thermal denaturing profiles were obtained using a Beckman Coulter DU® 640 Spectrophotometer with a high performance temperature controller and micro auto six holder. All hybridizations and UV melting experiments were carried out in 2.0×10^{-2} M PIPES (pH 7.0), 1.0×10^{-3} M EDTA, and 2.0×10^{-1} M NaCl. DNA oligonucleotides dA_{19} and dT_{19} were obtained from Integrated DNA Technologies (Coralville, IA) and purified using denaturing PAGE. Concentrations of oligonucleotide stocks were determined by using molar extinction coefficients 231,400 M^{-1} cm^{-1} for dA_{19} and 154,500 M^{-1} cm^{-1} for dT_{19} at 260 nm. RNA oligonucleotides rA_{19} and rU_{19} were obtained from Dharmacon RNA Technologies (Lafayette, CO) and purified using denaturing PAGE. Concentrations of oligonucleotide stocks were determined by using molar extinction coefficients 231,400 M^{-1} cm^{-1} for rA_{19} and 184,500 M^{-1} cm^{-1} for rU_{19} at 260 nm.

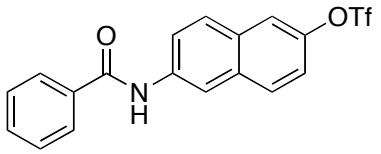
DNA double stranded (T·A) and triple-stranded (T·A·T) samples were prepared by mixing a 1:1 or 1:2 molar ratio of dA_{19} and dT_{19}, respectively. RNA triple-stranded (U·A·U) samples were prepared in a similar fashion with rA_{19} and rU_{19}. The concentration of ligand (10.0 μM) and total triple helix (1.0 μM) was maintained as a ratio of [ligand]:[triple helix] = 10.0. Samples were preformed by heating to 90 °C for 3 min then slowly cooled to room temperature over one hour. After reaching room temperature, the samples were further cooled in an ice bath for 20 min to reach temperatures favorable for triple helix formation. At the beginning of each thermal denaturing experiment, all samples were held at 5 °C for 30 min in order to stabilize preformed triple helixes. Samples were placed in a stoppered 1.0 cm path length cell and a background spectra (buffer) was subtracted from each sample. All melting curves were obtained upon increasing the temperature from 5 °C to 80 °C at a rate of 1.0 °C per min and absorbance measurements at 260 nm were performed every minute. Beckman Coulter software (provided with T_m Analysis Accessory for DU® Series 600 Spectrometers) determined the duplex melting temperatures by calculating the first derivative of the melting profile.

Synthetic Procedures & Characterization Data (^{1}H-NMR, ^{13}C-NMR, MS).



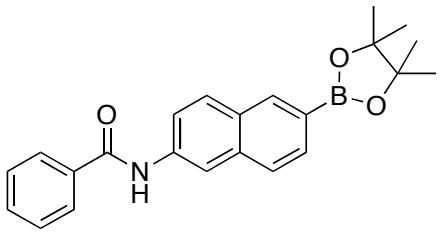
6-bromonaphthalen-2-yl trifluoromethanesulfonate (3): 6-bromo-2-naphthol (6 g, 27 mmol) was dissolved in anhydrous CH_2Cl_2 (100 mL) in a 500 mL round-bottom flask. Anhydrous pyridine (20 mL) was added, and the solution was cooled to –40ºC.

Trifluoromethanesulphonic anhydride (Tf_2O) (7.62 g, 27 mmol) was added drop wise by addition funnel. The reaction was gradually warmed to r.t. and stirred for an additional 2 h. TLC (CH_2Cl_2, R_f 0.9) showed the reaction was complete. The reaction was concentrated to dryness and then partitioned between saturated NH_4Cl and CH_2Cl_2. The organic layer was dried over Na_2SO_4 and purified by silica flash chromatography (100% CH_2Cl_2) to yield a clear oil (9.42 g, 98% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.30 (d, *J* = 2.0 Hz, 1H), 8.14 (d, *J* = 2.4 Hz, 1H), 8.08 (d, *J* = 9.2 Hz, 1H), 8.00 (d, *J* = 9.2 Hz, 1H), 7.74–7.14 (dd, *J1* = 8.8 Hz, *J2* = 2.0 Hz, 1H), 7.63–7.60 (dd, *J1* = 8.8 Hz, *J2* = 2.8 Hz, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 146.9, 133.2 131.5, 130.7, 130.2 (2), 129.8, 120.8, 120.7, 119.5. MS (EI) calculated for $C_{11}H_6BrF_3O_3S$ $[M^-]^+$ 353.92, found 354.0.



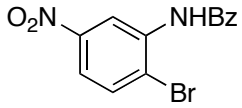
6-benzamidonaphthalen-2-yl trifluoromethanesulfonate (4): CuI (413 mg, 2.17 mmol), *N,N'*-dimethylethylenediamine (385 mg, 4.35 mmol), and K_2CO_3 (6 g, 43.5 mmol) were added to dry toluene (100 mL) in a 250 mL round-bottomed flask under argon. The mixture was stirred for 10 min to form the pre-catalyst, then compound **3** (7.7 g, 21.7 mmol) and benzamide (3.95 g, 32.63 mmol) were added under argon. The reaction was heated to 100ºC in a sealed flask for 48 h. TLC (CH_2Cl_2, R_f 0.4) showed the reaction was complete. After cooling the reaction to r.t., the crude was passed

through a plug of silica gel (200 mL) and washed with copious amounts CH_2Cl_2, then concentrated to dryness to afford an off-white solid (8.3 g, 96% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.57 (s, 1H), 8.60 (d, J = 1.2 Hz, 1H), 8.06 (t, J = 8.4 Hz, 3H), 8.01 (d, J = 7.2 Hz, 2H), 7.96–7.93 (dd, J_1 = 8.8 Hz, J_2 = 1.6 Hz, 1H), 7.64–7.54 (m, 4H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 166.0, 145.9, 138.2, 134.7, 132.6, 131.8, 130.4, 129.6, 128.6, 128.4, 127.7, 122.5, 120.0, 119.1, 116.2. MS (ESI) calculated for $C_{18}H_{12}F_3NO_4S$ $[M+H]^+$ 396.04, found 395.94.

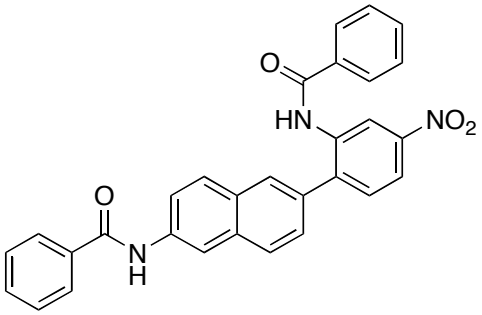


***N*-(6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)naphthalene-2-yl)benzamide (5)**: Pd(dppf)Cl_2 (52 mg, 0.063 mmol) and compound **4** (500 mg, 1.26 mmol) were dissolved in dry dioxane (10 mL) in a 100 mL round-bottom flask under argon. Et_3N (507 mg, 5 mmol), 4,4,5,5-tetramethyl-[1,3,2]dioxaborolane (644 mg, 5 mmol) were subsequently added. The reaction was stirred at 95 ºC under argon for 18 hours. TLC (CH_2Cl_2, R_f 0.3) showed the reaction was done. After cooling the reaction to r.t., it was passed through a plug of silica gel (50 mL), washed with CH_2Cl_2, and then concentrated to dryness. The crude was purified by flash column chromatography (60-100% CH_2Cl_2/hexanes) to yield a solid (306 mg, 65% yield) and triflate-reduced side product (53 mg, 17% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.50 (s, 1H), 8.47 (d, J = 1.2 Hz, 1H), 8.26 (s, 1H), 8.01-7.98 (m, 3H), 7.86–7.81 (m, 2H), 7.68 (d, J

= 8.0 Hz, 1H), 7.61-7.53 (m, 3H), 1.33 (s, 12H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 165.9, 138.0, 135.6, 135.0, 134.9, 131.7, 130.5, 129.3, 129.1, 128.4, 127.7, 126.7, 121.0, 116.1, 83.7, 24.7. MS (ESI) calculated for $C_{23}H_{24}BNO_3$ $[M+H]^+$ 374.18, found 374.14.



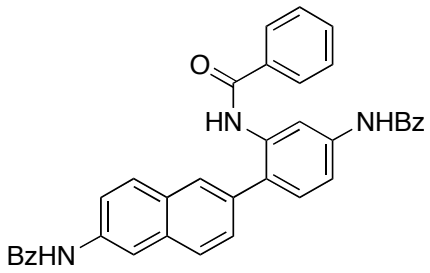
***N*-(2-bromo-5-nitrophenyl)benzamide (6)**: 2-bromo-5-nitroaniline (1.5 g, 6.9 mmol) and Et_3N (1.9 mL) were added to dry CH_2Cl_2 (15 mL) at 0ºC. Benzoyl chloride (1.46 g, 10.42 mmol) was added slowly. The reaction was refluxed for 12 h. TLC (80% CH_2Cl_2/hexanes, R_f 0.6) showed the reaction was complete. The reaction was partitioned between H_2O and CH_2Cl_2. The organic phase was dried over Na_2SO_4 and concentrated to dryness to afford a yellow solid (1.5 g, 70% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.32 (s, 1H), 8.48 (s, 1H), 8.04-8.01 (m, 4H), 7.66-7.55 (m, 3H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 165.3, 146.7, 137.5, 133.8, 133.2, 132.0, 128.4, 127.6, 127.2, 122.1, 121.7. MS (ESI) calculated for $C_{13}H_9BrN_4O_3$ $[M+H]^+$ 320.98, found 320.95.



***N*-(6-(*N*-(3-nitrophenyl)benzamide))naphthalene-2-yl)benzamide (7)**: Compound **6** (100 mg, 0.3 mmol) and compound **5** (117 mg, 0.3 mmol) were added to DMF (3 mL) and 1M aqueous Na_2CO_3 (1.5 mL) and flushed with argon. $Pd(dppf)Cl_2$ (12.7 mg, 0.015 mmol) was added under argon and the reaction was stirred at 80ºC for 16 h. TLC (80% CH_2Cl_2/hexane, R_f 0.3) showed the reaction was complete. After cooling the reaction to r.t., volatiles were removed under vacuum. The residue was dissolved in CH_2Cl_2 and passed through a plug of silica gel (40 mL), washed with copious amounts of CH_2Cl_2, and then concentrated to dryness. The product was purified by flash column chromatography (100% CH_2Cl_2) to yield a yellow solid (140 mg, 95% yield).

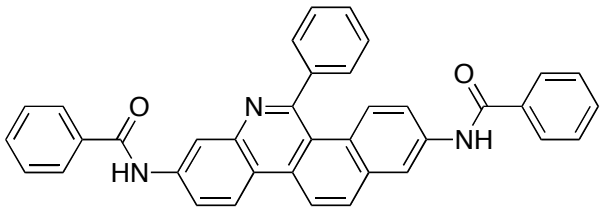
For larger scale reactions (932 mg, 2.5 mmol of boronic ester in this instance), the product can be isolated by precipitation. After removal of volatiles, the crude was washed with water (50 mL) and isolated by vacuum filtration. Purification of the product is aided by its insolubility in most solvents. Successive precipitation of the yellow solid in acetone removes most impurities. Any remaining product in the acetone mother liquor can be isolated by flash column chromatography. Yield: 828 mg (68%); yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.50 (s, 1H), 10.22 (s,

1H), 8.55 (d, *J* = 2.0 Hz, 1H), 8.51 (s, 1H), 8.24–8.21 (dd, *J1* = 8.8 Hz, *J2* = 2.4 Hz, 1H), 8.06 (s, 1H), 8.01 (d, *J* = 7.6 Hz, 2H), 7.96–7.86 (m, 3H), 7.82–7.79 (m, 3H), 7.63–7.59 (m, 2H), 7.55–7.52 (m, 3H), 7.47–7.43 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d6): δ 165.9, 165.8, 146.6., 143.6, 137.6, 136.2, 134.8, 134.0, 133.6, 132.9, 131.8, 131.7 (2), 129.8, 128.6, 128.4 (3), 127.7, 127.6 (2), 126.6, 122.0, 121.4, 121.0, 116.2. MS (ESI) calculated for $C_{30}H_{21}N_3O_4$ $[M+H]^+$ 488.16, found 488.03.



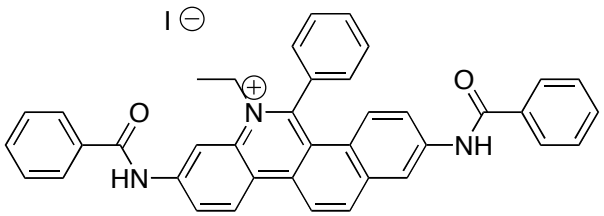
***N*-((6-(2,4-dibenzamide))naphthalene-2-yl)-benzamide (9)**: Compound **7** (175 mg, 3 mmol) was dissolved in EtOH (5 mL) in a 50 mL round-bottom flask. Hydrazine (80%, 0.1 mL) and 10% Pd/C (30 mg) were added under argon. The reaction was stirred at 65 ºC under H_2 (1 atm, balloon) for 4 h. TLC (5% MeOH/CH_2Cl_2, R_f 0.5) showed the reaction was complete. The reaction was cooled to room temperature and passed through a plug of Celite 500 (3 cm), washing with copious amount of MeOH. The filtrate was concentrated to dryness and briefly dried under high vacuum (< 1 torr). The crude was then dissolved in anhydrous DMF (5 mL) with Et_3N (0.5 mL) at 0ºC. Benzoyl chloride (0.078 mL, 0.72 mmol) was added to the flask. The reaction was stirred at r.t. for 12 h. TLC (5% MeOH/CH_2Cl_2, R_f 0.7) showed the reaction was

complete. The reaction was filtered through Whatman #50 filter paper, and then partitioned between saturated $NaHCO_3$ and CH_2Cl_2. The organic phase was dried over Na_2SO_4 and concentrated to dryness. The product was purified by flash chromatography (2% MeOH/CH_2Cl_2) to yield a solid (171 mg, 85% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.46 (s, 2H), 10.01 (s, 1H), 8.45 (s, 1H), 8.05-7.78 (m, 11H), 7.63-7.42 (m, 11H). ^{13}C NMR (100 MHz, DMSO-d6): δ 165.5, 165.4, 165.3, 138.6, 136.6, 135.1, 134.9, 134.7, 134.6, 134.3, 133.1, 131.2, 131.5, 131.4, 131.2, 130.3, 129.8, 128.3, 128.2, 128.1, 128.1, 127.5, 127.3, 127.2, 126.9, 126.8, 120.9, 119.6, 118.4, 116.0. MS (ESI) calculated for $C_{37}H_{27}N_3O_3$ $[M+H]^+$ 562.21, found 562.03.

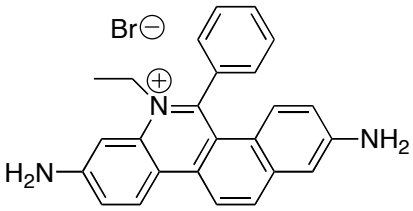


5-phenylbenzo[*i*]phenanthridine-2,8-dibenzyloxycarbamate (10): Compound **9** (260 mg, 0.46 mmol) was heated in $POCl_3$ (5 mL) at 100ºC for 18 hours. TLC (5% MeOH/CH_2Cl_2, R_f 0.3) showed the reaction was done. The reaction was cooled to 0ºC in an ice-water bath and $POCl_3$ was hydrolyzed with cold water. The solution was adjusted to pH 10 with concentrated NH_4OH. The resulting yellow precipitate (226 mg, 89% yield) was filtered through Whatman #50 filter paper and washed with water. ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.70 (s, 1H), 10.52 (s, 1H), 8.88 (d, J = 9.2 Hz,

1H), 8.85 (d, J = 9.2 Hz, 1H), 8.72 (s, 1H), 8.63 (s, 1H), 8.31 (d, J = 8.8 Hz, 1H), 8.14 (d, J = 9.2 Hz, 1H), 8.04 (d, J = 7.6 Hz, 1H), 7.99 (d, J = 7.6 Hz, 1H), 7.65– 7.51 (m, 13H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 165.7, 165.5, 158.5, 139.8, 137.0, 134.6, 134.5, 133.1, 132.9, 131.5 (2), 128.7, 128.5, 128.2, 127.6, 127.5, 125.5, 123.4, 120.7, 120.6, 119.6, 119.4, 119.3, 117.7, 117.4. MS (ESI) calculated for $C_{37}H_{25}N_3O_2$ $[M+H]^+$ 544.20, found 544.38.



5-phenylbenzo[*i*]phenanthridine-2,8-dibenzyloxycarbamate *N*-ethyl iodide (11): Compound **10** (163 mg, 0.3 mmol) and EtI (1 mL) were refluxed in anhydrous DMF (5 mL) for 20 h. TLC (10% MeOH/CH_2Cl_2, R_f 0.4) showed the reaction was done. The reaction was concentrated to dryness and purified by flash chromatography (1-10% MeOH/CH_2Cl_2) to afford a red solid (147 mg, 71% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 11.13 (s, 1H), 10.60 (s, 1H), 9.37 (d, J = 9.6 Hz, 1H), 9.25 (d, J = 1.2 Hz, 1H), 9.08 (d, J = 9.6 Hz, 1H), 8.74–8.72 (m, 2H), 8.53–8.51 (dd, J_1 = 9.2 Hz, J_2 = 1.2 Hz, 1H), 8.09–7.85 (m, 7H), 7.69–7.54 (m, 8H), 6.84 (d, J = 9.6 Hz, 1H), 4.75 (d, J = 6.8 Hz, 2H), 1.61 (t, J = 7.2 Hz, 3H). HRMS (FAB) calculated for $C_{39}H_{30}N_3O_2$ $[M]^+$ 572.2333, found 572.2348.



5-phenylbenzo[*i*]phenanthridine-2,8-diamine *N*-ethyl bromide (2): Compound **11** (200 mg, 0.28 mmol) was dissolved in 48% HBr (4 mL) and refluxed for 14 h. The reaction was concentrated to dryness and purified by flash chromatography (50 mL silica gel, 2-4% MeOH/CH_2Cl_2) to yield a red solid (100 mg, 80% yield). ^{1}H NMR (400 MHz, CD_3CN-d_3): δ 8.70 (d, J = 10.0 Hz, 1H), 8.50 (d, J = 9.2 Hz, 1H), 8.15 (d, J = 9.6 Hz, 1H), 7.82-7.74 (m, 3H), 7.56 (d, J = 6.8 Hz, 2H), 7.43 – 7.41 (dd, J_1 = 7.6 Hz, J_2 = 2.0 Hz, 2H), 7.08 (d, J = 2.8 Hz, 1H), 6.61 (d, J = 9.2 Hz, 1H), 6.55 – 6.52 (dd, J_1 = 9.6 Hz, J_2 = 2.8 Hz, 1H), 5.66 (s, 2H, NH_2), 4.67 (s, 2H, NH_2), 4.55 (q, J = 7.2 Hz, 2H), 1.46 (t, J = 6.8 Hz, 3H). HRMS (FAB) calculated for $C_{25}H_{22}N_3$ $[M]^+$ 364.1808, found 364.1808; UV (MeOH) λ_{max} = 512 nm (ε = 5200 cm^{-1} M^{-1}).

2.12 REFERENCES

1. Moser, H. E., Dervan, P. B., Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **1987**, *238*, 645–650.

2. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.-L., Thoung, N. T., Lhomme, J., Hélène, C., Sequence-specific recognition, photocrossllinking and cleavage of the DNA double helix by an oligo-[α]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **1987**, *15*, 7749–7760.

3. Thuong, N. T., Hélène, C., Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666–690.

4. Duval-Valentin, G., Thuong, N. T., Hélène, C., Specific inhibition of transcription by triple helix-forming oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504–508.

5. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., Wohlrab, F., The chemistry and biology of unusual DNA structures adopted by oligopurine·oligopyrimidine sequences. *FASEB J.* **1988**, *2*, 2939–2949.

6. Manor, H., Rao, B. S., Martin, R. G., Abundance and degree of dispersion of genomic $d(GA)_n$•$d(TC)_n$ sequences. *J. Mol. Evol.* **1988**, *27*, 96–101.

7. Dreyer, G. B., Dervan, P. B., Sequence-specific cleavage of single-stranded DNA: Oligodeoxynucleotide-EDTA•Fe(II). *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 968–972.

8. Sarkar, P. S., Brahmachari, S. K., Intramolecular triplex potential sequence within a gene down regulates its expression in vivo. *Nucleic Acids Res.* **1992**, *20*, 5713–5718

9. Escudé, C., Garestier, T., Sun, J.-S., Drug interaction with triple-helical nucleic acids. *Methods Enzymol.* **2001**, *340*, 340–357.

10. Felsenfeld, G., Davies, D. R., Rich, A., Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.

11. Radhakrishnan, I., Patel, D. J., DNA triplexes: solution structures, hydration sites, energetics, interactions, and function. *Biochemistry* **1994**, *33*, 11405–11416.

12. Radhakrishnan, I., Patel, D. J., Hydration sites in purine·purine·pyrimidine and pyrimidine·purine·pyrimidine DNA triplexes in aqueous solution. *Structure* **1994**, *2*, 395–405.

13. Radhakrishnan, I., Patel, D. J., Solution structure and hydration patterns of a pyrimidine·purine·pyrimidine DNA triplex containing a novel T·CG base-triple. *J. Mol. Biol.* **1994**, *241*, 600–619.

14. Kubinec, M., Wemmer, D., NMR evidence for DNA bound water in solution. *J. Am. Chem. Soc.* **1992**, *114*, 8739-8740.

15. Pilch, D. S., Levenson, C., Shafer, R. H., Structure, stability, and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. *Biochemistry* **1991**, *30*, 6081–6087.

16. Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, N. L., Faucon, B., Rougée, Montenay-Garestier, T., Bisagni, E., Hélène, C., Triple helix-specific ligands. *Science* **1992**, *256*, 1681–1684.

17. Pilch, D. S., Martin, M.-T., Nguyen, C. H., Sun, J.-S., Bisagni, E., Garestier, T., Hélène, C., Self-association and DNA-binding properties of two triple helix-specific ligands: comparison of benzo[*e*]pyridoindole and a benzo[*g*]pyridoindole. *J. Am. Chem. Soc.* **1993**, *115*, 9942–9951.

18. Escudé, C., Nguyen, C. H., Kukreti, S., Janin, Y., Sun, J.-S., Bisagni, E., garestier, T., Hélène, C., Rational design of a triple helix-specific intercalating ligand. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3591–3596.

19. Nguyen, C. H., Marchand, C., Delage, S., Sun, J.-S., Garestier, T., Hélène, C., Bisagni, E., Synthesis of 13*H*-benzo[6,7]- and 13*H*-benzo[4,5]indolo[3,2-*c*]-quinolines: a new strategy of potent specific ligands for triplex DNA. *J. Am. Chem. Soc.* **1998**, *120*, 2501–2507.

20. Escudé, C., Nguyen, C. H., Mergny, J.-L., Sun, J.-S., Bisagni, E., Garestier, T., Hélène, C., Selective stabilization of DNA triple helices by benzopyridoindole derivatives. *J. Am. Chem. Soc.* **1995**, *117*, 10212–10219.

21. Baudoin, O., Marchand, C., Teulade-Fichou, M.-P., Vigneron, J.-P., Sun, J.-S., Garestier, T., Hélène, C., Lehn, J.-M., Stabilization of DNA triple helices by cresecnt-shaped dibenzophenanthrolines. *Chem. Eur. J.* **1998**, *4*, 1504–1508.

22. Wilson, W. D., Tanious. F. A., Mizan, S., Yao, S., Kiselyov, A. S., Zon, G., Strekowski, L., DNA triple-helix specific intercalators as antigene enhancers: unfused aromatic cations. *Biochemistry* **1993**, *32*, 10614–10621.

23. Chaires, J. B., Ren, J., Henary, M., Zegrocka, O., Bishop, G. R., Strekowski, L., Triplex selective 2-(2-naphthyl)quinoline compounds: origins of affinity and new design principles. *J. Am. Chem. Soc.* **2003**, *125*, 7272–7283.

24. Strekowski, L., Say, M., Zegrocka, O., Tanious, F. A., Wilson, W. D., Manzel, L., Macfarlane, D. E., Bis-4-aminoquinolines: novel triple-helix DNA intercalators and antagonists of immunostimulatory CpG-oligodexoynucleotides. *Bioorg. Med. Chem.* **2003**, *11*, 1079–1085.

25. Keppler, M., Zegrocka, O., Strekowski, L., Fox, K. R., DNA triple helix stabilization by a naphthlquinoline dimer. *FEBS* **1999**, *447*, 223–226.

26. Lu, E., Peng, X., Song, F., Fan, J., A novel fluorescent sensor for triplex DNA. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 255–257.

27. Strekwoski, L., Gulevich, Y., Baranowski, T. C., Parker, A. N., Kiselyov, A. S., Lin, S.-Y., Tanious, F. A., Wilson, W. D., Synthesis and structure-DNA binding relationship analysis of DNA triple-helix specific intercalators. *J. Med. Chem.* **1996**, *39*, 3980–3983.

28. Haq, I., Ladbury, J. E., Chowdhry, B. Z., Jenkins, T. C., Molecular anchoring of duplex and triplex DNA by disubstituted anthracene-9,10-diones: calorimetric, UV metling, and competition dialysis studies. *J. Am. Chem. Soc.* **1996**, *118*, 10693–10701.

29. Kan, Y., Armitage, B., Schuster, G. B., Selective stabilization of triplex DNA by antrhaquinone sulfonamide derivatives. *Biochemistry* **1997**, *36*, 1461–1466.

30. Keppler, M. D., Read, M. A., Perry, P. J., Trent, J. O., Jenkins, T. C., Reszka, A. P., Neidle, S., Fox, K. R., Stabilization of DNA triple helices by a series of mono- and disubstituted amidoanthraquinones. *Eur. J. Biochem.* **1999**, *263*, 817–825.

31. Keppler, M. D., James, P. L., Neidle, S., Brown, T., Fox, K. R., DNA sequence specificity of triplex-binding ligands. *Eur. J. Biochem.* **2003**, *270*, 4982–4992.

32. Lee, J. S., Latimer, L. J. P., Hampel, K. J., Coralyne binds tightly to both T·A·T- and C•G•C^+- containing DNA triplexes. *Biochemistry* **1993**, *32*, 5591–5597.

33. Pilch, D. S., Breslauer, K. J., Ligand-induced formation of nucleic acid triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9332–9336.

34. Arya, D. P., Coffee, R. L., Willis, B., Abramovitch, A. I., Aminoglycoside–nucleic acid interactions: remarkable stabilization of DNA and RNA triple helices by neomycin. *J. Am. Chem. Soc.* **2001**, *123*, 5385–5395.

35. Xue, L., Charles, I., Arya, D. P., Pyrene-neomycin conjugate: dual recognition of a DNA triple helix. *Chem. Commun.* **2002**, 70–71.

36. Arya, D. P., Xue, L., Tennant, P., Combing the best in triplex recognition: synthesis and nucleic acid binding of a BQQ–neomycin conjugate. *J. Am. Chem. Soc.* **2003**, *125*, 8070–8071.

37. Ren, J., Chaires, J. B., Preferential binding of a 3,3′-diethyloxadicarbocyanine to triplex DNA. *J. Am. Chem. Soc.* **2000**, *122*, 424–425.

38. Vacquier, V. D., Brachet, J., Chromosomal abnormalities resulting from ethidium bromide treatment. *Nature* **1969**, *222*, 193–195.

39. Lehrman, E. A., Crothers, D. M., An ethidium-induced double helix of poly d(A)·poly r(U). *Nucleic Acids Res.* **1977**, *4*, 1381–1392.

40. Waring, M. J., Complex formation between ethidium bromide and nucleic acids. *J. Mol. Biol.* **1965**, *13*, 269–282.

41. Bresloff, J. L., Crothers, D. M., Equilibrium studies of ethidium–polynucleotide interactions. *Biochemistry* **1981**, *20*, 3547–3553.

42. Waring, M. J., Stabilization of two-stranded ribohomopolymer helices and destabilization of a three-stranded helix by ethidium bromide. *Biochem. J.* **1974**, *143*, 483–486.

43. Scaria, P. V., Shafer, R. H., Binding of ethidium bromide to a DNA triple helix. *J. Biol. Chem.* **1991**, *266*, 5417–5423.

44. Mergny, J. L., Collier, D., Rougée, M., Montenay-Garestier, T., Hélène, C., Intercalation of ethidium bromide into a triple-stranded oligonucleotide. *Nucleic Acids Res.* **1991**, *19*, 1521–1526.

45. Leudtke, N. W., Liu, Q., Tor, Y., On the electronic structure of ethidium. *Chem. Eur. J.* **2005**, *11*, 495–508.

46. Müller, W., Interactions of heteroaromatic compounds with nucleic acids. *Eur. J. Biochem.* **1975**, *54*, 267–277.

47. Espino, G., Kurbangalieva, A., Brown, J. M., Aryl bromide/triflate selectivities reveal mechanistic divergence in palladium-catalysed couplings; the Suzuki–Miyaura anomaly. *Chem Commun.* **2007**, 1742–1744.

48. Klapars, A., Hunag, X., Buchwald, S. L., A general and efficient copper catalyst for the amidation of aryl halides. *J. Am. Chem. Soc.* **2002**, *124*, 7421–7428.

49. Murata, M., Oyama, T., Watanabe, S., Masuda, Y., Palladium-catalyzed borylation of aryl halides or triflates with dialkoxyborane: a novel and facile synthetic route to arylboronates. *J. Org. Chem.* **2000**, *65*, 164–168.

50. Brimble, M. A., Lai, M. Y. H., Suzuki-Miyaura homocoupling of naphthyl triflates using bis(pinacolato)diboron: approaches to the biaryl skeleton of crisamicin A. *Org. Biomol. Chem.* **2003**, *1*, 2084–2095.

51. Giroux, A., Han, Y., Prasit, P., One pot biaryl synthesis *via in situ* boronate formation. *Tet. Lett.* **1997**, *38*, 3841–3844.

52. Hunter, C. A., Lawson, K. R., Perkins, J., Urch, C. J., Aromatic interactions. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651–669.

53. Le Pecq, J. B., Paoletti, C., Interaction of ethidium hydrobromate (EH) with polyribonucleotides. Applications to the study of hybridization reactions. *C. R. Hebd. Seances Acad. Sci.* **1965**, *260*, 7033–7036.

Chapter 3

Design and Synthesis of Neamine and 2-Deoxystreptamine Mimics

---

3.1 PROKARYTOIC TRANSLATION AND AMINOGLYCOSIDES

The rapid emergence of multidrug resistant bacteria is a dire public health concern, where resistance phenotypes appear within months to a few short years after the clinical introduction of most antibiotics.[1] Bacterial strains resistant to last resort antibiotics vancomycin and *Linezolid*™ have been reported, while analysis of current drug pipelines for major pharmaceutical companies reveals few new antimicrobial drugs.[2–6] This is an unabating crisis as bacteria continuously develop new resistance mechanisms, and economic pressures dissuade drug companies from engaging in new antibiotic discovery programs.[6–8] The discovery of new antibiotic classes could potentially kill current drug resistant bacteria,[4,9] but a majority of clinically-administered antimicrobials still utilize molecular scaffolds known for over half a century.

Antibiotics either disrupt the growth of (bacteriostatic) or kill bacteria (bacteriocidal) by acting on different metabolic or reproductive pathways in the cell,[7] including inhibiting the synthesis of the cell wall, DNA, RNA and proteins. Many early clinically administered antibiotics, including streptomycin, tetracycline and erythromycin all act on the ribosome to disrupt protein synthesis.[10] Since proteins are

involved with all biological functions in the cell, disruption of this integral process seems a logical strategy to elicit pathogen death.

Interestingly, most antibiotics that target the ribosome only make contacts with ribosomal RNA and not protein components. Since the mid-twentieth century, one class of antibiotics, the aminoglycosides, has been known to disrupt the fidelity of protein synthesis (known also as translation);[11] however, the actual mechanism was not fully understood until the advent of modern NMR and high resolution X-ray crystallography within the past decade.[10,12] As higher resolution X-ray crystal structures are obtained, scientists are gaining greater understanding of the ribosome's molecular machinery during peptide synthesis and the effects of small molecule ligands on its normal function.[13] A brief description of translation is discussed for greater insight into the deleterious effects aminoglycosides play in normal ribosomal function.[11,14,15]

Prokaryotic translation occurs at the 70S ribosome, composed of a larger 50S and smaller 30S subunit.[16] The cellular components involved in protein synthesis are directed by a messenger RNA (mRNA) template previously synthesized from DNA during transcription. Codons in mRNA, a nucleotide triplet, code specifically for different aminoacyl-tRNAs carrying a particular amino acid according to a "genetic code" elucidated by many groups in the 1960's.[17–19]

Initiation of translation occurs by assembly of the "initiation complex" which includes the 30S and 50S ribosomal subunits, mRNA template, a set of initiation factors, and the first tRNA, fMet-tRNA^{fMet} (transfer RNA-*N*-formylmethionine

complex) which always leads the growing polypeptide chain (Figure 3.1). In the ribosome, there are three binding sites for tRNAs: the aminoacyl (A-site), peptidyl (P-site) and exit (or E site). The A- and P-site are structurally composed of both the 30S and 50S subunits while the E-site is found mostly in the larger 50S subunit. Normally, aminoacyl-tRNA enters the ribosome by binding to the A-site; however, fMet-tRNA^{fMet} is the only aminoacyl-tRNA that specifically binds the P-site first.

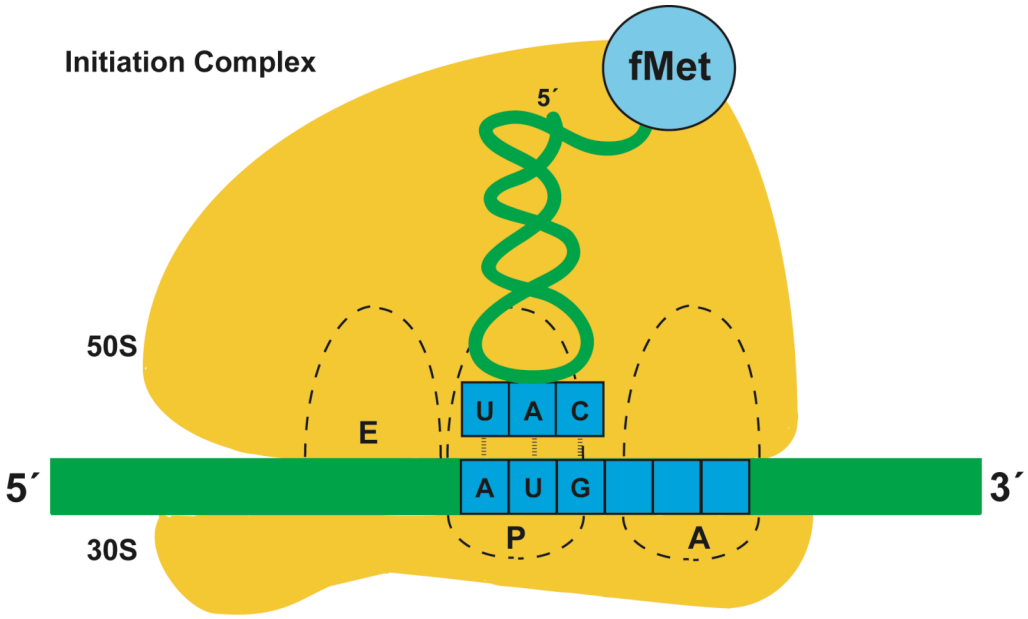


Figure 3.1. The translation initiation complex. The mRNA template (green) resides between the large 50S and smaller 30 subunit of the ribosome (yellow). The initiation codon (AUG), first three blue squares) mediates binding of fMet-tRNA^{fMet} to the P-site instead of the A-site. Initiation factors have been omitted for clarity. (Figures 3.1 – 3..5 were adapted from Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*. 3rd ed.; Worth Publishers: New York, 2000.)

The elongation cycle is a 3-step processes involving the "initiation complex," additional aminoacyl-tRNAs, and a set of elongation factors (Figure 3.2). First, the cognate aminoacyl-tRNA binds the A-site (as determined by the mRNA template). Second, the amino acid in the A-site forms a peptide bond by nucleophilic attack of

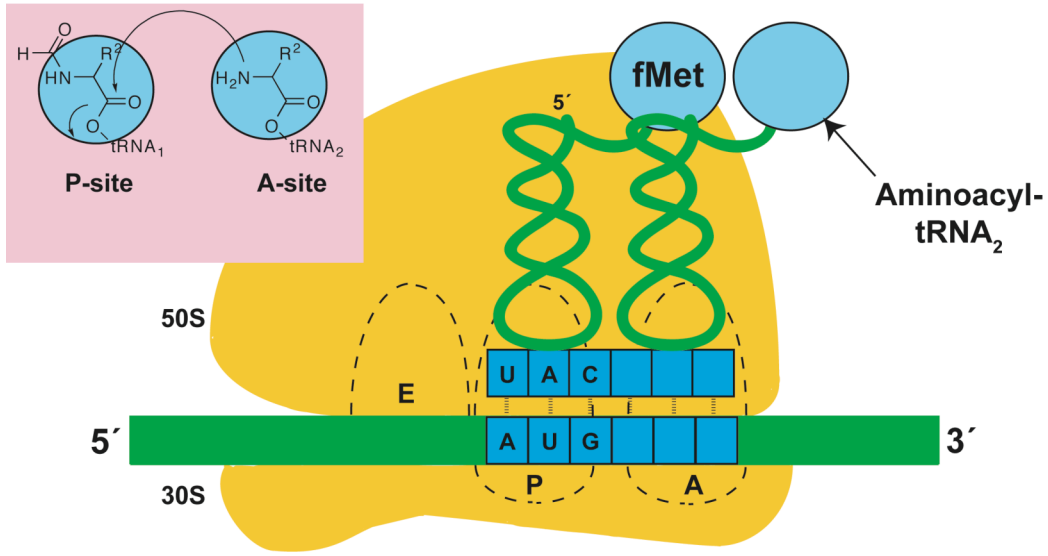


Figure 3.2. First stage of the elongation phase of translation. A second aminoacyl tRNA binds to its cognate sequence in the empty A-site. *Inset*. The amino acid of the second tRNA in the A-site forms a peptide bond by nucleophilic attack of fMet.

the fMet in the P-site, breaking the ester bond between fMet and the initial tRNA (Figure 3.3). The final step is translocation where the ribosome moves one codon in the 3′-direction of the mRNA template. This ribosomal shift causes the tRNAs, both attached to their respective codons along the mRNA, to move towards the E-site. Specifically, the deacylated $tRNA^{fMet}$ now occupies the E-site after movement from the P-site, while the didpeptidyl-tRNA now resides in the P-site after leaving the A-site (Figure 3.4). Elongation can now repeat with the A-site empty and the mRNA codon reading frame accessible for the next cognate aminoacyl-tRNA.

Termination of polypeptide synthesis is signaled by one of three codons in mRNA (UAA, UAG, UGA). A termination or release factor occupies the empty A-site, followed by hydrolysis of the peptide-tRNA bond, and release of the polypeptide

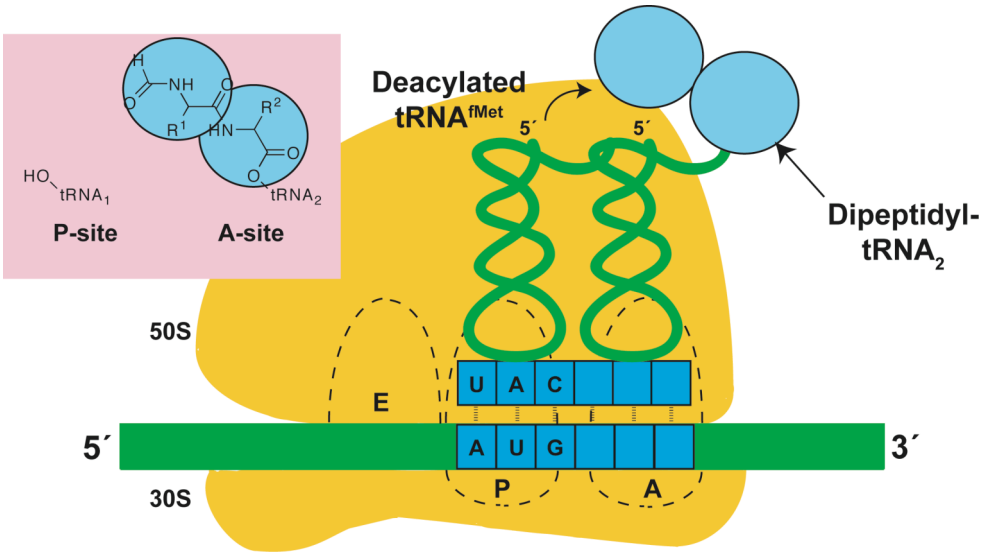


Figure 3.3. Second stage of the elongation phase of translation. The dipeptide now is bonded to the tRNA residing in the A-site, leaving a deacylated tRNA^{fMet} in the P-site. *Inset*. The resulting dipeptide (A-site) and free hydroxyl group (P-site).

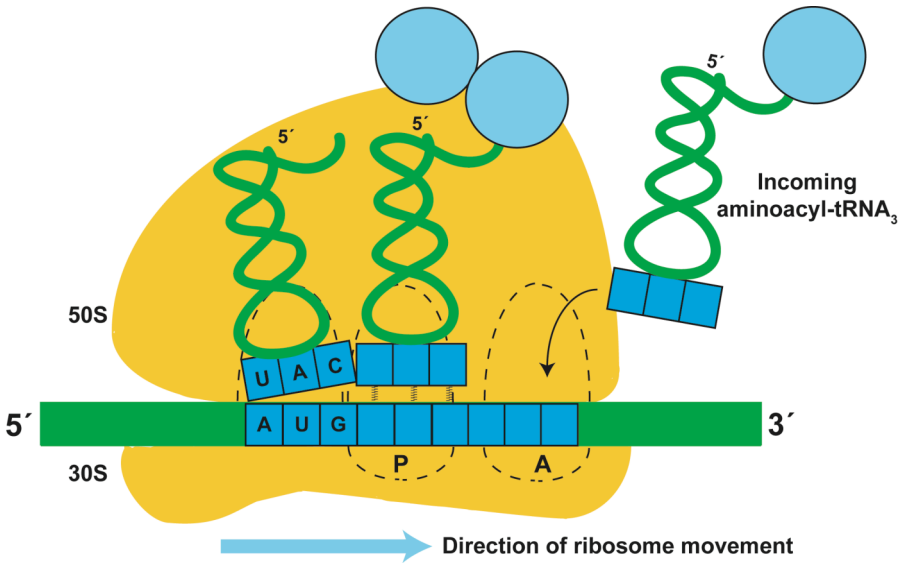


Figure 3.4. Third stage of the elongation phase of translation. Along the mRNA template, a 5′ to 3′ movement of the ribosome shifts the deacylated tRNA^{fMet} into the E-site and the dipeptidyl tRNA into the P-site, leaving the A-site empty for the next cognate aminoacyl tRNA. Elongation repeats until a termination codon in mRNA is reached.

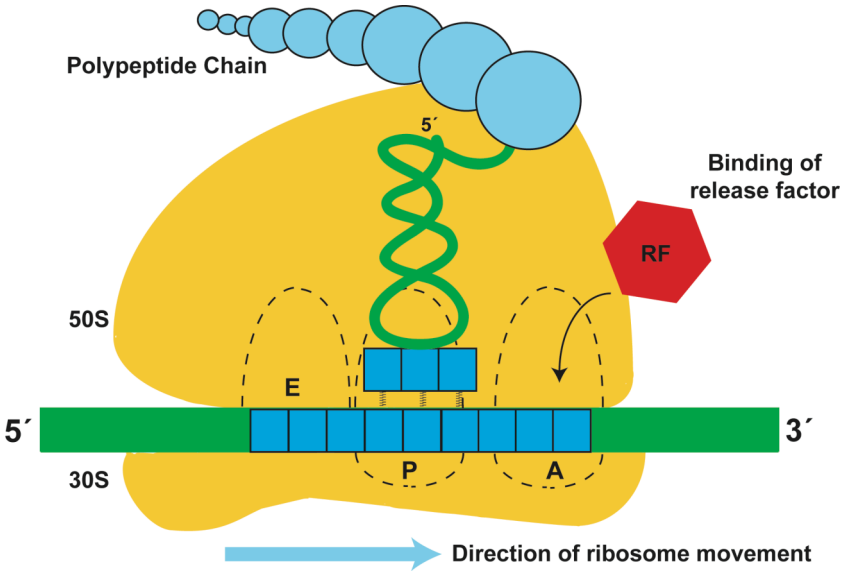


Figure 3.5. Termination of translation. A release factor binds the empty A-site signaling the end of translation and dissociation of the entire complex.

and tRNA (Figure 3.5). Finally, cofactors, tRNA, ribosomal subunits, and mRNA dissociate for later use in polypeptide synthesis.

The structural basis for the deleterious effects caused by aminoglycosides on peptide synthesis was unknown until the seminal work of Noller and co-workers.[20] It was revealed through footprinting studies that the antibiotics bind the ribosomal RNA of the A-site. The exact contacts made between the A-site rRNA and aminoglycosides were determined by the development of smaller RNA systems[21] and NMR studies.[12,22] Specifically, the binding of paromomycin induced a conformational change in the internal loop of the prokaryotic A-site as observed by NMR.[12] Further studies revealed that three adenine residues, A1408, A1492 and A1493, are displaced towards the minor groove upon antibiotic binding.[23] Puglisi and colleagues hypothesized these adenines play a role in recognition of cognate aminoacyl-tRNA.[24] Upon binding of

the anticodon stem loop to the mRNA helix in the ribosomal A-site, bases A1492 and A1493 make contacts with the mRNA backbone which provides a possible mechanism for discrimination by the ribosome (Figure 3.6).

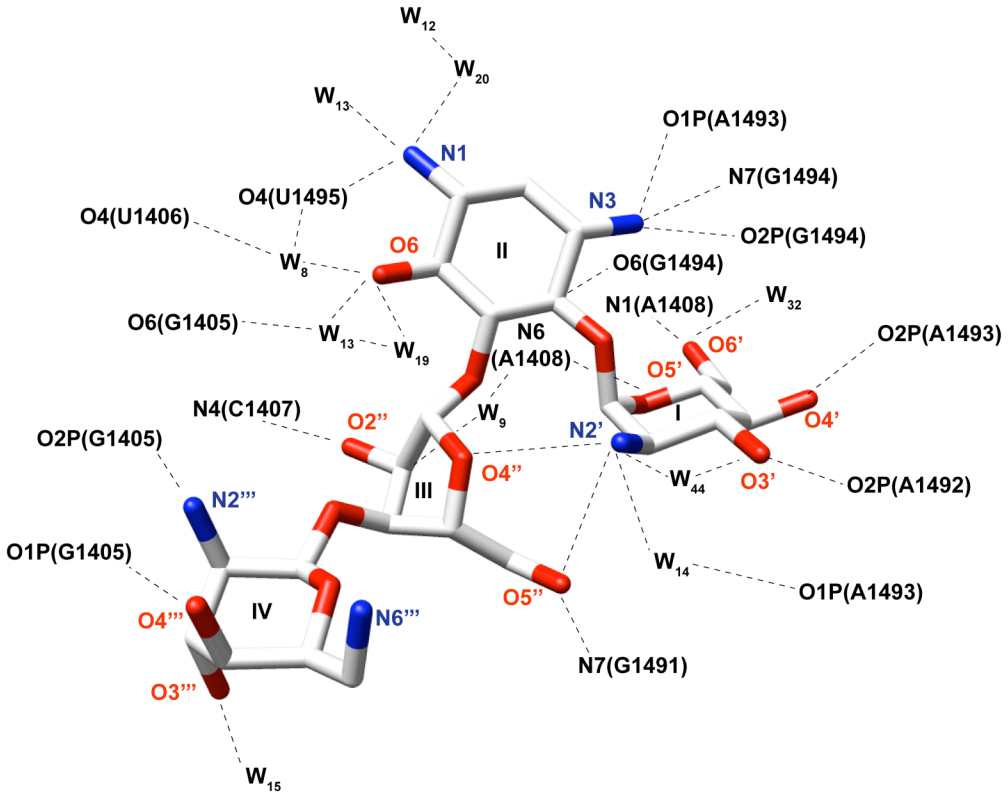


Figure 3.6. Contacts made between paromomycin and the prokaryotic A-site as determined by X-ray crystallography. The aminoglycoside with rings (I–IV) and nitrogen (blue) and oxygen (red) atoms are specified. The RNA atoms are numbered according to *E. coli* numbering. W corresponds to water. Hydrogen-bonding interactions are represented by dashed lines. PDB ID. 1J7T. (Adapted from Vicens, Q., Westhof, E. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **2001**, *9*, 647–658.)

Two years later in 2001, a high resolution X-ray crystal structure of tRNA and paromomycin docked in the A-site provided greater understanding of the role aminoglycosides play in affecting translation.[25] In the absence of aminoglycosides,

residues A1492 and A1493 rotate freely in and out of the RNA helix. Upon binding of the mRNA template by cognate tRNA in the ribosome, these same bases flip out into the minor groove. Rotated out the helix, A1493 and A1492, interact with the first two bases in the tRNA anticodon–mRNA codon helix, respectively. The interaction is specific enough for the ribosome to discriminate between any Watson-Crick or mismatched base pairing. One specific contact observed is hydrogen-bonding between the 2′-hydroxyl of A1493 and the 2′-hydroxyl of the first nucleotide of the mRNA codon (Figure 3.7a).

Paromomycin and other aminoglycosides are detrimental to this decoding process because they cause similar ribosomal structural changes in the absence of cognate tRNA, preventing the ribosome from controlling recognition (Figure 3.7b). The flipping of A1492 and A1493 is energetically favorable upon cognate tRNA binding, whereas binding of near-cognate tRNA causes energetically less favorable structural changes in the ribosome. Upon binding of paromomycin, the ligand pays part of the energetic cost of inducing structural change, allowing for near-cognate tRNA to bind. Since the energetic barrier for discrimination is lowered, the ribosome is not able to ensure fidelity in protein synthesis. The increasing concentration of malformed proteins then normally results in bacterial cell death.

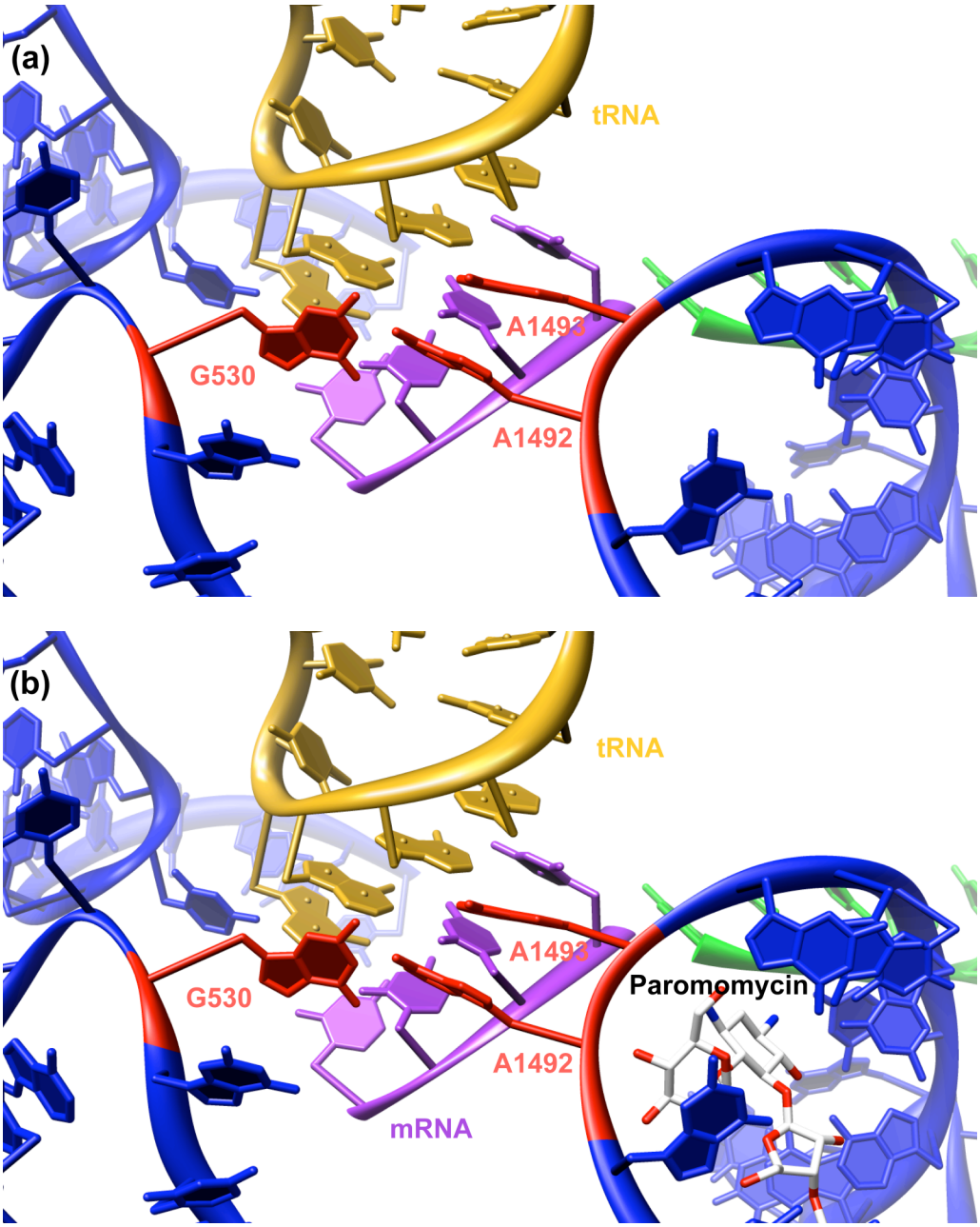


Figure 3.7. Structure of the prokaryotic A-site docked with a tRNA, paromomycin and mRNA. (a) In the absence of paromomycin, the anticodon stem loop of cognate tRNA (yellow) binds the A-site (16S RNA in blue) to its complementary mRNA codon (purple). The ribosome recognizes the cognate tRNA-mRNA match by conformational changes with A1492, A1493 and G530. (b) In the presence of paromomycin, A1492 and A1493 are also rotated out of the helix in a similar manner, disrupting the normal cognate tRNA recognition processes by the ribosome. PDB ID: 1IBL. (Ogle, J. M., Brodersen, D. E., Clemons Jr., W. M., Tarry, M. J., Carter, A. P., Ramakrishnan, V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001**, *292*, 897–902).

3.2 STRUCTURE OF AMINOGLYCOSIDES

Aminoglycosides are a class of pseudo-oligosaccharide antibiotics that have been clinically used for over 50 years,[26] and all are characterized by a central aminocyclitol moiety. The neomycin and tobramycin/kanamycin class of aminoglycosides contain a 2-deoxystreptamine (2-DOS) aminocyclitol (Figure 3.8a–c). Specifically, neomycin is substituted at the 4- and 5-positions of 2-DOS while the tobramycin/kanamycin family is substituted at the 4- and 6-positions. In both families, ring I is the aminosugar moiety glycosidically bonded at the 4-position. 2-DOS is considered ring II, while ring III is bonded either at the 5- or 6-position of 2-DOS for the neomycin or tobramycin/kanamycin family, respectively. Ring IV is glycosidically bonded to ring III, in the case of the neomycin family.

Atypical aminoglycoside antibiotics include streptomycin, apramycin, spectinomycin, and hygromycin which all contain various streptamine-based cores (Figure 3.8d). The streptamine core of streptomycin is substituted at the 4-position with guanidinium groups at the 1- and 3-positions. Hygromycin contains a mono-substituted 2,4-dideoxystreptamine core while apramycin contains a 4-substituted 2-DOS aminocyclitol. Additionally, spectinomycin contains a fused aminocyclitol core. Since antibiotics are secondary metabolites developed by specific bacteria as a defense mechanism, other microbes have developed individual enzymes to selectively recognize each individual scaffold to combat these efforts.

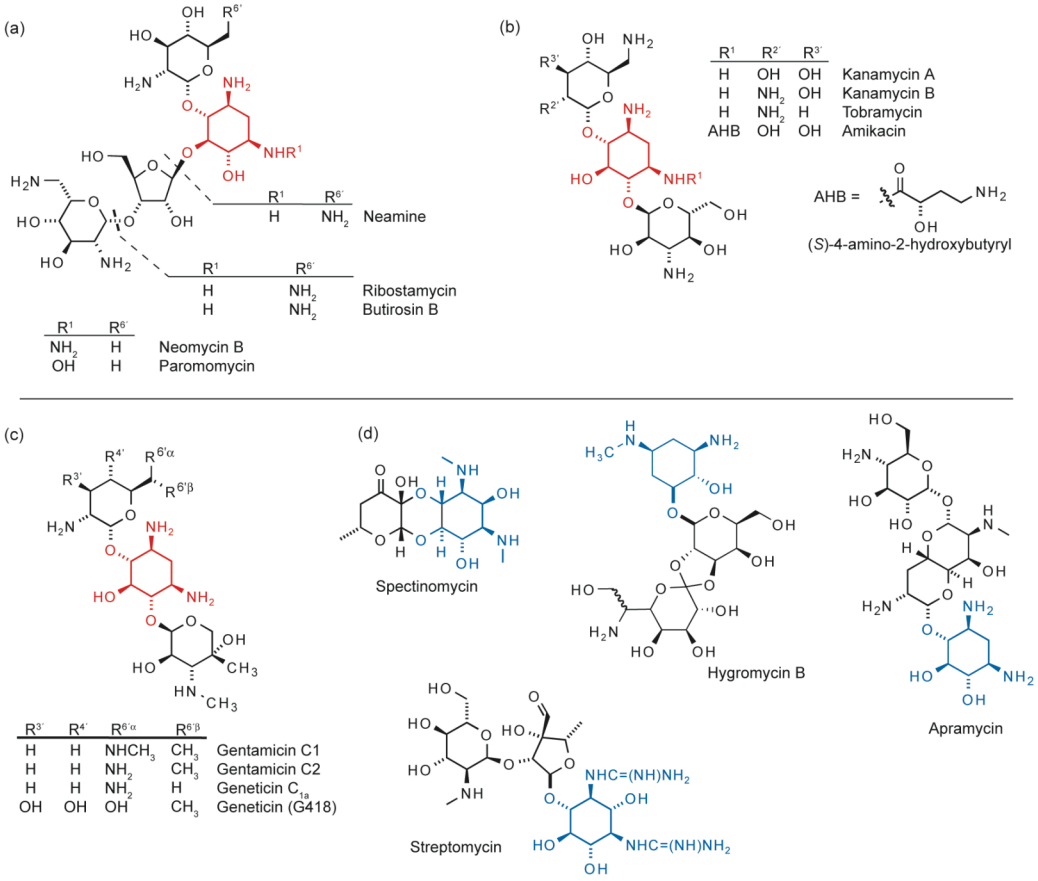


Figure 3.8. Structures of different aminoglycoside families. (a) Aminoglycosides in the neomycin family, characterized by a 4,5-disubstituted 2-DOS (red) core. (b) Aminoglycosides in the kanamycin/tobramycin family characterized by a 4,6-disubstitued 2-DOS (red) core. (c) Aminoglycosides from the gentamicin family, also characterized by a 4,6-disubstituted 2-DOS (red) core. (d) Atypical aminoglycosides with differing streptamine cores (blue).

3.3 THE MECHANISM OF AMINOGLYCOSIDE RESISTANCE

Bacteria utilize three main strategies for antibiotic resistance: a) drug efflux, b) modification of the antibiotic, and c) modification of the drug target.[7,27] Although active drug transport out of the cell is the primary mechanism for drug resistance, a majority of aminoglycoside resistant bacteria enzymatically modify the drug itself or RNA bases in the 16S rRNA A-site.[27,28]

Aminoglycoside Modification. Aminoglycosides are highly functionalized molecules that make specific contacts with its RNA target. Modification of functional groups can potentially disrupt binding while decreasing specificity and affinity for the A-site. Three chemical modifications have been identified to be caused by unique enzymes observed in resistant bacteria: a) *O*-adenylation, b) *O*-phosphorylation, and c) *N*-acetylation. For example, kanamycin has been observed to be regioselectively modified by various enzymes on ring I, which is known to make essential hydrogen bonding contacts with the phosphate backbone and A1408.[29] Studies with *O*-adenylyltransferase ANT(4′) reveal an ATP driven modification of the 4′-position. This secondary alcohol has been shown to make critical hydrogen bonding contacts with the phosphorous backbone of A1493. The resulting AMP modification drastically reduces binding.[27] Additionally, modified aminoglycosides can still be effluxed from the cell as a secondary resistance mechanism. In another example, *O*-phosphorylation by APH(3′) of *Enterococcus faecalis* acts on the 3′-hydroxyl of kanamycin A. X-ray crystal structures have shown this alcohol to interact with the phosphate of nucleotide A1492.[29] Finally, *N*-acetylation of the primary amine in neomycin and kanamycin is modified by *N*-acetyltransferase AAC(6′). Recall, this amine makes a pseudo-Watson-Crick base pair with A1408, which is an anchor point for binding. A summary of modifications on kanamycin B is shown in Figure 3.9, along with the numerous modifying enzymes that are detrimental to aminoglycoside binding.

RNA Target Modification. In addition to aminoglycoside modification, RNA target modification is a strategy utilized by drug resistant bacteria to lower drug binding affinity.[27] One predominant modification is methylation of key rRNA bases required for

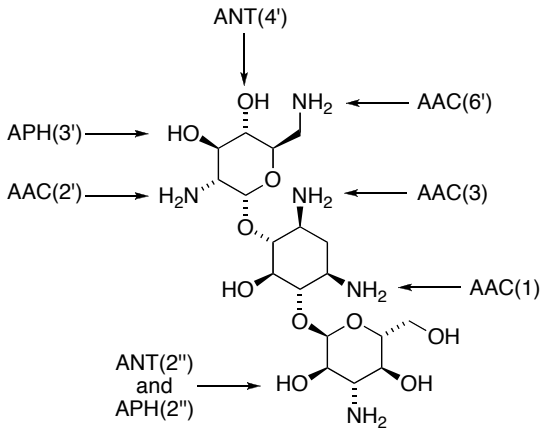


Figure 3.9. Modification sites of various resistance enzymes on kanamycin B. ANT, aminoglygocisde nucleotidyltransferase; APH, aminoglycoside phosphotransferase; AAC, aminoglycoside acetyltransferase. The primed notation indicates the location of modification by the enzyme.

aminoglycoside recognition, catalyzed by rRNA methyltransferases. Methylation at N1 of 1408 has been observed to confer resistance against kanamycin, tobramycin and apramycin while not disrupting gentamicin's bacteriocidal effects.[27] Some gentamicin producing bacteria prevent suicide by methylation of the N7 position of G1405 which conveys resistance to all 4,6-disubstituted 2-DOS aminoglycosides. From X-ray crystal structures, the 2″-NH_2 group on ring III of tobramycin and kanamycin class aminoglycosides makes a direct interaction with N7 of G1405.[29] Methylation presumably disrupts this interaction, therefore disrupting binding.

3.3 AMINOGLYCOSIDE MIMETICS

The development of synthetic or semi-synthetic aminoglycoside mimics that still elicit a similar mode of action but with a different scaffold to elude recognition by modifying enzymes may be effective against current drug resistant bacteria.[30] Numerous groups have synthesized aminoglycoside mimics that have provided insights into the structure–activity relationship of ligand–A-site interactions. Important examples of aminoglycoside mimics will be presented to demonstrate the strategies and molecular scaffolds used to date. Mimics and derivatives are organized by research group since similar synthetic methodologies and scaffolds dominate the chemistry of certain laboratories. However, this discussion is not exhaustive as new scaffolds are reported monthly. Moreover, the groups selected have provided particular inspiration for the design of our own aminoglycoside mimics.

Thomas Hermann & Anadys Group. Thomas Hermann and colleagues at Anadys Pharmaceuticals (San Diego, CA), and more recently the University of California, San Diego, have actively synthesized a number of structure-guided aminoglycoside mimics (Figure 3.10). Many mimics have been designed from X-ray crystal and NMR complexes of natural aminoglycosides docked within the ribosomal A-site. In one example, a novel series of paromamine derivatives were synthesized with different hydrogen bond accepting and donating groups substituted at the 5′ and 6′ positions of the glucosamine ring (ring I).[31] Evaluated in an in vitro transcription-translation assay, many had IC_{50} values one order of magnitude lower than natural paromamine. More importantly, the 6′ position was revealed to be extremely pivotal

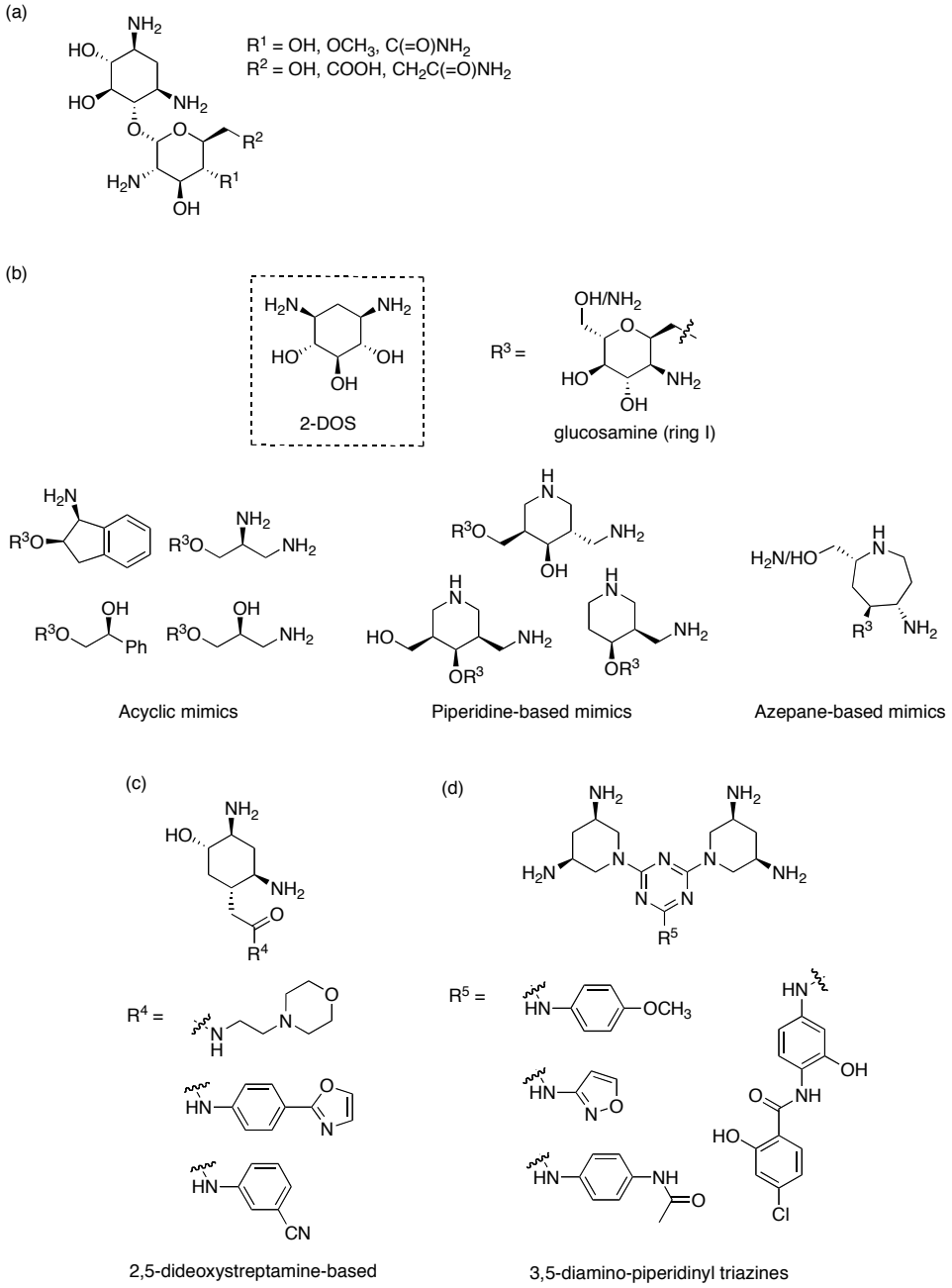


Figure 3.10. Examples of aminoglycoside and 2-DOS mimics synthesized by Hermann and colleagues. (a) Paromamine derivatives with differing substituents at the 5′ and 6′ positions in order to probe the hydrogen-bonding interactions of ring I with RNA. (b) Families of 2-DOS mimics based and heterocyclic and acyclic scaffolds. (c) 2,5-Dideoxystreptamine-based neamine mimics. (d) 3,5-Diamino-piperidinyl triazines as a scaffold to mimic 2-DOS and aminoglycosides.

for RNA recognition and sensitive to functional group substitution. Recent efforts have been focused on utilizing different heterocycles and short acyclic polyamines to mimic the 2-DOS ring, a core component in aminoglycosides. Small libraries of aminomethyl piperidines,[32] aminoazepanes,[33] 3,5-diaminopiperidines and 3,5-diamino-piperidinyl triazines (DAPT) conjugates,[34,35] as well as 2,5-dideoxystreptamine[36] and short polyamines and alcohols[37] have all been synthesized and evaluated utilizing translation inhibition assays (Figure 3.10b). Notably, derivatives containing the DAPT scaffold have been proven to inhibit translation to a moderate degree and are being investigated further as a potential drug candidate.[34]

Chi-Huey Wong. In the late 1990's, the Wong Group hypothesized that hydroxyamine motifs in aminoglycosides are responsible for their selective recognition of RNA structures. Conceivably, the 1,2- and 1,3- hydroxyamine substructures found in aminoglycosides would be able to recognize the phosphodiester backbone as well as form a "Hoogsteen-type" hydrogen bonding interaction with base pairs (Figure 3.11a). Several model compounds were synthesized with 1,2- and 1,3-hydroxyamine substructures and were shown through NMR studies to bind phosphate and sulfate ions in a bidentate manner as hypothesized (Figure 3.11b).[38] In addition, using the aminoglucopyranoside core as an initial scaffold, a small molecule library with various tripeptide side chains were assembled and screened rapidly by surface plasmon resonance (SPR) against the prokaryotic A-site (Figure 3.11c). Several of the library compounds had affinities similar to neamine, suggesting the hydroxyamine motif is an important RNA recognition element, despite the large variations in amino

acid side chains.[39] Dimeric aminoglycosides were also designed to both increase affinity to the prokaryotic A-site but also as a strategy to circumvent bacterial resistance modifying enzymes (Figure 3.11d). In one example, synthesized neamine and nebramine dimers had lower MIC (minimum inhibition concentration) values compared to tobramycin in translation assays with resistant *P. aeruginosa*.[40]

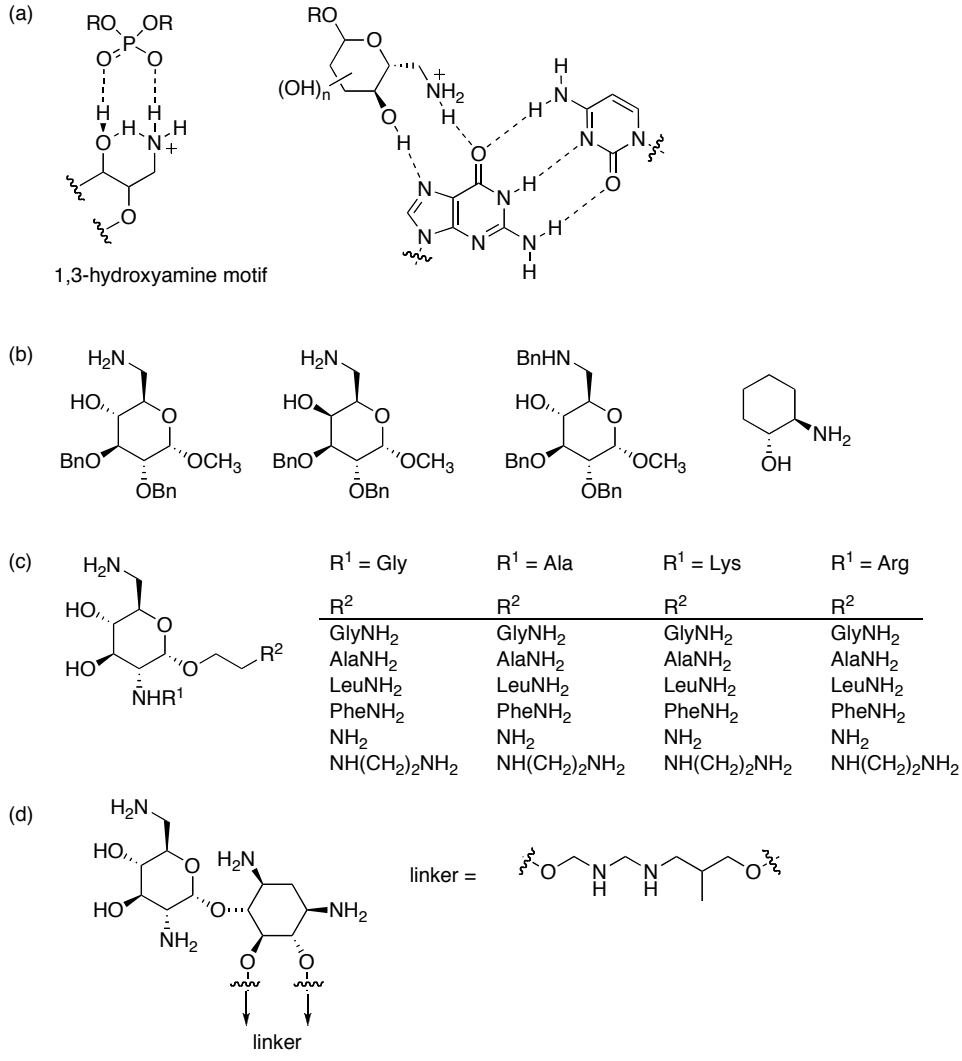


Figure 3.11. Aminoglycoside mimics synthesized by Wong and colleagues. a) The 1,3-hydroxyamine motif hypothesized to mediate RNA–ligand recognition. b) 1,3-Hydroxyamine motifs tested for binding to RNA. c) Tripeptide aminoglycoside conjugates. d) Dimeric 2-DOS aminoglycosides with representative linker.

Larger libraries of synthetic aminoglycosides based off the neamine core were systematically screened. One study probed the role of ring IV in neomycin in A-site binding (Figure 3.12a).[41] The idose ring was replaced with polyamine tails, or analogs containing only one hydroxyl or amino group. Interestingly, a derivative with only alcohols had the same binding affinity as ribostamycin which completely lacks ring IV. Another derivative with two polyamines was found to have affinity similar to neomycin B, the tightest A-site binder.[41] From these observations, Wong suggests that ring IV does not play a significant role in determining affinity or specificity.[41] Another synthetic library was made where the position of amino groups was varied around the glycopyranose ring of neamine (Figure 3.12b).[42] Polyamines and heterocycles were also attached to the 5- and 6-position of 2-deoxystreptamine to mimic the larger neomycin structure (Figure 3.12c). Although many bound an A-site construct more tightly than neamine as determined by SPR, they were worse in inhibiting translation. This suggests the lack of correlation between binding and antibiotic activity.[43] Through the years, the development of methodologies to rapidly synthesize different oligossacharides allowed Wong and co-workers to form combinatorial libraries of tobramycin.[44] In this case, the third ring attached at the 6-position of 2-DOS was effectively replaced with sugars of differing alcohol and amine composition utilizing thioglycoside building blocks. Facile entry to kanamycin B analogues was also made possible by recently developed selective glycosylation methodology.[45] Both cases provided analogs with affinities similar to their respective parent compound at best.

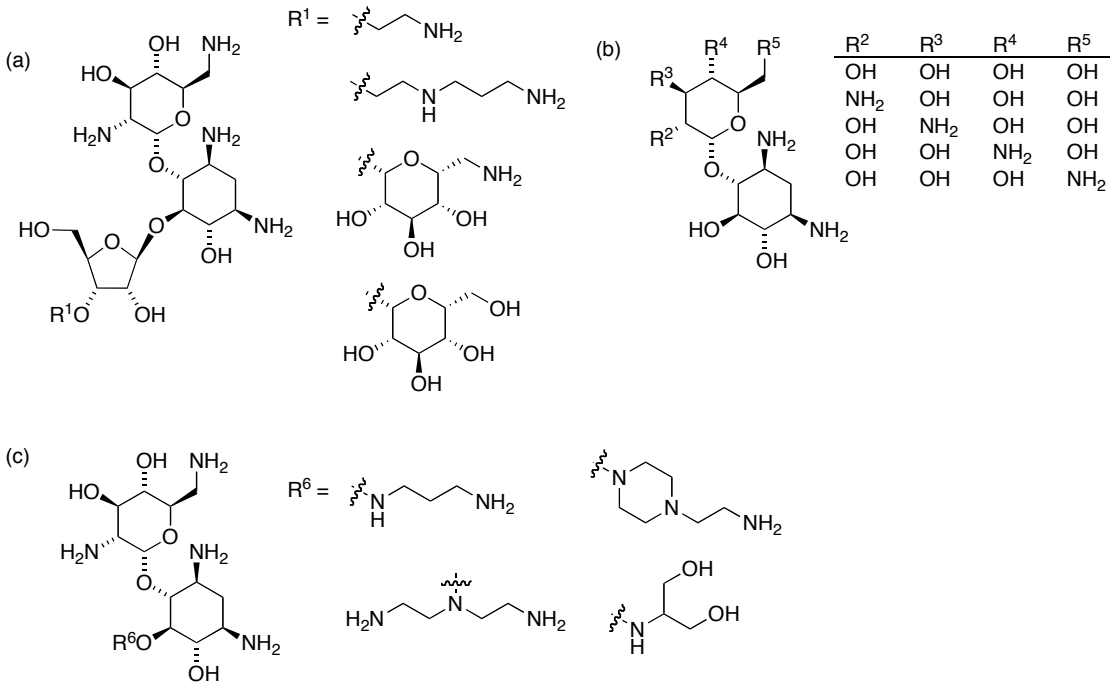
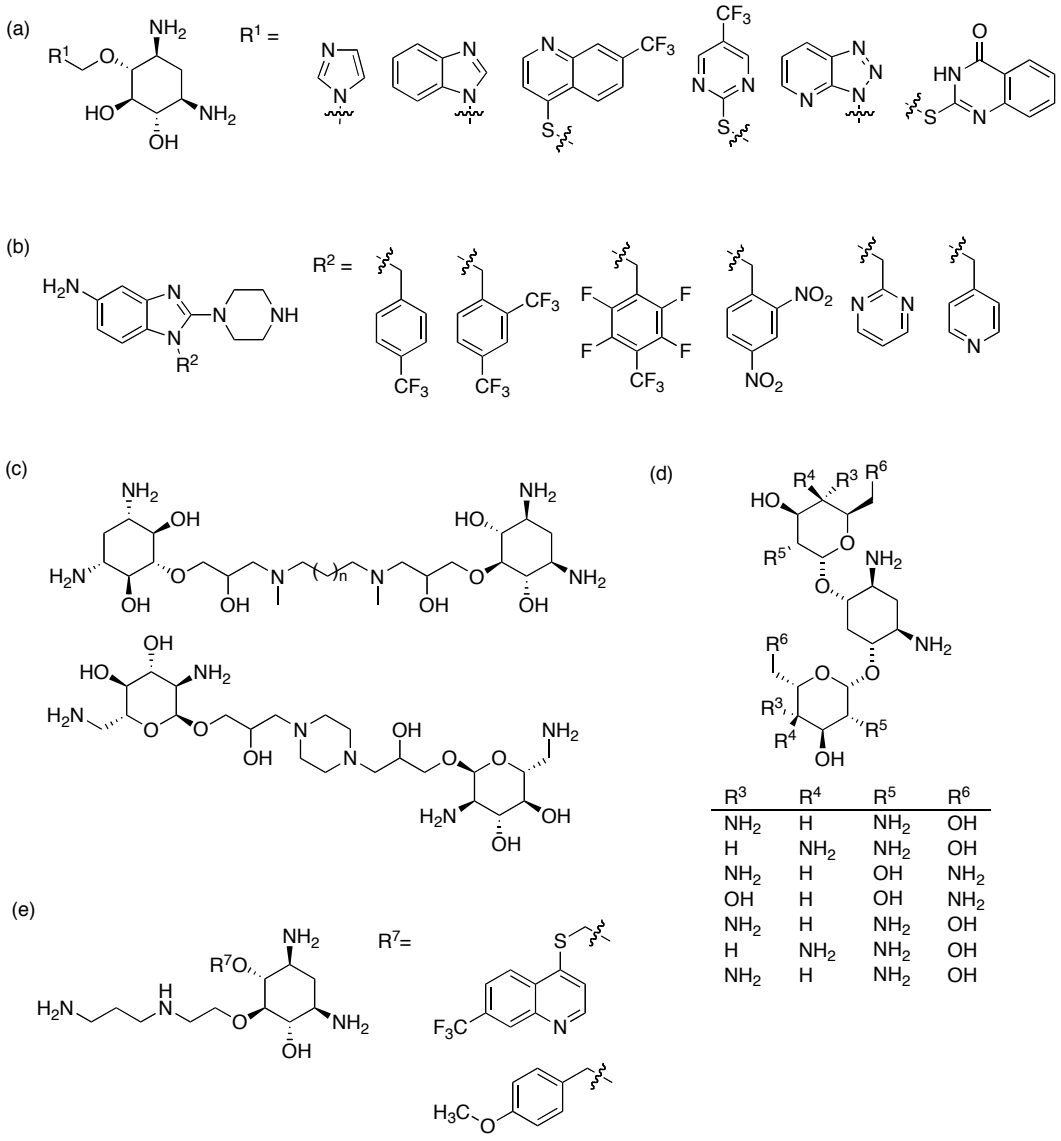


Figure 3.12. Additional aminoglycoside mimics synthesized by Wong and colleagues. a) Linear amines and sugars that mimic ring IV of the neomycin family. b) Systematic substitution of amines in ring I of neamine. c) Acyclic and cyclic polyamines and polyols used to mimic the neamine.

Richard Griffey & Isis Group. Many of the aminoglycoside mimics synthesized by Griffey and coworkers at Isis Pharmaceuticals (Carlsbad, CA) have utilized numerous heterocycles to mimic the neamine core. More importantly, rapid screening has been possible through the use of their proprietary ESIMS-based affinity assay.[46] In one example, nearly 20 different heterocycles were screened to mimic ring I of paromomycin (Figure 3.13a).[47] Replacement of ring I with [7-(trimethyl)-4-quinolinyl]sulfanyl was evaluated as the best mimic in the small library with moderate affinity for the A-site. Although binding affinity was less than paromomycin, the quinolinylsulfanyl derivative bound tighter than apramycin, kanamycin and tobramycin. Separately, a benzimidazole family of molecules synthesized by solution



R^3	R^4	R^5	R^6
NH_2	H	NH_2	OH
H	NH_2	NH_2	OH
NH_2	H	OH	NH_2
OH	H	OH	NH_2
NH_2	H	NH_2	OH
H	NH_2	NH_2	OH
NH_2	H	NH_2	OH

Figure 3.13. Examples of aminoglycoside mimics synthesized by Isis. a) Heterocycle 2-DOS conjugates. b) Benzimidazole derivatives that mimic ring I. c) 2-DOS dimers with different heterocyclic linkers. d) 2,5-Dideoxystreptamine-based aminoglycoside mimics. e) Non-carbohydrate-based aminoglycoside mimics.

and solid phase chemistry after a 2-piperazinyl-benzimidazole derivative was found to competitively displace glucosamine (ring I) from the A-site (Figure 3.13b).[48] This scaffold was used as an anchor for the small library. Other mimics included dimerically linked 2-DOS and carbohydrates found in aminoglycosides (Figure

3.13c),[49] 2,5-dideoxystreptamine neomycin mimics (Figure 3.13d),[50] and the development of a carbohydrate-free aminoglycoside with binding affinity approaching that of paromomycin (Figure 3.13e).

Cheng-Wei Tom Chang. A novel class of aminoglycosides replaces the ring III ribose with a pyranose ring to form the synthetic pyranmycins (Figure 3.14).[52] Assayed against *E. coli*, a thorough SAR of over twenty derivatives discovered two pyranmycins with MIC_{50} values similar to neomycin B. A hybrid of pyranmycin and amikacin, termed pyrankacin, was synthesized utilizing the same strategy and found to be an effective broad-spectrum antibiotic, even having efficacy against MRSA (methicillin resistant *S. aureus*).[53]

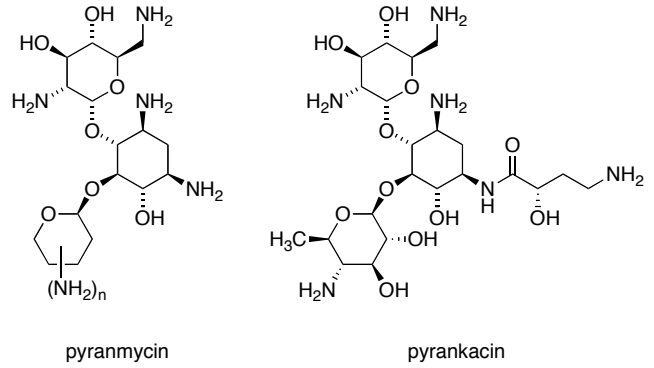


Figure 3.14. Representative pyranmycin structure along with pyrankacin, a hybrid of pyranamycin and amikacin.

Yitzhak Tor. In addition to the structure-activity relationship studies probing the role electrostatics play in RNA recognition (see Chapter 1.12), Tor and co-workers have developed many aminoglycoside derivatives that not only target the prokaryotic A-site but also other RNA targets including the HIV-1 RRE and HIV-1 TAR. Mimics have included neomycin–acridine conjugates,[54] guanidinoglycosides,[55] dimeric

aminoglycosides,[56] conformationlly constrained aminoglycosides[57] and nucleobase–aminoglycoside conjugates.[58] Neomycin–acridine was designed to exploit both intercalation and electrostatic modes of binding (Figure 3.15a).[54] Through fluorescence anisotropy and gel motility assays, the conjugate was discovered to have an order of two magnitude higher affinity than neomycin B for the HIV-1 RRE, enough to effectively disrupt binding of the Rev protein. Another class of derivatives, the guanidinoglycosides, were inspired by the pivotal role guanidinium groups play in recognition along RNA–protein interfaces (Figure 3.15b).[55] Additionally, in comparison to the ammonium group, guanidinium is a more polar, highly basic and rigid functional group capable of directing hydrogen-bonding. Tor and colleagues synthesized a series of perguanidinylated tobramycin, kanamycin and neomycin derivatives. All had higher affinities for the HIV-1 RRE than their natural counterparts, suggesting the guanidinium group can be incorporated in future ligands for increasing RNA binder selectivity and affinity. Separately, symmetric and asymmetric tobramycin, kanamycin and neomycin dimers were shown to inhibit hammerhead ribozyme cleavage more effectively than their monomer counterparts.[56] These findings presented yet another strategy for targeting RNA: covalently-linking two moderately good, highly charged RNA binders can potentially increase their affinity and act as a new scaffold for ligand design.

Recent efforts by the research group have focused on understanding the conformation adapted by aminoglycosides in different RNA targets such as the HIV TAR and prokaryotic A-site.[57] Models of paromomycin docked in the TAR and A-

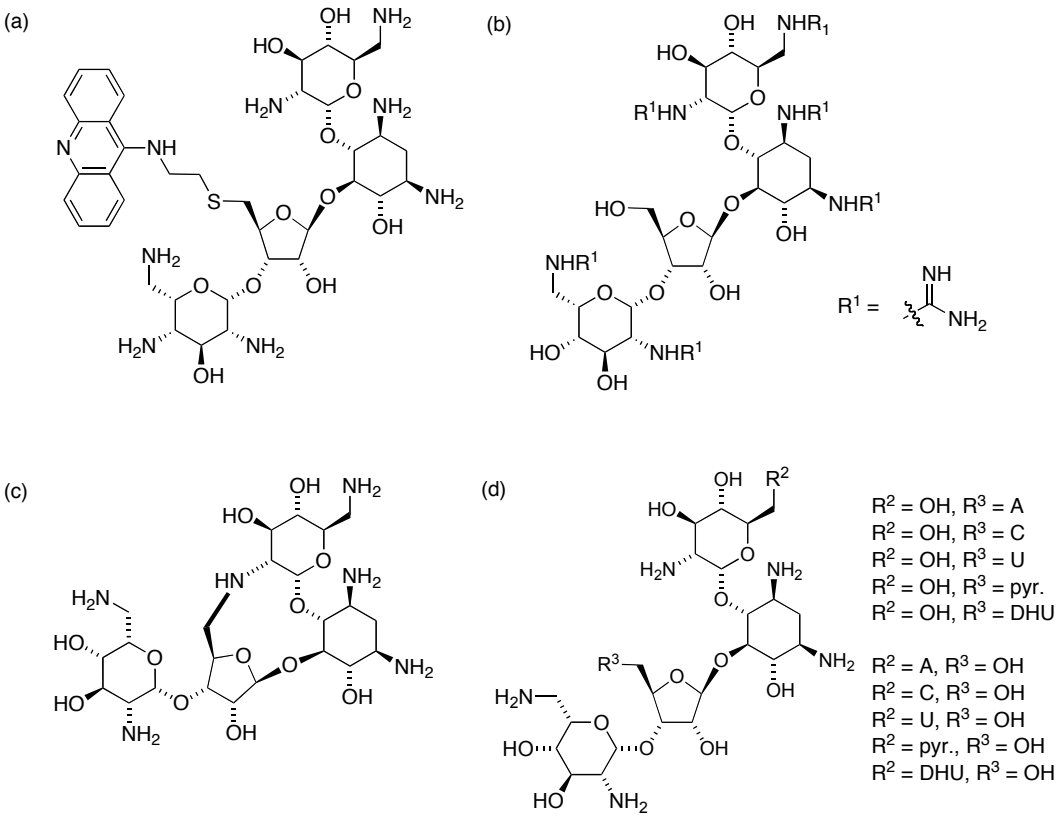


Figure 3.15. Modified aminoglycoside structures synthesized by Tor. (a) Neomycin–acridine conjugate. (b) Neomycin guanidinoglycoside. (c) Conformationally constrained neomycin. (d) Nucleoside–aminoglycoside conjugate. Pyr., 4-pyridone; DHU, dihydrouracil.

site revealed the ligand adapts a different conformation in each RNA due to rotation around the glycosidic bonds. Researchers hypothesized that by covalently linking ring I and III, neomycin would mimic a conformation similar to that adopted upon A-site binding, thereby achieving selectivity between the TAR and A-site (Figure 3.15c). Although a high degree of selectivity was not realized with this constrained analog, studies revealed the degree of flexibility in these RNA structures upon binding of low molecular weight ligands; specifically, that the A-site is a more discriminating RNA target while TAR RNA is more "plastic" and accommodating. The discriminating

behavior of the TAR and A-site were also supported by another study with nucleoside–aminoglycoside conjugates (Figure 3.15d).[58] The conjugates were designed to selectively target the A-site, but had a higher affinity for the TAR possibly due to its accommodating structure.

In summary, the variety of scaffolds utilized to mimic aminoglycoside structure have yielded ligands that bind their RNA target or inhibit translation on the same order of magnitude as their natural counterparts. These scaffolds have included heterocycles, aminoglycosides with variation in connectivity, dimeric aminoglycosides, and libraries of sugar derivatives. The slight structural changes in each of these mimetic classes have provided informative SARs; spurring further development of next generation ligands from encouraging data that natural scaffolds can be adequately mimicked with specially designed molecules.

3.4 TARGETING UNPAIRED A1408 WITH NEAMINE MIMICS

The design of our neamine mimics was based on X-ray crystal structures of paromomycin docked in the prokaryotic A-site.[59] The complex reveals a pseudo-base pairing interaction between ring I of paromomycin and unpaired A1408 in the binding pocket. It may be possible to exploit this base-pairing interaction as an "anchor point" to increase the affinity of aminoglycoside mimics. Conceivably, a mimic that incorporated a scaffold similar to thymine or uracil would allow for favorable interactions similar to an A•T or A•U Watson-Crick base pair. To this end, we

initiated the synthesis of neamine mimics that could potentially base-pair with unpaired residue A1408 (another group has reported a similar strategy).[60]

Our synthetic scheme was designed to be modular to allow for the use of a common precursor. The 2-deoxystreptamine (2-DOS) ring was retained, similar to natural neamine. In order to mimic thymine or uracil, chloropyrimidines were chosen to replace ring I (Figure 3.16). The aryl chlorides act as both centers of diversity and attachment for the 2-DOS ring. Functionalization can be achieved through nucleophilic aromatic substitution (S_NAr) under alkaline condition with various nucleophiles capable of hydrogen-bonding, including amines, methylamines and azides.

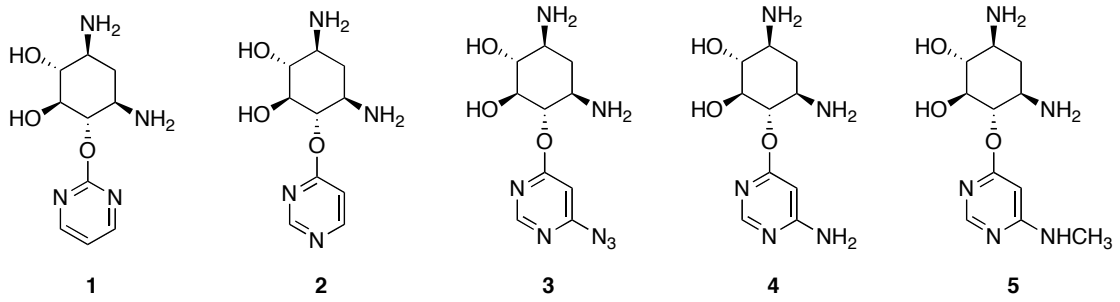
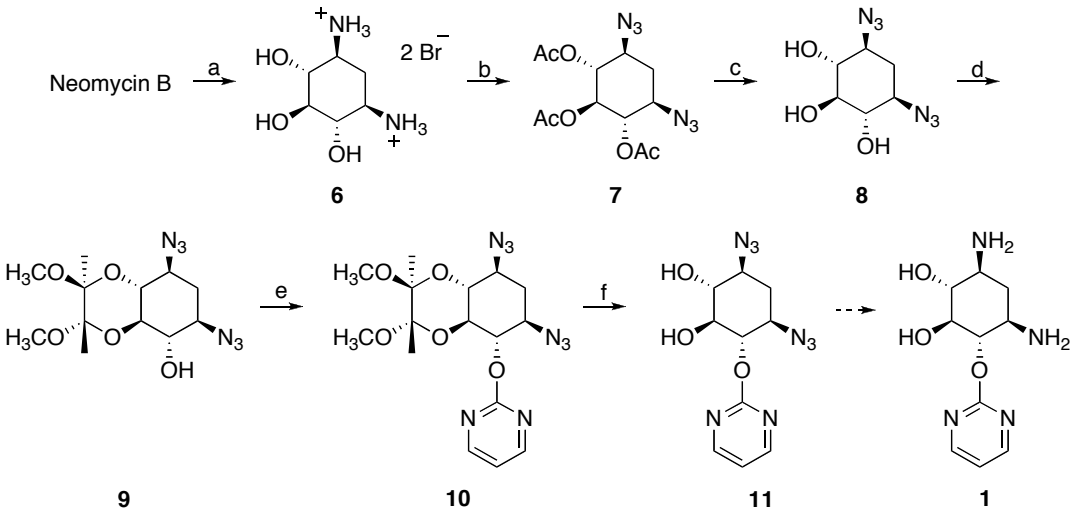


Figure 3.16. Proposed 2-DOS–pyrimidine conjugates as neamine mimics.

Synthesis of the neamine mimics **1–5** commenced with acidic hydrolysis of neomycin B sulfate as a source of 2-DOS (Scheme 3.1). Neomycin B was refluxed in 48% HBr over 16 h to afford a thick, black sludge.[61] After iterative treatments with decolorizing charcoal, precipitation from MeOH yielded the hydrogen bromide salt of 2-DOS **6**. Next, protection of the amino groups by conversion to azides was achieved by standard protocols designed by Wong and colleagues.[62] Coincidentally, a

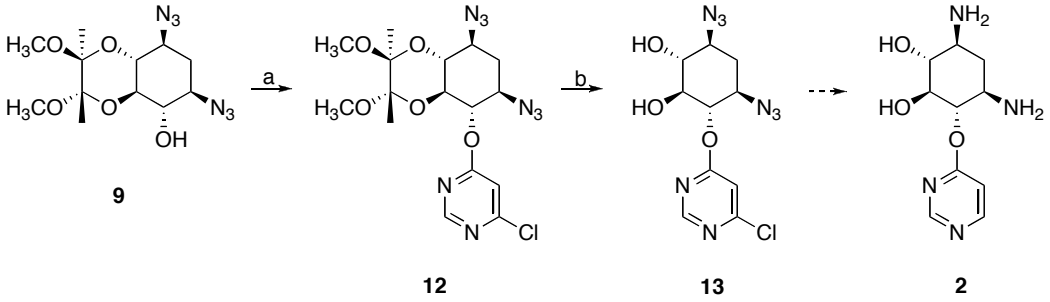


Scheme 3.1. Synthesis towards **1**. *Reagents and conditions*. (a) 48% HBr, reflux, 18 h, 63%; (b) *i*. TfN_3, $ZnCl_2$, Et_3N, $H_2O/CH_3OH/CH_2Cl_2$ (3:10:3), rt, 3 h; *ii*. Ac_2O, pyr., 5 mol% DMAP, rt, 3 h, 75%; (c) *i*. 30 mol% $NaOCH_3$, CH_3OH, rt, 12 h; ii. DOWEX 50WX4-100 H^+, 95%; (d) $CH(OCH_3)_3$, 2,3-butanedione, (±)-CSA, CH_3OH, reflux, 36 h, 96%; (e) NaH, 2-chloropyrimidine, DMF, –78 °C to rt, 6 h, 60%; (f) TFA, rt, 16 h, 91%. *Abbreviations*. DMAP, 4-(dimethylamino)pyridine; (±)-CSA, *DL*-10-camphorsulfonic acid; TFA, trifluoroacetic acid.

trifluoromethanesulfamide byproduct eludes simultaneously with the azide product during silica gel chromatography. Acetylation of the three alcohols to form a more nonpolar peracetylated product **7** allowed for easier purification. The acetyl groups were removed under strongly basic conditions to afford **8** in quantitative yield. Two of the three alcohols were unselectively masked using 2,3-butanedione developed by Ley and coworkers to form an enantiomeric mixture of the diketal **9**.[63] The secondary alcohol **9** was subsequently treated with NaH and 2-chloropyrimidine to form the 2-DOS-heterocycle conjugate **10**. Acidic treatment with TFA removed the diketal protecting group to afford **11**, while subsequent attempts to reduce the azide were unsuccessful. Reduction under Staudinger conditions with tributylphosphine in addition with catalytic hydrogenation resulted in the formation of the free amine **1** as

well as other impurities. Purification proved difficult even after reverse phase silica gel chromatography as degredation was detected. Isolation of pure **1** was never achieved, and had been intended to act as a standard to measure affinity gained from base stacking interactions since the heterocycle lacked exocyclic hydrogen-bonding functional groups.

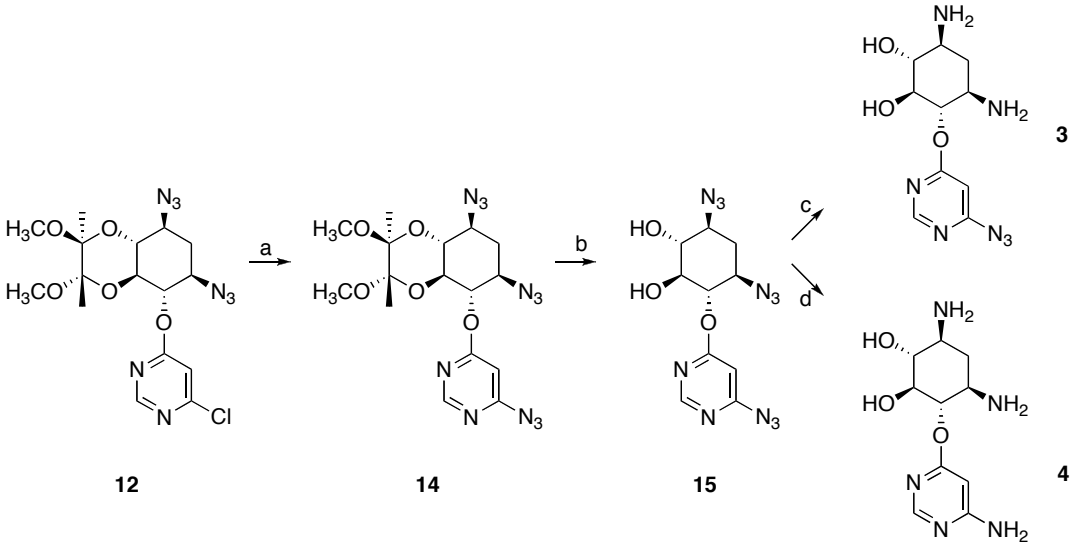
Other 2-DOS-heterocycle conjugate derivatives could be synthesized from the secondary alcohol **9** with a variety of commercially available chloropyrimidines. To obtain conjugate **2**, alcohol **9** was treated in a similar manner as before with NaH, then subsequently stirred with 4,6-dichloropyrimidine to afford **12** (Scheme 3.2). Removal of the Ley diketal proceeded under acidic conditions with TFA and CH_2Cl_2 to afford the diol **13**. Again, reduction of the azide proved difficult and **2** was never obtained in pure form.



Scheme 3.2. Synthesis towards **2**. *Reagents and conditions*. (a) NaH, 4,6-dichloropyrimidine, 0 °C to rt, 6 h, 88%; (b) TFA, rt, 3 h, 64%.

The electrophilic aryl chloride of conjugate **12** was an excellent precursor for reaction with a variety of nucleophiles. In one instance, the chloropyrimidine **12** was reacted with sodium azide to form **14** (Scheme 3.3). Hydrolysis of the Ley ketal under acidic conditions afforded the free alcohol **15** followed by selective azide reduction to

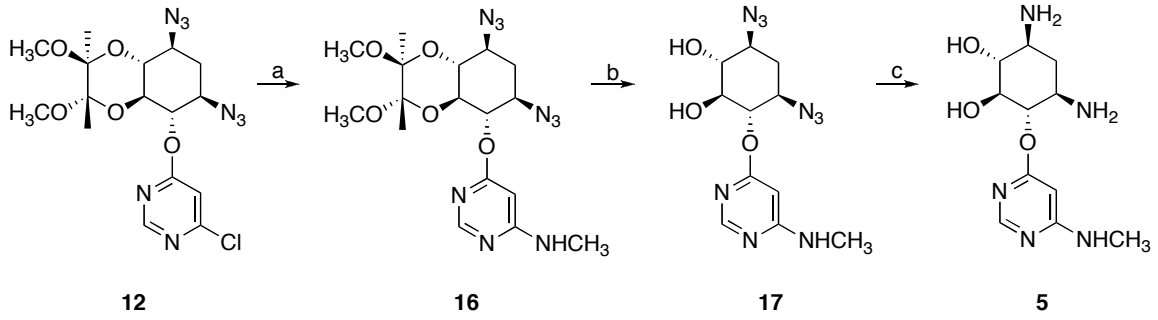
afford the free amine **3**. Reduction of the aryl azide was achieved by extended reaction times to obtain the aryl amine **4**. Additionally, nucleophilic aromatic substitution was performed on **12** with nucleophilic methylamine affording **16** (Scheme 3.4). Deprotection was again achieved by sequential acid hydrolysis to **17** and hydrogenation followed by chromatography to obtain **5**. Interesting to note is that clean reduction of the azide was achieved under hydrogenating conditions with conjugates possessing a functionalized pyrimidine.



Scheme 3.3. Synthesis of **3** and **4**. (a) NaN_3, DMF, 80 °C, 3 h, 96%; (b) TFA:H_2O (9:1), 0 °C to rt, 16 h, 7%; (c) 5% Pd/C, H_2 (1 atm), Et_3N, EtOH, rt, 1h, 25%; (d) 5% Pd/C, H_2 (1 atm), several h, Et_3N, EtOH, 54%.

Preliminary affinity studies with an established fluorescent assay using a 2-aminopurine labeled A-site RNA construct showed no binding with **3** and **4**.[64] Aggregation was only observed in the fluorescence emission spectra at high ligand concentrations. Conceivably, the two-ringed conjugates are too small to elicit substantial hydrogen-bonding contacts and stacking interactions within the A-site

binding pocket. Additionally, binding may occur but does not elicit the necessary structural perturbations to result in a fluorescence intensity change with the assay. Mimics with a larger scaffold may be necessary for RNA recognition, as neamine is the absolute smallest molecule known to selectively bind the A-site.



Scheme 3.4. Synthesis of **5**. *Reagents and conditions*. (a) 2 M CH_3NH_2 in MeOH, 60 °C, 12 h, 78%; (b) TFA, r.t., 3 h, 97%; (c) 5% Pd/C, H_2 (1 atm), EtOH, r.t., 18 h, 48%.

3.5 DESIGN AND SYNTHESIS OF 2-DEOXYSTREPTAMINE HOMOLOGUES

Along with mimicking ring I of paromomycin, synthetic efforts were commenced to mimic the 2-DOS ring found prevalently in most aminoglycosides. Aforementioned, Hermann and colleagues have synthesized numerous small heterocyclic libraries to mimic the 2-DOS ring. To understand the relationship between structure and binding, we proposed synthesizing a 2-DOS homologue with a primary alcohol at the 6-position. The primary alcohol would be conjugated to a triazine heterocycle, a scaffold which was previously shown to have some RNA recognition properties.[34,35] The extension of the 2-DOS core may provide greater understanding of the spatial requirements for RNA recognition when compared to the

natural counterpart. Moreover, since the previous neamine mimics were hypothesized to be too small for effective hydrogen bonding in RNA recognition, we proposed two 2-DOS homologues should be coupled to one triazine ring to potentially increase hydrogen-bonding and base stacking interactions (Figure 3.17).

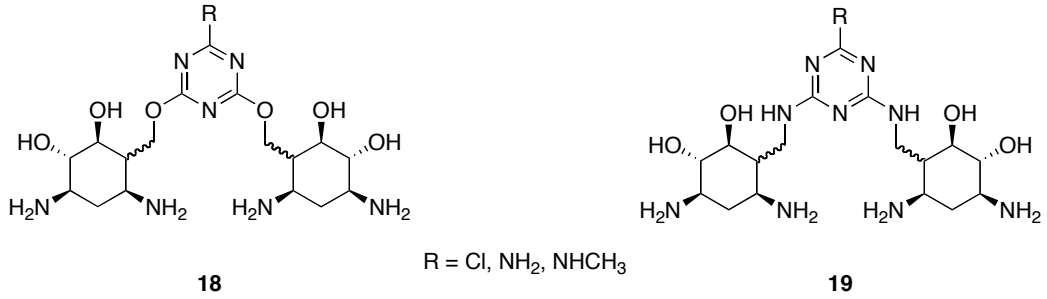
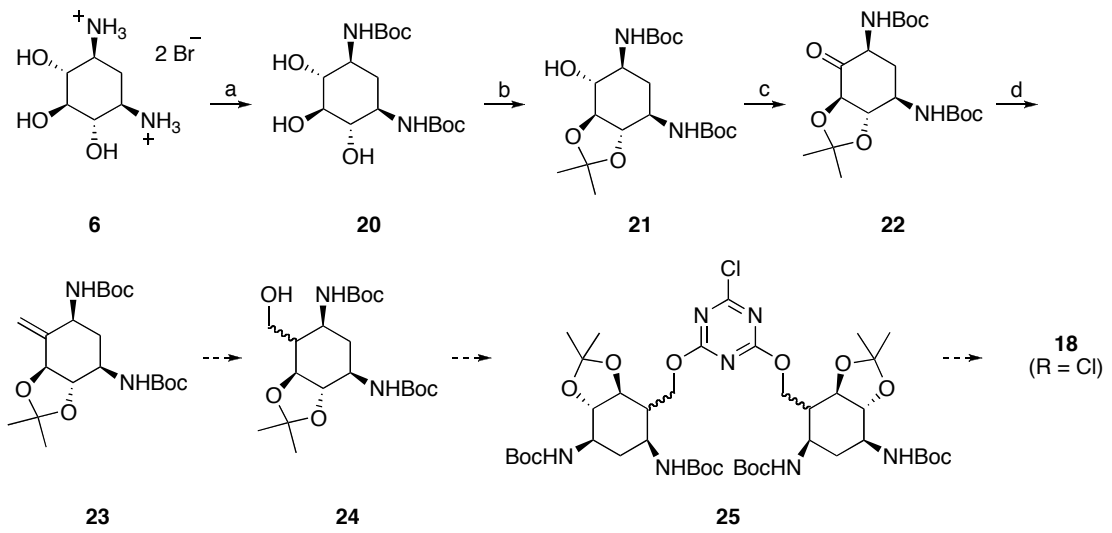


Figure 3.17. Triazine–2-DOS homologue conjugates (only one enantiomer shown).

The synthesis of the DOS homologue also commenced from the hydrogen bromide salt of 2-DOS **6**, similar to the neamine mimics (Scheme 3.5). In a water and dioxane mixture, the 2-DOS salt was stirred with triethylamine and di-tertbutyldicarbonate (Boc anhydride) overnight to form the white solid **20** in moderate yield. (A synthetic route using a *N*-benzyloxycarbonyl-protected 2-DOS was also designed but abandoned due to complications with solubility from the carbamate groups.) In contrast to a diketal, an enantiomeric mixture of acetanide **21** was formed by reaction with 2,2-dimethoxypropane and catalytic (±)-CSA to protect the two alcohols. Swern oxidation of the secondary alcohol to the ketone **22** proceeded in modest yield, followed by a Wittig reaction with methyl(triphenylphosphonium) bromide to afford the terminal alkene **23**.



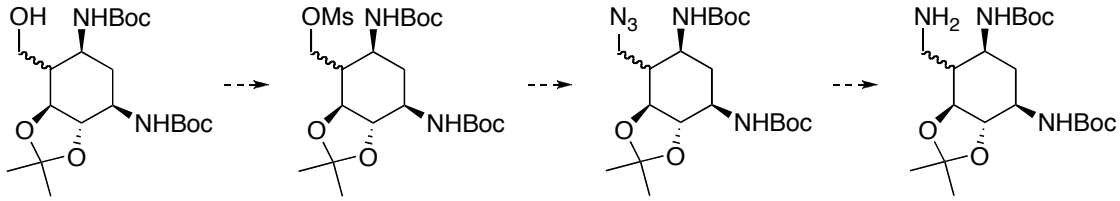
Scheme 3.5. Synthesis towards **18**. (a) Boc_2O, Et_3N, H_2O/dioxane, 60 °C, 2 h, 59%; (b) 2,2-dimethoxypropane, (±)-CSA, rt, 30 min, 67%; (c) *i.* $(COCl)_2$, DMSO, –78 °C, *ii.* Et_3N, rt, 16 h, 46%; (d) CH_3PPh_3Br, KO*t*Bu, reflux, 4 h, 70%.

Unfortunately, attempts to proceed further from this intermediate have currently stalled. Original synthetic plans were designed to hydrolyze the alkene to afford the primary alcohol; however all standard hydroboration conditions have failed (Scheme 3.5). No reaction occurred with the use of 9-borabicyclo[3.3.1]nonane (9-BBN);[65] separately, starting material consumption was only observed with BH_3•THF as a hydroborylating agent.[66] This suggests sterics may play a role in reactivity. Subsequent oxidation during hydroboration with NaOH and H_2O_2, however, resulted in product degradation. Repeated attempts to isolate a product have failed with only ESI-MS suggesting the possible existence of the primary alcohol **24**.

In the last segment of the proposed synthetic scheme, the primary alcohol would be treated under alkaline conditions to react with a chlorotriazine to form the aryl ether **25** (Scheme 3.5). Additional substitutions around the triazine core could be

achieved by increasing the temperature and reactant ratio. Regardless of the number of substitutions, universal deprotection could be accomplished through mild acid hydrolysis of the acetanide and Boc groups to afford the free amine.

An additional variation of the proposed scheme was to transform the primary alcohol to an amine in order to increase the nucleophilicity of the 2-DOS homologue (Scheme 3.6). Mesylation of **24** to form **26** followed by reaction with nucleophilic sodium azide affording **27**. The azide would then be reduced under standard Staudinger or catalytic hydrogenation conditions to afford the free amine **28**. The primary amine could then be treated in a similar fashion as the primary alcohol to be coupled to the triazine core.



24 **26** **27** **28**

Scheme 3.6. Synthesis towards **28**.

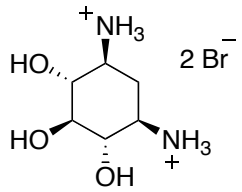
3.6 CONCLUSIONS

The synthesis of small molecules that mimic the action of antibiotics serves two purposes: a) to discover new scaffolds that are not recognized by current resistance mechanisms; and b) to understand RNA–ligand interactions for SAR development to potentially influence the design of future high affinity ligands. As more drug resistant bacteria appear, the need for new antibiotics to overcome

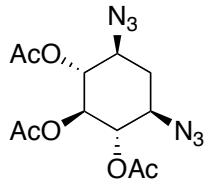
resistance mechanisms increases. Frustratingly, as new antibiotics are clinically administered, bacteria almost immediately acquire resistance. The development of antibiotics and bacterial resistance is a vicious cycle, but a greater understanding of the structural requirements for ligand recognition of RNA may give scientists an advantage.

Two successes are important to note in the development of new antibiotics. Recall from Chapter 1, in 2000, *Linezolid*™ was the first fully synthetic antibiotic introduced in over 50 years. Also, platensimycin, a recently discovered antibiotic, has a molecular scaffold not seen in current antibiotics.[9] Currently, combination therapies are being employed, similar to antiretroviral drug regiments, where antibiotics with different modes of action are administered to patients with drug resistant pathogens.[7] Conceivably, attacking two or more biosynthetic pathways can still adversely affect resistant pathogens with the current drug supplies. All are examples of ways to cope with an ever growing problem, but the current fear is additional alternatives may not be available in the future.

3.7 EXPERIMENTAL



2-deoxystreptamine dihydrogen bromide (6).[61] Neomycin B sulfate (50 g, 55.0 mmol) was dissolved in 48% HBr (300 mL). The reaction was slowly heated to reflux over several hours (~ 160 ºC) as the yellow solution turned black. The black solution was left to reflux for 18 h overnight. After 18 h, the reaction was concentrated to a thick, black residue. The residue was dissolved in H_2O (approx. 250 mL) and mixed with activated, decolorizing charcoal (5 g) and stirred at 60 ºC for 3 h. After 3 h, the reaction was filtered and the resulting black aqueous solution was again decolorized with activated charcoal (5 g) while stirring in an oil bath at 60 ºC. After an additional 3 h, the solution was filtered to afford a clear aqueous solution. The aqueous solution was concentrated to a small volume (~ 10–20 mL) to which copious amounts of MeOH were added causing precipitation of a white solid. The solid was filtered and repeatedly washed with MeOH. Small amounts of ether were added to assist with drying and removal of organics to give the product (11.1 g, 63%). ^{1}H NMR (400 MHz, D_2O): δ 3.44 (t, J = 9.6 Hz, 2H), 3.30 (t, J = 9.2 Hz, 1H), 3.21 (ddd, J_1 = 4.4 Hz, J_2 = 12.8 Hz, J_3 = 12.8 Hz, 2H), 2.36 (ddd, J_1 = 4.4 Hz, J_2 = 4.4 Hz, J_3 = 12.4 Hz, 1H), 1.72 (q, J = 12.8 Hz, 1H). ^{13}C NMR (100 MHz, D_2O): δ 74.9, 72.5, 50.3, 28.3. MS (APCI) was calculated for $C_6H_{13}N_2O_3$ $[M+H]^+$ 163.10, found 163.01.



2-Deoxy-1,3-diazido-4,5,6-tri-*O*-acetylstreptamine (7).[42,62]

(a) *Trifluoromethanesulfonyl (triflyl) azide solution.* The triflic azide solution was made by dissolving sodium azide (1560 mg, 24.0 mmol, 12 equivalents to 2-DOS, 6 equivalents per amine), in 5 mL H_2O (solubility of NaN_3 is 0.4 g/mL H_2O). CH_2Cl_2 (5 mL) was added and the biphasic mixture was stirred vigorously and lowered into an ice bath. Trifluoromethanesulfonic anhydride (1.98 mL, 12.0 mmol, 6 equiv) was slowly added. The reaction was stoppered and stirred at 0 ºC for 2 h. Saturated sodium bicarbonate solution was slowly added while stirring until CO_2 evolution ceased. Reaction was transferred to separatory funnel and organic layer separated. The aqueous layer was washed 2 × 12.5 mL CH_2Cl_2 (total volume of CH_2Cl_2 used for reaction and extraction equaled 5× the volume of H_2O used). Organic layers were combined and washed once more with saturated $NaHCO_3$ solution (10 mL). Organic solution used as is without further purification.

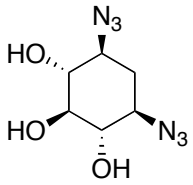
(b) *Substrate & Catalyst Solution.* 2-deoxystreptamine dihydrogen bromide salt (648 mg, 2.0 mmol) and $ZnCl_2$ (5.5 mg, 2.0 mol%) was dissolved in 25 mL H_2O (equal volume of CH_2Cl_2 used for TfN_3 solution). Triethylamine (1.7 mL, 12.0 mmol, 6 equiv) was slowly dripped in as the solution turned white. Methanol (80 mL) was added for homogeneity to preserve a 3:10:3 (H_2O/MeOH/CH_2Cl_2) solvent ratio.

Finally, the TfN_3 solution (approx. 25 mL) was slowly added over several minutes. The biphasic mixture was stirred vigorously at r. t. for 3 h.

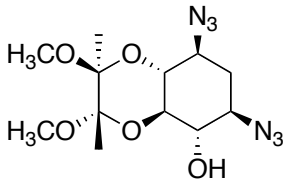
The reaction was quenched with 1N NaOH (30 mL) and organic solvents removed by vacuum. The organic phase was extracted 3 × 75 mL EtOAc. Organic layers were combined, dried over Na_2SO_4 and concentrated down to white solid (605 mg).

The crude (605 mg, assumed 2.8 mmol) was dissolved in dry pyridine (8 mL) in a stoppered, oven-dried flask. The flask was purged with Ar and lowered into ice bath. Acetic anhydride (1.2 mL, 12.7 mmol, 4.5 equiv), was added by syringe under Ar. Finally, 4-(dimethylamino)pyridine (17.3 mg, 0.14 mmol, 5 mol%) was added. The flask was purged with Ar and let warm to r. t. while stirring over 3 h. Upon completion, the reaction was quenched with MeOH (approx. 30 mL) and concentrated to a pink residue. The residue was dissolved in EtOAc (75 mL) and extracted 1 × 30 mL 1N HCl, 1 × 30 mL saturated solution $NaHCO_3$, 1 × 30 mL brine. Organic phases combined, dried over Na_2SO_4 and concentrated to an off-white solid. Further removal of triflic amide byproduct was performed by dissolving the solid in EtOAc (approx 40 mL) and extracting 2 × 20 mL 1N NaOH and 1 × 20 mL brine. Organic layers were again combined, dried over Na_2SO_4, and concentrated to a white solid. The crude was chromatographed by silica with a gradient of 15% to 30% EtOAc in hexanes to affor a white solid (510 mg, 75%). R_f 0.2 (20% EtOAc/hexane). ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.05 – 5.02 (m, 3H), 3.60 – 3.53 (m, 2H), 2.30 (ddd, J_1 = 13.6 Hz, J_2 = 4.4 Hz, J_3 = 4.4 Hz), 2.06 (s, 6H), 1.97 (s, 3H), 1.55 (ddd, J_1 = 12.8 Hz, J_2 = 12.8 Hz, J_3 =

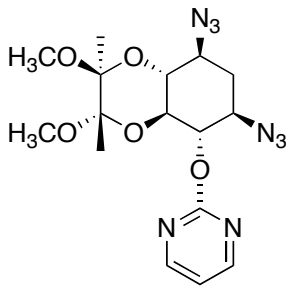
13.2 Hz). ^{13}C NMR (400MHz, $CDCl_3$): δ169.6, 169.2, 73.4, 71.4, 57.9, 31.8, 20.6, 20.5. MS (ESI) calculated for $C_{12}H_{16}N_6O_6$ $[M + Na]^+$ 363.10, found 362.98.



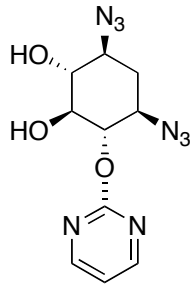
2-Deoxy-1,3-diazidostreptamine (8). 2-Deoxy-1,3-diazido-4,5,6-tri-*O*-acetyl-streptamine (6.13 g, 18.0 mmol) was dissolved in dry MeOH (80 mL) while stirring under Ar and lowered into an ice bath. Finally, a catalytic amount of sodium methoxide (292 mg, 5.4 mmol, 10 mol% per acetate) was added to the stirring solution. The reaction was left to stir for 12 h and followed by thin layer chromatography. After 12 h, the reaction was neutralized with DOWEX 50WX4-100 (H^+) ion-exchange resin. The solution was concentrated under reduced pressure to give an off-white product (3.67 g, 95%): R_f 0.2 (50% ethyl acetate/hexanes). ^{1}H NMR (400 MHz, CD_3OD): δ3.43 – 3.37 (m, 2H), 3.28 – 3.24 (m, 3H), 2.13 (ddd, J_1 = 12.8 Hz, J_2 = 4.4 Hz, J_3 = 4.4 Hz, 1H), 1.27 (q, J = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, CD_3OD): δ77.7, 77.0, 62.3, 33.5. Spectroscopic data matches values reported in literature.[62]



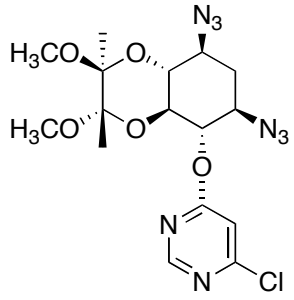
Synthesis of **9**.[63] 2-Deoxy-1,3-diazidostreptamine (3.64 g, 17.0 mmol), trimethyl orthoformate (6.14 mL, 56.1 mmol, 3.3 equiv.), 2,3-butanedione (1.64 mL, 18.7 mmol, 1.1 equiv.), and (±)-camphorsulfonic acid (434 mg, 1.87 mmol, 11 mol%) were added to dry MeOH (20 mL) in an oven-dried flask under argon. The solution was heated to reflux in an oil bath and let stir for 36 h. After following by chromatography, the reaction was quenched with triethylamine (5 mL). Solvents were removed under reduced pressure to afford a brown oily residue. The crude was purified without extraction by silica gel flash chromatography (20% ethyl acetate:hexanes) to afford a light, white solid (5.12 g, 96% yield). R_f 0.15 (20% ethyl acetate:hexanes). ^{1}H NMR (400 MHz, CDCl3): δ3.50 (m, 4H), 3.38 (m, 1H), 3.31 (s, 3H), 3.28 (s, 3H), 2.80 (broad s, 1H), 2.14 (ddd, J_1 = 16 Hz, J_2 = 4.4 Hz, J_3 = 4.4 Hz, 1H), 1.33 (s, 3H), 1.32 (s, 3H), 1.29 (q, J = 13.2 Hz, 1H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 99.7, 99.6, 73.5, 72.1, 71.0, 60.3, 56.9, 48.1, 48.0, 32.5, 17.5, 17.4. MS (APCI) was calculated for $C_{12}H_{20}N_6O_5$ $[M–N_2]^-$ = 300.14, found 300.90.



Synthesis of **10**. Compound **9** (75 mg, 0.24 mmol) was dissolved in dry DMF (5 mL) under argon in oven-dried glassware. Next, the flask was lowered into a –78 ºC dry ice bath. Sodium hydride (6.3 mg, 0.26 mmol) was added and stirred until the evolution of H_2 gas ceased. Finally, 2-chloropyrimidine (30.3 mg, .26 mmol) was added and the reaction was allowed to warm to ambient temperature while stirring over 6 h as the reaction was followed by thin layer chromatography. The reaction was quenched by addition of MeOH (10 mL) and concentrated to a clear oil under reduced pressure. The residue was redissolved in EtOAc (20 mL) and extracted 1 × 10 mL 1N HCl, 1 × 10 mL saturated solution $NaHCO_3$, and 1 × 10 mL brine. The organic layers were combined, dried over Na_2SO_4 and concentrated under reduced pressure. The crude oil was purified by flash chromatography (1:3 EtOAc:hexanes) to afford a white solid (58.4 mg, 60%): R_f 0.55 (50% EtOAc/hexanes). ^{1}H NMR (300 MHz, $CDCl_3$): δ8.50 (d, J = 4.5 Hz, 2H), 6.94 (t, J = 4.8 Hz, 1H), 5.51 (t, J = 9.9 Hz, 1H), 3.83 (t, J = 9.6, 1H), 3.70 (t, J = 9.9, 1H), 3.61 (m, 2H), 3.29 (s, 3H), 3.17 (s, 3H), 2.22 (ddd, J_1 = 13.2, J_2 = 4.8, J_3 = 4.8 Hz, 1H), δ1.45 (q, J = 12.3, 1H), δ1.28 (s, 3H), δ1.10 (s, 3H). ^{13}C NMR (75 MHz, $CDCl_3$): δ164.8, 159.0, 115.4, 99.6, 99.3, 75.6, 72.4, 69.7, 59.2, 56.6, 48.1, 48.0, 47.9, 32.5, 17.5, 17.4. MS (ESI) calculated for $C_{16}H_{22}N_8O_5$ [M + H^+] 407.17, found 406.92.

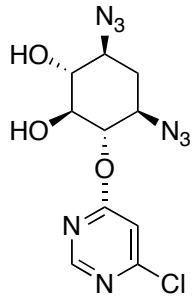


Synthesis of **11**. To neat TFA (7 mL) was added **10** (437 mg, 1.1 mmol). The reaction was stirred at r. t. overnight. After 18 h, the acid was removed under reduced pressure to afford an yellow oil and subsequently purified by silica gel flash chromatography (50% – 75% EtOAc/hexanes) to afford a white solid (288 mg, 91%). ^{1}H NMR (400 MHz, CD_3OD): δ 8.56 (d, J = 4.8 Hz, 1H), 7.11 (t, J = 4.8 Hz, 2H), 5.37 (t, J = 10.0 Hz, 1H), 3.78 – 3.71 (m, 1H), 3.57 (t, J = 9.2 Hz, 1H), 3.52 (m, 1H), 3.39 (t, J = 9.2 Hz, 1H), 2.21 (dt, J_1 = 13.2 Hz, J_2 = 4.4 Hz, 1H), 1.43 (q, J = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, CD_3OD): δ 165.5, 159.6, 115.9, 79.5, 76.8, 74.4, 61.0, 59.4, 32.2. MS (ESI) calculated for $C_{10}H_{12}N_8O_3$ $[M+H]^+$ 293.10, found 292.99.

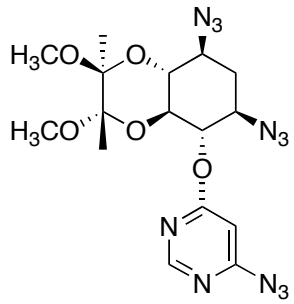


Synthesis of **12**. Compound **9** (624 mg, 2.0 mmol) was dissolved in dry DMF (7 mL) and stirred in an oven-dried flask under Ar. The flask was lowered into an ice bath

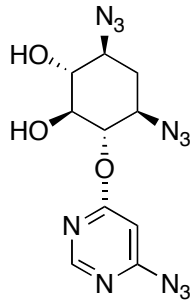
and sodium hydride (53 mg, 2.2 mmol) was carefully added under Ar. After the evolution of H_2 gas ceased, 4,6-dichloropyrimidine (328 mg, 2.2 mmol) was added. The flask was purged with Ar for 5 min and allowed to stir at r. t. for 6 h. After following the reaction for 6 h by thin layer chromatography, the reaction was quenched with MeOH (25 mL). Solvent was removed under reduced pressure. The resulting orange residue was dissolved in EtOAc (100 mL) and extracted 1 × 50 mL 1N HCl, 1 × 50 mL saturated solution $NaHCO_3$, 1 × 50 mL brine. Organic layers were combined, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by silica gel flash chromatography (15% ethyl acetate:hexanes) to afford a white solid product (776 mg, 88% yield). R_f 0.40 (20% ethyl acetate:hexanes). ^{1}H (400 MHz, $CDCl_3$), mixture of diastereomers: δ8.59 (s, 1H), 6.79 (s, 1H), 5.58 (t, J = 9.6 Hz, 1H), 3.77 (t, J = 9.6 Hz, 1H), 3.71 (t, J = 10 Hz, 1H), 3.57 (m, 2H), 3.23 (s, 3H), 3.17 (s, 3H), 2.26 (ddd, J_1 = 13.6, J_2 = 4.4, J_3 = 4.4 Hz, 1H), 1.48 (q, J = 13.2 Hz, 1H), 1.32 (s, 3H), 1.16 (s, 3H). ^{13}C (100 MHz, $CDCl_3$): δ170.1, 161.4, 158.0, 107.5, 99.9, 99.7, 75.4, 72.4, 69.5, 59.2, 56.6, 48.1, 47.8, 32.4, 17.4, 17.3. MS (ESI) calculated for $C_{16}H_{18}ClN_8O_5$ [M + H^+] 441.13, found 440.92.



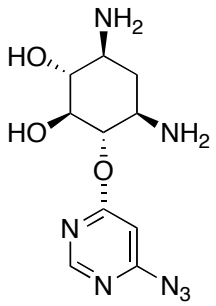
Synthesis of **13**. Compound **12** (330 mg, 0.75 mmol) was dissolved in TFA (5 mL) at r.t. and stirred for 3 h. After the reaction was complete, the acid was removed *in vacuo* to afford a yellow oil. The crude was dissolved in EtOAc (50 mL) and washed 2 × 30 mL saturated solution of $NaHCO_3$ and 1 × 20 mL brine. The organic phase was dried over Na_2SO_4 and concentrated under reduced pressure to afford a yellow residue which was purified by silica gel flash chromatography (50% EtOAc/hexanes) to afford a white, amorphous solid (156 mg, 64%). ^{1}H NMR (400 MHz, CD_3OD): δ 8.57 (s, 1H), 7.02 (s, 1H), 5.45 (t, J = 9.6 Hz, 1H), 3.74 (m, 1H), 3.54 (t, J = 9.2, 1H), 3.47 (m, 1H), 3.40 (t, J = 9.2 Hz, 1H), 2.23 (ddd, J_1 = 13.2, J_2 = 4.4 Hz, J_3 = 4.4 Hz, 1H), 1.45 (q, J = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, CD_3OD): δ 171.4, 161.3, 158.1, 108.2, 79.6, 76.7, 74.3, 61.0, 59.2, 32.1. MS (ESI) calculated for $C_{10}H_{11}ClN_8O_3$ $[M+H]^+$ 327.06, found 326.94.



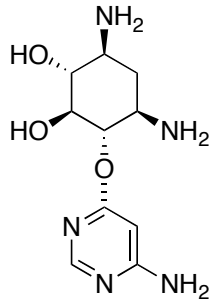
Synthesis of **14**. To DMF (8 mL) were added **12** (440 mg, 1.0 mmol) and sodium azide (325 mg, 5.0 mmol). The mixture was heated to 80 °C and stirred over several hours (approximately 3 h). After following the reaction by chromatography, solvents were removed under reduced pressure. The orange residue was dissolved in EtOAc (100 mL) and washed 3 × 50 mL with brine. Organic phases were combined and dried over Na_2SO_4. EtoAc was removed *in vacuo* to afford an orange solid (432 mg, 96%). R_f 0.45 (20% EtOAc/hexanes). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.50 (s, 1H), 6.14 (s, 1H), 5.51 (t, J = 9.6 Hz, 1H), 3.68 (m, 2H), 3.51 (m, 2H), 3.26 (s, 3H), 3.10 (s, 3H), 2.18 (ddd, J_1 = 13.5 Hz, J_2 = 4.5, J_3 = 4.5 Hz, 1H), 1.41 (q, J = 12.9 Hz, 1H), 1.25 (s, 3H), 1.10 (s, 3H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 170.3, 163.0, 157.6, 99.6, 99.4, 95.1, 74.9, 72.3, 69.5, 59.1, 56.6, 47.9, 47.7, 32.3, 17.4, 17.3. MS (ESI) calculated for $C_{16}H_{21}N_{11}O_5$ $[M+H]^+$ 448.17, found 447.87.



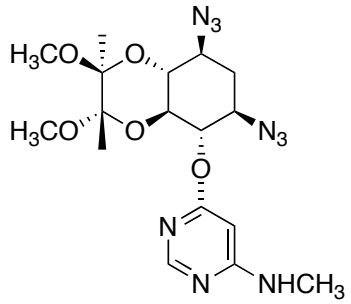
Synthesis of **15**. Compound **14** (335.6 mg, 0.75 mmol), was dissolved in a 0 °C TFA:H_2O (9:1) solution (10 mL). Reaction was allowed to stir overnight at r. t. After following the reaction by thin layer chromatography, solvent was removed under reduced pressure to afford a yellow oil. The crude was purified by silica gel chromatography (40% EtOAc/hexanes) to afford a white product (108 mg, 43% yield). R_f 0.25 (40% EtOAc:hexanes). A side product was also isolated and determined to be the azide-protected 2-deoxystretpamine moiety **8** which is the result of acidic hydrolysis of the pyrimidine ring (18 mg, 7% yield). ^{1}H NMR (400 MHz, CD_3OD): δ 8.45 (s, 1H), 6.26 (s, 1H), 5.31 (t, J = 9.6 Hz, 1H), 3.65 – 3.59 (m, 1H), 3.45 (t, J = 9.2 Hz, 1H), 3.42 – 3.30 (m, 2H), 2.12 (ddd, J = 12.8, 4.0, 4.0 Hz, 1H), 1.36 (q, J = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, CD_3OD): δ172.5, 164.6, 158.9, 96.7, 80.0, 77.6, 75.3, 61.9, 60.3, 33.0. MS (ESI) was calculated for $C_{10}H_{11}N_{11}O_3$ $[M + H]^+$ 334.10, found 334.09.



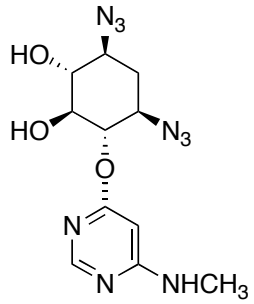
Synthesis of **3**. Compound **15** (76 mg, .23 mmol) was dissolved in EtOH (5 mL) and Et_3N (70 μL, 2.2 equiv.) while stirring under Ar. 5% Palladium on carbon (approximately 40 mg) was added to the reaction flask and purged with Ar for 5 min. The flask was then flushed with H_2 gas for an additional 5 min and then placed under a H_2 atmosphere (with balloon, approximately 1 atm). Solution was allowed to stir at r. t. for 1 h. After 1 h and disappearance of starting material, the reaction was filtered through a pad of Celite and washed with copious amounts of H_2O and MeOH. Removal of solvent under reduced pressure afforded a white solid. The crude was purified using C_{18} reverse phase flash chromatography conditions (2% saturated NH_4OH, 10% H_2O, 10% MeOH, 78% CH_3CN) resulting in isolation of a yellow solid. Addition of methanol caused precipitation of impurities which were filtered through glass wool with water. Removal of water under reduced pressure afforded a white solid (16 mg, 25% yield). ^{1}H NMR (400 MHz, D_2O): δ7.98 (s, 1H), 5.91 (s, 1H), 4.78 (t, J = 10 Hz, 1H), 3.67 – 3.61 (m, 1H), 3.44 (t, J = 8 Hz, 1H), 3.12 (t, J = 9.6 Hz, 1H), 2.70 – 2.63 (m, 1H), 2.05 (ddd, J_1 = 12.8, J_2 = 4.4, J_3 = 4.4 Hz, 1H), 1.27 (q, J = 12.8 Hz, 1H). ^{13}C NMR (100 MHz, D_2O): δ 169.7, 165.6, 157.6, 87.5, 79.4, 76.4, 74.0, 59.6, 50.0, 33.0. MS (ESI) calculated for $C_{10}H_{15}N_7O_3$ $[M + H]^+$ 282.12, found 281.94.



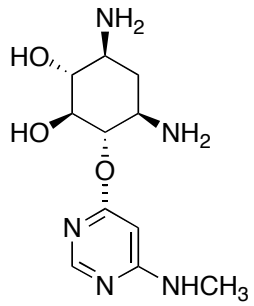
Synthesis of **4**. Compound **15** (97 mg, .29 mmol) was dissolved in EtOH (7 mL) and Et_3N (89 μL, .64 mmol). The flask was stoppered and purged with Ar. Quickly, 5% palladium on carbon (~ 30 mg) was added to the flask which was purged further with Ar. Finally, the flask was evacuated as H_2 gas was streamed through with a balloon. The flask was then placed under a H_2 atomosphere with a balloon and let stir at r. t. for several hours. After following the reaction by chromatography, the reaction was filtered through a pad of Celite 500 with copious amounts of water. The water was removed under reduced pressure to afford an off white solid which was further purified by HPLC (76 mg of TFA salt, 54% yield). ^{1}H NMR (400 MHz, D_2O): δ 7.97 (s 1H), 5.89 (s, 1H), 4.62 (t, *J* = 9.2 Hz, 1H), 3.41 (t, J = 9.2 Hz, 1H), 3.21 (t, *J* = 8.8 Hz, 1H), δ 2.91 (ddd, *J1* = 10.8, J_2 = 10.8, J_3 = 4.0 Hz, 1H), 2.78 (ddd, J_1 = 10.8, J_2 =10.8, J_3 = 4.0 Hz, 1H), 1.98 (ddd, J_1 = 12.8, J_2 = 4.4, J_3 = 4.4 Hz, 1H), δ1.28 (q, *J* = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, D_2O): δ170.4, 165.6, 157.7, 87.5, 81.5, 76.5, 74.6, 50.7, 49.7, 35.1. MS (ESI) calculated for $C_{10}H_{17}N_5O_3$ $[M+H]^+$ = 256.13, found 255.98.



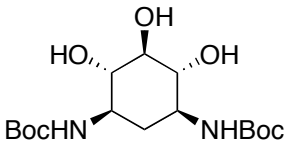
Synthesis of **16**. Compound **12** (660 mg, 1.5 mmol) was dissolved in a 2.0 M solution of methylamine in MeOH (7 mL) in a pressure tube. The vessel was heated to 60 ºC and let stir for 12 h. Upon the disappearance of starting material, the solvent was removed under reduced pressure. The white residue was dissolved in EtOAc (75 mL) and extracted 1 × 30 mL 1N HCl, 1 × 30 mL saturated solution $NaHCO_3$, and 1 × 30 mL brine. The organic layer was dried over Na_2SO_4 and the EtOAc was removed under reduced pressure to afford a white solid (511 mg, 78% yield). R_f 0.20 (35% ethyl acetate:hexanes). ^{1}H NMR (400 MHz, $CDCl_3$): δ8.19 (s, 1H), 5.62 (s, 1H), 5.60 (broad s, 1H), 5.45 (t, J = 9.6 Hz, 1H), 3.70 (m, 2H), 3.54 (m, 2H), 3.28 (s, 3H), 3.16 (s, 3H), 2.83 (d, J = 5.2 Hz, 3H), 2.18 (ddd, J_1 = 13.6 Hz, J_2 = 4.4 Hz, J_3 = 4.4 Hz, 1H), 1.42 (q, J = 12.8 Hz, 1H), 1.28 (s, 3H), 1.15 (s, 3H). ^{13}C NMR (100 MHz, $CDCl_3$): δ169.7, 165.2, 157.2, 99.7, 99.5, 84.2, 74.0, 72.4, 69.7, 59.5, 56.7, 47.9, 47.7, 32.4, 28.3, 17.4, 17.3. MS (ESI) calculated for $C_{17}H_{25}N_9O_5$ $[M+H]^+$ 436.20, found 435.99.



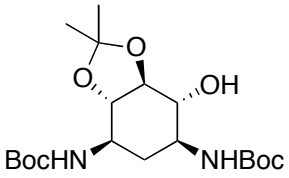
Synthesis of **17**. Compound **16** (298 mg, 0.69 mmol) was dissolved in TFA (6 mL) and stirred at r. t. for 3 h. After following the reaction by chromatography, the acid was removed under reduced pressure to afford a yellow oil that was subsequently purified by silica gel flash chromatography (80% EtOAc/hexanes) to afford a white solid (215 mg, 97%), Rf = 0.33 (100% EtOAc). ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.15 (s, 1H), 7.36 (s, 1H), 5.76 (s, 1H), 5.80 – 4.80 (broad s, 2H), 5.15 (t, J = 9.2 Hz, 1H), 3.72 (m, 1H), 3.48 (m, 1H), 3.35 (t, J = 9.2 Hz, 1H), 3.29 (t, J = 9.2 Hz, 1H), 2.75 (s, 3H), 2.04 (ddd, J_1 = 8.0 Hz, J_2 = 4.0 Hz, J_3 = 4.0 Hz, 1H), 1.37 (q. J = 12.0 Hz, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 164.6, 158.6, 158.3, 156.4, 77.9, 76.1, 73.8, 60.5, 59.0, 31.8, 27.9. MS (ESI) calculated for $C_{11}H_{15}N_9O_3$ $[M+H]^+$ 322.13, found 322.00.



Synthesis of **5**. A flask was charged with absolute EtOH (10 mL) and **17** (129 mg, 0.40 mmol), stoppered and flushed with Ar. Quickly, 5% Pd/C (approximately 30 mg) was added and the flask was again flushed with Ar for 5 min. The flask was next flushed with H_2 gas for an additional 5 min, and finally placed under a H_2 gas atmosphere using a balloon. The reaction was allowed to stir at r. t. overnight. After 18 h, the heterogeneous solution was filtered through a pad of cellulose and washed with copious amounts of EtOH. The resulting filtrate was concentrated to a brown residue and purified by analytical HPLC (gradient: 0% to 7.5% CH_3CN in H_2O over 15 min on a C_{18} preparatory column; flow rate 3.0 mL/min). The trifluoroacetate salt was isolated (96.0 mg, 48%). ^{1}H NMR (400 MHz, D_2O): δ 8.17 (s, 1H), 6.09 (s, 1H), 5.23 (t, J = 8.8 Hz, 1H), 3.58 – 3.48 (m, 1H), 3.52 (t, J = 9.2 Hz, 1H), 3.45 (t, J = 9.6 Hz, 3.20 – 3.14 (m, 1H), 2.36 – 2.31 (dt, J_1 = 12.8 Hz, J_2 = 4.0 Hz, 1H), 1.76 (q, J = 12.8 Hz, 1H). HRMS (ESI) calculated for $C_{11}H_{19}N_5O_3$ $[M+H]^+$ 270.1482, found 270.1479.

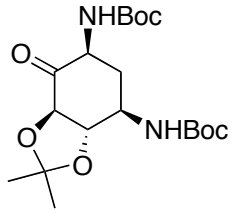


Synthesis of **20**. To H_2O (70 mL) was added 2-deoxystreptamine dihydrogen bromide salt **6** (4860 mg, 15.0 mmol) and Et_3N (12.5 mL, 90.0 mL). The opaque solution was stirred until clear. Dioxane (90 mL) was subsequently added (90 mL) followed by di-tert-butyl dicarbonate (8184 mg, 37.5 mL). The solution was stirred vigorously at 60 °C for 2 h. Afterwards, a concentrated NH_4OH solution (approximately 12 mL) was added and the reaction was allowed to continue stirring at 60 °C for an additional 30 min. The solution was then concentrated *in vacuo* to afford a while solid which was dissolved in hot MeOH (approximately 100 mL) and purified by recrystallization. Three iterative recrystallizations from hot MeOH and cooling to –20° C afforded a white powder (3235 mg, 59%). ^{1}H NMR (DMSO-d_6): δ 6.58 (d, J = 8.0 Hz, 2H), 4.80 (d, J = 4.0 Hz, 1H), 4.57 (d, J = 4.0 Hz, 2H), 3.16 – 3.04 (m, 2H), 3.02 – 2.90 (m, 3H), 1.80 – 1.70 (m, 1H), 1.36 (s, 18H), 1.07 (q, J = 12.4 Hz, 1H). Spectroscopic values matches those reported in literature.



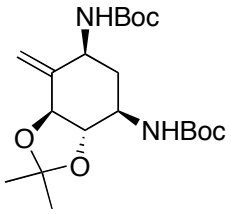
Synthesis of **21**. To anhydrous CH_3CN (150 mL) was added Compound **20** (1087 mg, 3.0 mmol), 2,2-dimethoxypropane (1.66 mL, 13.5 mmol), and (±)-camphorsulphonic

acid (70 mg, 0.3 mmol). The flask was flushed with Ar and stirred at r. t. for 30 min as the white suspension slowly dissolved and became a clear, yellow liquid. After 30 min, the reaction was complete and quenched with anhydrous Et_3N (approximately 0.75 mL). Upon addition of the base, the solution became colorless. Solvent was removed *in vacuo* to afford a white solid which was dissolved in CH_2Cl_2 (100 mL) and washed 1 × 100 mL 5% w/v citric acid solution, 1 × 100 mL saturated $NaHCO_3$ solution, and 1 × 100 mL brine. The organic phase was dried over Na_2SO_4 and concentrated to a white solid. Finally, the crude was taken up in *n*-hexanes and white precipitate was collected by vacuum filtration to afford a white powder (818 mg, 67% yield). ^{1}H NMR (400 MHz, DMSO-d_6): δ 6.97 (d, J = 8.4 Hz, 1H), 6.74 (d, J = 8.4 Hz, 1H), 5.09 (d, J = 5.6 Hz, 1H), 3.58 – 3.46 (m, 1H), 3.40 – 3.34 (m, 2H), 3.30 – 3.20 (m, 2H), 1.83 (dt, J_1 = 13.2 Hz, J_2 = 4.8 Hz, 1H), 1.36 (s, 18H), 1.32 (s, 3H), 1.31 (s, 3H), 1.13 (q, J = 12.4 Hz, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 155.3, 154.8, 109.6, 81.2, 77.9, 77.8, 77.6, 71.1, 52.7, 47.9, 36.7, 28.3, 28.2, 26.9, 26.8. MS (ESI) calculated for $C_{19}H_{34}N_2O_7$ $[M+H]^+$ 403.24, found 402.85.



Synthesis of **22**. To perform the Swern oxidation, oxalyl chloride (1.29 mL, 15.0 mmol) was added to anhydrous –78 °C CH_2Cl_2 (125 mL). A separate solution of

anhydrous DMSO (2.13 mL) in anhydrous CH_2Cl_2 (10.0 mL) was also cooled to –78 °C and slowly added to the first, while maintaining a temperature below –60 °C as gas slowly evolved. The mixture was allowed to stir at –78°C for 10 min. Afterwards, a third solution of CH_2Cl_2 (100 mL) cooled to –78°C with alcohol **21** (1207 mg, 3.0 mmol) was added to the reaction slowly while maintaining a temperature below –60 °C as gas also evolved. The solution was stirred for an additional 1.5 h at –78 °C. Finally, Et_3N that had been previously cooled to –78 °C was slowly added to the reaction, again maintaining a temperature below –60 °C. The reaction was allowed to stir an additional 1.5 h at –78 °C, then slowly warmed to r. t. overnight. After 18 h, the reaction was diluted with CH_2Cl_2 (50 mL) and washed 1 × 150 mL 5% w/v citric acid, 1 × 150 mL saturated solution of $NaHCO_3$, 1 × 150 mL brine. The organic phase was separated and dried over Na_2SO_4 and concentrated *in vacuo* to afford a yellow residue. The crude was purified by silica gel chromatography (3% MeOH/CH_2Cl_2) to afford a yellowish-orange solid (553 mg, 46%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.12 (d, J = 8.4 Hz, 1H), 6.96 (d, *J* = 8.4 Hz, 1H), 4.63 (d, *J* = 8.0 Hz, 1H), 4.41 – 4.38 (m, 1H), 4.15 – 4.00 (m, 1H), 3.51 (t, *J* = 10.4 Hz, 1H), 2.10 – 2.02 (m, 1H), 1.38 (s, 9H), 1.37 (s, 9H), 1.36 (s, 3H), 1.32 (s, 3H), 1.08 (app t, *J* = 6.8 Hz, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 199.4, 155.1, 154.8, 111.1, 79.7, 79.3, 78.2, 78.0, 54.4, 47.8, 35.8, 28.2, 28.1, 26.9, 26.3. MS (ESI) calculated for $C_{19}H_{32}N_2O_7$ $[M+H]^+$ 401.22, found 400.80.



Synthesis of **23**. *Ylide formation*: To anhydrous benzene (75 mL) was added methyltriphenylphosphonium bromide (2143 mg, 6.0 mmol) and potassium tert-butoxide (673 mg, 6.0 mmol). The mixture was heated to reflux for 2 h as the solution turned deep yellow in color.

Wittig reaction: After 2 h, the ketone **22** (801 mg, 2.0 mmol) was added to the ylide solution and the pale yellow reaction was returned to reflux for 1.5 h. The reaction was stopped and solvent removed *in vacuo* to afford a yellow solid. The crude was dissolved in CH_2Cl_2 (150 mL) and washed 1 × 150 mL 5% w/v citric acid, 1 × 150 mL saturated solution $NaHCO_3$, and 1 × 150 mL brine. The phases were separated, the organic layer was dried over Na_2SO_4, and concentrated down to a yellow solid. The crude was purified by silica gel chromatography (2% MeOH/CH_2Cl_2) to afford a white solid (565 mg, 70%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.10 (d, J = 8.8 Hz, 1H), 7.03 (d, J = 8.4 Hz, 1H), 4.85 (app s, 1H), 4.73 (app s, 1H), 4.10 – 4.02 (m 1H), 3.96 (d, J = 9.2 Hz, 1H), 3.78 – 3.68 (m, 1H), 3.21 (t, J = 9.6 Hz, 1H), 1.90 – 1.84 (m, 1H), 1.38 (s, 9H), 1.37 (s, 9H), 1.34 (s, 6H), 1.22 (q, J = 12.0 Hz, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 155.0, 154.8, 144.4, 110.3, 101.3, 81.0, 78.3, 77.9, 77.7, 49.0, 48.1, 30.7, 28.2 (2), 27.0, 26.8. MS (ESI) calculated for $C_{20}H_{34}N_2O_6$ $[M+H]^+$ 399.24, found 398.82.

3.8 REFERENCES

1. Davies, J. Bacteria on the rampage. *Nature* **1996**, *383*, 219–220.

2. *Staphylococcus aureus* with reduced susceptibility to vancomycin–United States. *MMWR Morb Mortal Wkly Rep* **1997**, *46*, 765–766.

3. *Staphylococcus aureus* resistant to vancomycin–United States, 2002. *MMWR Morb Mortal Wkly Rep* **2002**, *51*, 565–567.

4. Tsiodras, S., Gold, H. S., Sakoulas, G., Eliopoulous, G. M., Wennersten, C., Venkataraman, L., Moellering, R. C., Ferraro, M. J. Linezolid resistance in clinical isolate of *Staphylococcus aureus*. *Lancet* **2001**, *358*, 207–208.

5. Spellberg, B., Powers, J. H., Brass, E. P., Miller, L. G., Edwards, J. E. Trends in antimicrobial drug development: implication for the future. *Antimicrob. Res. Dev.* **2004**, *38*, 1279–1286.

6. Shlaes, D. M., Projan, S. J., Edwards, J. E. Antibiotic discovery: state of the state. *ASM News* **2004**, *70*, 275–281.

7. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **2000**, *406*, 775–781.

8. Leeb, M. Antibiotics: a shot in the arm. *Nature* **2004**, *431*, 892–893.

9. Wang, J., Soisson, S. M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Patiner, R., Parthasarathy, G., Tang, Y. S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya, H., Ondeyka, J., Herath, K., Zhang, C., Hernandez, L., Allocco, J., Basilio, A., Tormo, J. R., Genilloud, O., Vicente, F., Pelazez, F., Colwell, L., Lee, S. H., Michael, B., Felcetto, T., Gill, C., Silver, L. L., Hermes, J. D., Bartizal, K., Barrett, J., Schmatz, D., Becker, J. W., Cully, D., Singh, S. B. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **2006**, *441*, 358–361.

10. Carter, A. P., Clemons, W. M., Broderson, D. E., Morgan-Warren, R. J., Wimberly, B. T., Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340–348.

11. Davies, J., Davis, B. D. Misreading of ribonucleic acid code words induced by aminoglycoside antibiotics. *J. Biol. Chem.* **1968**, *243*, 3312–3316.

12. Fourmy, D., Recht, M., Blanchard, S., Puglisi, J. Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996**, *274*, 1367–1371.

13. Yusupova, G., Jenner, L., Rees, B., Moras, D., Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **2006**, *444*, 391–394.

14. Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*. 3rd ed.; Worth Publishers: New York, 2000.

15. Green, R., Noller, H. F. Ribosome and translation. *Annu. Rev. Biochem.* **1997**, *66*, 679–716.

16. (The Svedberg unit, S, is the sedimentation coefficient which is partially determined by hydrodynamic factors such as molecular weight, unit shape and size. Since sedimentation is dependent on nonlinear factors, their coefficients are not additive).

17. Crick, F. H. C. The genetic code: III. *Sci. Am.* **1966**, *215*, 55–62.

18. Nirenberg, M. W. The genetic code: II. *Sci. Am.* **1963**, *208*, 80–94.

19. Cedergren, R., Miramontes, P. The puzzling origin of the genetic code. *Trends Biochem. Sci.* **1996**, *21*, 199–200.

20. Moazed, D., Noller, H. F. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **1987**, *327*, 389–394.

21. Purohit, P., Stern, S. Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. *Nature* **1994**, *370*, 659–662.

22. Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D., Puglisi, J. D. RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J. Mol. Biol.* **1996**, *262*, 421–436.

23. Fourmy, D., Yoshizawa, S., Puglisi, J. D. Paromomycin binding induces a local conformational change in the A-site of 16S rRNA. *J. Mol. Biol.* **1998**, *277*, 333–345.

24. Yoshizawa, S., Fourmy, D., Puglisi, J. Recognition of the codon-anticodon helix by ribosomal RNA. *Science* **1999**, *285*, 1722–1725.

25. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Tarry, M. J., Carter, A. P., Ramakrishnan, C. V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001**, *292*, 897–902.

26. Umezawa, H., Hooper, I. R., *Aminoglycoside antibiotics*. Springer-Verlag: New York, 1982; Vol. 62, p 368.

27. Magnet, S., Blanchard, J. S. Molecular insights into aminoglycoside action and resistance. *Chem. Rev.* **2005**, *105*, 477–497.

28. Vakulenko, S. B., Mobashery, S. Versatility of aminoglycosides and prospects for their future. *Clin. Microbiol. Rev.* **2003**, *16*, 430–450.

29. François, B., Russell, R. J. M., Murry, J. B., Aboul-ela, F., Masquida, B., Vicens, Q., Westhof, E. Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.* **2005**, *33*, 5677–5690.

30. Sucheck, S. J., Wong, A. L., Koeller, K. M., Boehr, D. D., Draker, K., Sears, P., Wright, G. D., Wong, C.–H. Design of bifuunctional antibiotics that target bacterial rRNA and inhibit resistance-causing enzymes. *J. Am. Chem. Soc.* **2000**, *122*, 5230–5231.

31. Simonsen, K. B., Ayida, B. K., Vourloumis, D., Takahashi, M., Winters, G. C., Barluenga, S., Qamar, S. Shandrick, S., Zhao, Q., Hermann, T. Novel paromamine derivatives exploring shallow-groove recognition of ribosomal-decoding-site RNA. *ChemBioChem* **2002**, *3*, 1223–1228.

32. Simonsen, K. B., Ayida, B. K., Vourloumis, D., Winters, G. C., Takahashi, M., Shandrick, S., Zhao, Q., Hermann, T. Piperidine glycosides targeting the ribosomal decoding site. *ChemBioChem* **2003**, *4*.

33. Barluenga, S., Simonsen, K. B., Littlefield, E. S., Ayida, B. K., Vourloumis, D., Winters, G. C., Takahashi, M., Shandrick, S., Zhao, Q., Han, Q., Hermann, T. Rational design of azepane-glycoside antibiotics targeting the bacterial ribosome. *Bioorg. Med. Chem. Lett.* **2003**, *14*, 713–718.

34. Zhou, Y., Gregor, V. E., Sun, Z., Ayida, B. K., Winters, G. C., Murphy, D., Simonsen, K. B., Vourloumis, D., Fish, S., Froelich, J. M., Wall, D., Hermann, T. Structure-guided discovery of novel aminoglycoside mimetics as antibacterial translation inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49*, 4942–4949.

35. Zhou, Y., Sun, Z., Froelich, J. M., Hermann, T., Wall, D. Structure-activity relationships of novel antibacterial translation inhibitors: 3,5-diamino-piperidinyl triazines. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5451–5456.

36. Vourloumis, D., Takahashi, M., Winters, G. C., Simonsen, K. B., Ayida, B. K., Barluenga, S., Qamar, S., Shandrick, S., Zhao, Q., Hermann, T. Novel 2,5-dideoxystreptamine derivatives targeting the ribosomal decoding site RNA. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3367–3372.

37. Vourloumis, D., Winters, G. C., Takahasi, M., Simonsen, K. B., Auida, B. K., Shandrick, S., Zhao, Q., Hermann, T. Novel acyclic deoxystreptamine mimetics targeting the ribosomal decoding site. *ChemBioChem* **2003**, *4*, 879–885.

38. Hendrix, M., Alper, P. B., Priestley, E. S., Wong, C.–H. Hydroxyamines as a new motif for the molecular recognition of phosphodiesters: implication for aminoglycoside–RNA interactions. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 95–98.

39. Wong, C. H., Hendrix, M., Manning, D. D., Rosenbohm, C., Greenberg, W. A. A library approach to the discovery of small molecules that recognize RNA: use of a 1,3-hydroxyamine motif as core. *J. Am. Chem. Soc.* **1998**, *120*, 8319–8327.

40. Agnelli, F., Sucheck, S. J., Marby, K. A., Rabuka, D., Yao, S.–L., Sears, P. S., Liang, F.–S., Wong, C.–H. Dimeric aminoglycosides as antibiotics. *Angew. Chem. Int. Ed.* **2004**, *43*, 1562–1566.

41. Alper, P. B., Hendrix, M., Sears, P., Wong C.–H. Probing the specificity of aminoglycoside–ribosomal RNA interactions with designed synthetic analogs. *J. Am. Chem. Soc.* **1998**, *120*, 1965–1978.

42. Greenberg, W. A., Priestley, S., Sears, P. S., Alper, P. B., Rosenbohm, C., Hendrix, M., Hung, S.-C., Wong, C.-H. Design and synthesis of new aminoglycoside antibiotics containing neamine as an optimal core stucture: correlation of antibiotic activity with in vitro inhibition of translation. *J. Am. Chem. Soc.* **1999**, *121*, 6527–6541.

43. Kaul, M., Barbieri, C. M., Pilch, D. S. Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *J. Am. Chem. Soc.* **2006**, *128*, 1261–1271.

44. Liang, F.-S., Wang, S.-K., Nakatani, T., Wong, C.-H. Targeting RNAs with tobramycin analogues. *Angew. Chem. Int. Ed.* **2004**, *43*, 6496–6500.

45. Chou, C.-H., Wu, C.-S., Chen, C.-H., Lu, L.-D., Kulkarni, S. S., Wong, C.-H., Hung, S.-C. Regioselective glycosylation of neamine core: a facile entry to kanamycin B related analogues. *Org. Lett.* **2004**, *6*, 585–588.

46. Griffey, R. H., Hofstadler, S. A., Sannes-Lowery, K. A., Ecker, D. J., Crooke, S. T. Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10129–10133.

47. Ding, Y., Hofstadler, S. A., Swayze, E. E., Risen, L., Griffey, R. H. Design and synthesis of paromomycin-related heterocycle-substituted aminoglycoside mimetics based on a mass spectrometry RNA-binding assay. *Angew. Chem. Int. Ed.* **2003**, *42*, 3409–3412.

48. He, Y., Yang, J., Wu., B., Robinson, D., Sprankle, K., Kung P.–P., Lowery, K., Mohna, V., Hofstadler, S., Swayze, E. E., Griffey, R. Synthesis and evaluation of novel bacterial rRNA-binding benzimidaoles by mass spectromtery. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 695–699.

49. Wu, B., Yang, J., Robinson, D., Hofstadler, S., Griffey, R., Swayze, E. E., He, Yun Synthesis of linked carbohydrates and evaluation of their binding for 16S RNA by mass spectrometry. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3915–3918.

50. Seeberger, P. H., Baumann, M., Zhang, G., Kanemitsu, Swayze, E. E., Hofstadler, S. A., Griffey, R. H. Synthesis of neomycin analogs to investigate aminoglycoside–RNA interactions. *Synlett* **2003**, 1323–1325.

51. Wang, X., Migawa, M. T., Sannes-Lowery, K. A., Swayze, E. E. The synthesis and 16S A-site rRNA recognition of carbohydrate-free aminoglycosides. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4919–4922.

52. Chang, C. W. T., Hui, Y., Elchert, B., Wang, J., Li, J., Rai, R. Pyranmycins, a novel class of aminoglycosides with improved acid stability: the SAR of *D*-pyranoses on ring III of pyranmycin. *Org. Lett.* **2002**, *4*, 4603–4606.

53. Rai, R., Chen, H., Czyryca, P. G., Li, J., Chang, C.-W. T. Design and synthesis of pyrankacin: a pyranmycin class of broad-spectrum aminoglycoside antibiotics. *Org. Lett.* **2006**, *8*, 887–889.

54. Kirk, S. R., Luedtke, N. W., Tor, Y. Neomycin–acridine conjugate: a potent inhibitor of Rev–RRE binding. *J. Am. Chem. Soc.* **2000**, *122*, 980–981.

55. Luedtke, N. W., Baker, T. J., Goodman, M., Tor, Y. Guanidinoglycosides: a novel family of RNA ligands. *J. Am. Chem. Soc.* **2000**, *122*, 12035–12036.

56. Wang, H., Tor, Y. Dimeric aminoglycosides: design, synthesis and RNA binding. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1951–1956.

57. Blount, K. F., Zhao, F., Hermann, T., Tor, Y. Conformational constraint as a means for understanding RNA–aminoglycoside specificity. *J. Am. Chem. Soc.* **2005**, *127*, 9818–9829.

58. Blount, K. F., Tor, Y. A tale of two targets: differential RNA selectivity of nucleobase–aminoglycoside conjugates. *ChemBioChem* **2006**, *7*, 1612–1621.

59. Vicens, Q., Westhof, E. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **2001**, *9*, 647–658.

60. Pathak, R., Böttger, E. C., Vasella, A. Design and synthesis of aminoglycoside antibiotics to selectively target 16S ribosomal RNA position 1408. *Helv. Chim. Acta* **2005**, *88*, 2967–2985.

61. Gerogiadis, M. P., Constantinou-Kokotou, V. Synthesis of amino acid derivatives of neamine and 2-deoxystreotamine to be used as mutasynthons. *J. Carbohydr. Chem.* **1991**, *10*, 739–748.

62. Hyffeler, P. T., Liang, C.-H., Koeller, K. M., Wong, C.-H. The chemistry of amine-azide interconversion: catalytic diazotransfer and regioselective azide reduction. *J. Am. Chem. Soc.* **2002**, *124*, 10773–10778.

63. Douglas, N. L., Ley, S. V., Osborn, H. M. I., Owen, D. R., Priepke, H. W. M., Sarriner, S. L. Direct protection of 1,2-diols from a-diketones. *Synlett* **1996**, *1996*, 793.

64. Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. Monitoring molecular recognition of the ribosomal decoding site. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 3177–3182.

65. Burgess, K., Ohlmeyer, M. J. Manipulation of substrate-controlled diastereoselectivities in hydroboration in acyclic allylamine derivatives. *J. Org. Chem.* **1991**, *56*, 1027–1036.

66. Coombs, R. V., Koletar, J., Danna, R. P., Mah, H. Synthesis of 9,11-disubstituted 19-nor-steroids. *J. Chem. Soc., Perkin Trans. 1* **1975**, *9*, 792–798.

CHAPTER 4

Development of a Fluorescent HIV-1 DIS Binding Assay

4.1 THE IMPACT OF HIV ON HUMANS

In 1981, doctors in Los Angeles reported a rare form of pneumonia in small groups of homosexual men along with rare forms of cancer in other parts of the country.[1] The ailment was eventually termed Acquired Immunodeficiency Syndrome (AIDS), due to the increase in secondary infections found in patients with suppressed immune systems. Numerous misconceptions in the general public during the early years of this pandemic included the belief that the disease only affected homosexuals and was easily transmitted through casual contact. Due to the severe symptoms and high mortality rates, identification of the cause was paramount. In 1983 French scientists isolated and identified a lymphadenopathy-associated virus (LAV) as the cause of AIDS,[2,3] which is now known as the Human Immunodeficiency Virus (HIV). American scientists later confirmed these results in 1984.[3]

Over the past 30 years, the spread of HIV in modern countries has dramatically decreased due to the elucidation of transmission modes, administration of antiretrovirals, and establishment of health education campaigns. Notwithstanding, the reduction in HIV infection and treatment in third world countries remains elusive. The rapid appearance of resistant HIV-variants,[4–11] adverse effects of current treatments,[12,13] and recent indications for HIV "superinfections"[14,15] all highlight the difficulty of anti-

HIV drug development. Projections suggest that the number of individuals infected with HIV may double (and possibly triple) by the end of the decade. At the end of 2006, the World Health Organization (WHO) and UNAIDS Project reported that over 40 million people worldwide were infected with HIV or have symptoms concurrent with the full onset of AIDS infection.[16]

Current anti-HIV treatments have been focused on disrupting or modulating key steps in the retroviral life cycle. The current widespread use of highly active antiretroviral therapy (HAART) incorporates a combination of at least three drugs that independently target different stages of replication.[17,18] The majority of clinically administered drugs target the viral proteins reverse transcriptase and protease, while new classes of antiretroviral drugs currently in clinical trials target viral fusion, binding and viral-host genome integration.[19] Another focal point within the lifecycle for drug design encompasses the disruption of sequence-specific interaction between regulatory proteins and their respective viral RNA binding sites. The development of small molecules that block these interactions may be detrimental to viral replication and provide insight into the development of new therapeutics.[20–22] Accordingly, the inhibition of Tat–TAR and Rev–RRE interaction with small molecules has been widely studied in our laboratories[23,24] and in other research institutions.[25–27] However, current clinically-administered retrovirals do not target these RNA–protein interactions, but instead different stages and phases of the HIV lifecycle.

4.2 THE HIV LIFE CYCLE

Retroviruses do not possess the cellular machinery necessary for self-replication; instead, they must infect host cells and redirect cellular resources for the production of new viral particles (refer to Figure 4.1 for retroviral lifecycle).[28] HIV primarily infects T-helper lymphocytes and macrophages. First, during the process of viral attachment, two sets of proteins within the viral envelope, gp120 and gp41, bind either to CD4 receptors or to CXCR4 co-receptors on T-helper lymphocytes, or CCR5 co-receptors on macrophages. The next step in the cycle involves viral penetration or fusion where the viral nucleocapsid is injected into the host cell. An interplay of glycoproteins and CD4/CCR5 receptors allows for fusion of the viral envelope and cell membrane, thereby enabling direct contact between the host cell's cytoplasm and contents of the viral capsid. Finally, uncoating of the injected nucleocapsid releases the small viral genome into the host's cytoplasm, making it available for reverse transcription and integration into the host genome.

In this process, a RNA-dependent DNA polymerase, reverse transcriptase (RT), catalyzes the synthesis of DNA complementary to the viral RNA necessary for insertion into the host genome. After synthesis of the first DNA strand, the enzyme degrades the viral RNA and catalyzes the synthesis of a second DNA strand that is complementary to the first. This *c*DNA (complementary DNA) is then available for host integration. The newly synthesized *c*DNA is then transported into the cell's nucleus for integration with the host's DNA. Once inside the host nucleus, the viral protein integrase site specifically cleaves two bases at the 3′ ends of the double helical

viral DNA exposing the 3′-hydroxyl groups. Integrase then catalyzes a transesterification reaction by nucleophilic attack of the *c*DNA phosphodiester bond. Subsequent coupling and processing of any mismatches or gaps during integration is catalyzed by the host's own DNA repair machinery.

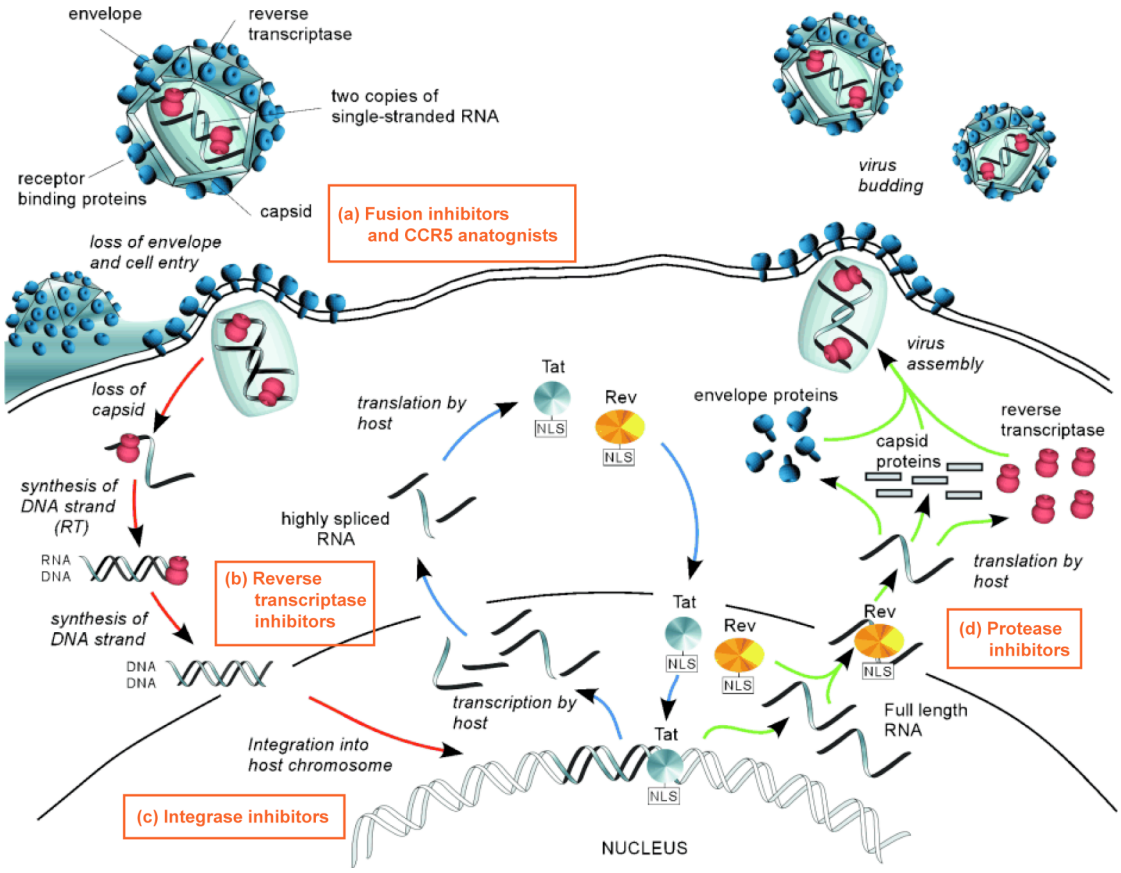


Figure 4.1. HIV life cycle. Stages of the life cycle where current and upcoming drugs interfere with replication have been highlighted in red (a–d). (Figure adapted from Tor, Y. Targeting RNA with small molecules. *ChemBioChem* **2003**, *4*, 998–1007).

After initial integration, the host cell is transformed into a provirus, which can remain latent from months to years until activation, by factors still unknown. Upon activation, transcription factors catalyze the synthesis of full length viral RNA which is

then exported out of the nucleus for packaging during viral assembly, or partially spliced and used as templates for HIV protein synthesis. The proteins necessary for viral assembly and replication are produced in the cytoplasm. One of these proteins, Tat, is responsible for continuing cellular transcription of viral mRNA. Tat is synthesized in the cytoplasm and imported into the nucleus where it binds to the transactivating region (TAR), located near the 5′-end of RNA and stimulates the rapid production of full length viral RNA. Another protein, Rev, assists in transport of the newly synthesized viral genome. Similar to Tat, Rev is transported into the nucleus and binds to a region of RNA called the Rev Response Element (RRE). Rev facilitates the export of full length viral RNA into the cytoplasm for packaging into new viral particles, preventing any degradation before incorporation into new virons.[29] Studies have shown that in the absence of Rev peptide, only short, spliced mRNA sequences are produced.

During assembly and eventual release of the virion, viral proteins synthesized from the *gag* and *pol* genes in the cytoplasm are further processed and cleaved by viral proteases. Concurrently, two RNA strands and necessary viral proteins are packaged into the new viral particle which subsequently buds, forming a new viral envelope separating the virion from the host cytoplasm. After separation from the host cell, the immature viral particle undergoes a maturation process where the nucleocapsid forms around the two viral RNA genomes and further processing of proteins occurs to allow for binding to another lymphocyte or macrophage.

Intriguingly, signaling for encapsidation of the viral genome may be mediated by the dimerization of the two full-length RNA genomes.[30] While there is considerable disagreement on the requirement for dimerization and its role in signaling viral packaging,[31] the presence of a dimerized genome has been observed in most retroviruses. The conservation of this interaction suggests that dimerization may play a pivotal role in viral replication. Despite the identification of where dimerization occurs along the retroviral genome (see Chapter 4.4),[32] there is no conclusive evidence for the precise time of onset of in vivo dimerization during the lifecycle.[31] Isolation of murine leukemia virions (MuLV) shortly after budding has revealed that RNA is in a dimeric form, although physically different than dimeric RNA in more matured viral particles.[33] Another group has reported the visualization of RNA dimers in newly infected cells; this observation suggests that dimerization occurs early in viral replication and not during encapsidation.[34]

4.3 THERAPEUTICS TARGETING THE HIV LIFE CYCLE

Many current and upcoming drugs target different steps in the viral lifecycle. Blocking cell-surface receptors has been recognized as a potential strategy for antiviral development. Recently, Pfizer has developed a cell entry inhibitor that acts as a CCR5 antagonist; this inhibitor is named *Maraviroc*™ and has currently passed Phase III clinical trials.[35] Preventing injection of viral material into the host cell has also been targeted with fusion inhibitors such as *Enfuvirtide*™, which has shown to be effective in individuals with multidrug resistance.[36]

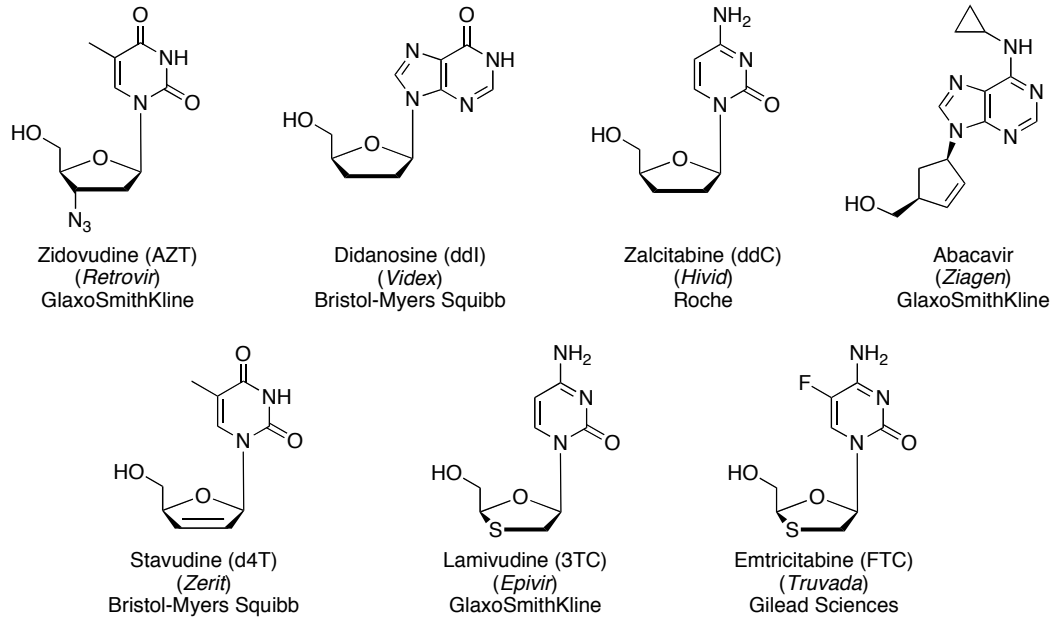
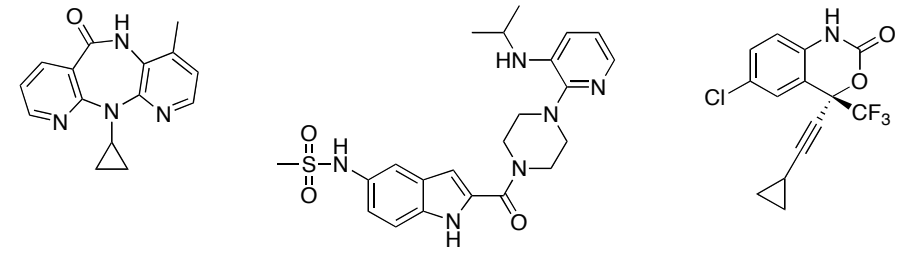


Figure 4.2. Structures of nucleoside reverse transcriptase inhibitors. The common chemical name, abbreviation, marketing name (in italics), and manufacturer are indicated.

Reverse transcription in viral replication is the most therapeutically targeted stage in the HIV lifecycle. Reverse transcriptase inhibitors are categorized into two classes of compounds based on either a nucleoside or non-nucleoside scaffold. Nucleoside reverese transcriptase inhibitors (NRTIs) are mimics of natural nucleosides that are incorporated into the growing proviral DNA strand and terminate transcription (Figure 4.2). Termination is caused by the lack of a 3′ hydroxyl moiety on the ribose sugar, preventing genome extension. The first NRTI indicated for the treatment of HIV was AZT (3′-azido-2′,3′-dideoxythymidine, marketed as *Retrovir™)*. Other NRTIs administered as a single drug or in a combination regimen include Didanosine, Emtricitabine, and Abacavir. In contrast, NNRTIs (non-nucleoside reverse transcriptase inhibitors) inhibit activity by binding to the active site of reverse



Nevirapine
(*Viramune*)
Boehringer Ingelheim

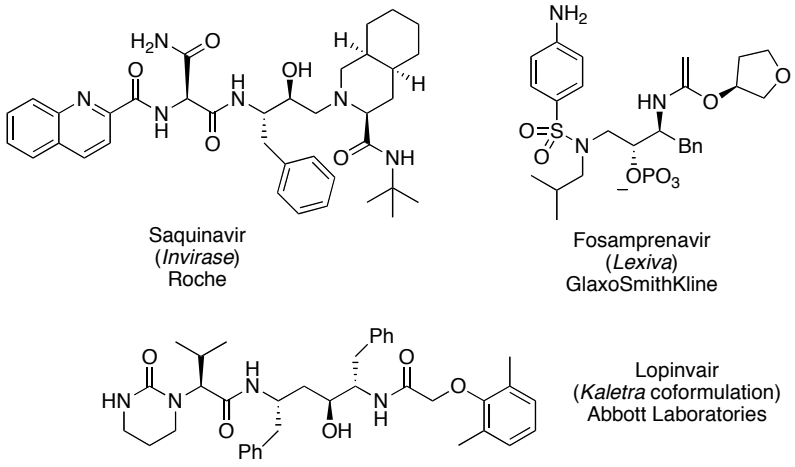
Delavirdine
(*Rescriptor*)
Pfizer

Efavirenz
(*Sustiva*)
Bristol-Myers Squibb

Figure 4.3. Structures of non-nucleoside reverse transcriptase inhibitors. The common chemical name, marketing name (in italics), and manufacturer are indicated.

transcriptase. NNRTIs in current use include Nevirapine, Delaviridine, and Efaviren (Figure 4.3).

At the present time, there is no clinical administration of integrase inhibitors; however, many are involved in phase II and III testing. The most likely drug candidate to gain FDA approval by the end of 2007 is Merck's *Raltegravir*™.[35] Another class of clinically utilized antiretrovirals is protease inhibitors, which interfere with the normal function of protease enzymes (Figure 4.4). Finally, while no therapeutics in clinical



Saquinavir
(*Invirase*)
Roche

Fosamprenavir
(*Lexiva*)
GlaxoSmithKline

Lopinvair
(*Kaletra* coformulation)
Abbott Laboratories

Figure 4.4. Structures of protease inhibitors. The common name, marketing name (in italics), and manufacturer are indicated.

trails target the Dimerization Initiation Site (DIS), structural information recently demonstrated a strong resemblance to the prokaryotic A-site, the binding site and target of aminoglycoside antibiotics.[37] This suggests possible application to the HIV-1 DIS of antibiotics currently targeting the prokaryotic ribosome.

4.4 HIV-1 DIMERIZATION: STRUCTURE & IMPORTANCE

In 1967, retroviruses were first recognized to contain two copies of their viral genome after sedimentation studies on the Rous sarcoma virus (RSV) and murine leukemia virus (MuLV) discovered cellular particles that were larger than expected.[38] Later in 1975, electron microscopy studies on feline retroviruses revealed Y- or T-shaped structures near the center of the RNA genome, which suggested that the juncture might be a linkage of two identical RNA strands.[39] Dimeric RNA has been isolated in vivo from tissue cultured virons and analyzed by Northern blotting,[40] while in vitro, RNA oligonucleotides have been synthesized and observed to form dimers.[41] While mutation of the DIS sequence does not halt the assembly of a dimer in vivo, defects in RNA packaging in new viral particles were observed.[30,42] Additionally, engineered mutants containing one viral genome were able to package the monomer into a new virion, but were unable to replicate further.[31,43,44]

The dimeric nature of retroviral genomes have been shown to play an important role in key HIV life cycle events, including reverse transcription, encapsidation, and recombination.[45] Two copies of viral genome are suggested to be advantageous during replication. One study proposes a dimeric genome allows for reverse transcriptase

(RT) to "read" from the other genome in the event that the RNA strand becomes damaged.[46] Moreover, a diploid genome may possibly increase genetic recombination and retroviral evolution.[47]

Dimerization occurs at stem loop 1 (SL1 or DIS), the first of four characterized stem loops found on the ca. 120 nt long HIV Ψ-packaging site (Figure 4.5a).[48,49] This conserved RNA stem loop initiates dimerization by forming a "kissing loop", where six complementary nucleotides from each RNA strand associate through base pairing (Figure 4.5b). The nine nucleotide loop varies in sequence depending on subtype.[50] Subtypes A and G, found primarily in Southeast Asia and Northwest Africa have a loop sequence of 5′–AG**GUGCAC**A–3′, while subtypes B and D are found in Europe, the Pacific, and North and Central America and are characterized by the loop sequence 5′–AA**GCGCGC**A–3′ (nucleotides in bold denote the bases directly involved with "kissing loop" hydrogen bonding). Subtype C is found in India, South and East Africa, and has a loop sequence of 5′–AA**GCGCGC**U–3′. Finally, the subtypes F and H are found predominantly in South America and Central Europe and have loop sequences of 5′–AA**GUGCAC**A–3′. The purines that flank the six nucleotides involved in kissing loop formation have been reported to assist in stabilizing its tertiary structure.[51]

This kissing loop complex, however, is a meta-stable structure and has been shown in vivo to isomerize to the more stable extended duplex structure, a process catalyzed by the NCp7 encapsidation protein (Figure 4.5b).[52] Additionally, in vitro studies have shown the kissing loop complex to thermally isomerize in the absence of enzyme under appropriate buffer conditions to the extended duplex structure.[53–55]

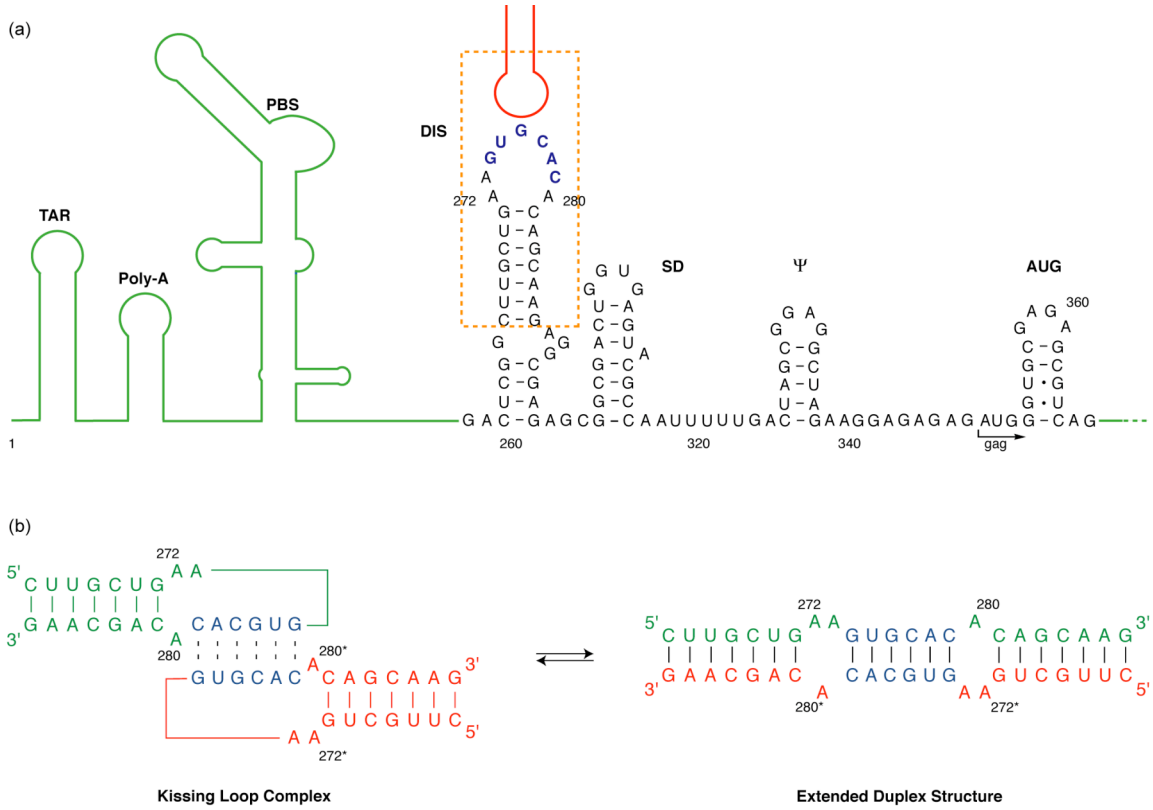


Figure 4.5. a) Schematic representation of the 5′-end of the HIV RNA genome and the Dimerization Initiation Site (DIS) (orange box). The sequence of the Ψ-packaging region is explicitly shown. b) Rearrangement of the DIS kissing loop complex to an extended duplex structure (note the different orientation compared to (a)). Strands are colored separately (red and green) as well as bases involved in kissing loop base-pairing (blue). Abbreviations: TAR, *trans*-acting response element; poly-A, polyadenylation signal; PBS, primer-binding site; SD, major splice-donor site; Ψ (psi), main part of the RNA packaging signal; AUG, translation initiation codon of the *gag* gene.

Although dimerization is an integral part of the HIV lifecycle, the DIS has not been extensively studied as a therapeutic target. Since mutations in the DIS have been shown to dramatically diminish viral reproduction,[56,57] small molecules that can disrupt or alter this process may constitute a viable antiretroviral therapeutic strategy.[31,50] Despite the availability of crystal structures of short DIS oligonucleotides complexed with aminoglycosides,[37,58,59] little is known about the recognition properties of HIV

DIS in solution. To this end, a sensitive fluorescence DIS assay has been developed to screen potential binders in "real time".

4.5 DESIGN AND DEVELOPMENT OF A REAL-TIME HIV-1 DIS FLUORESCENCE ASSAY

In 2003, Ehresmann and colleagues reported the primary, secondary, and tertiary structures of the HIV DIS and those of the bacterial A-site are remarkably similar, including the kissing loop and extended duplex structures (Figure 4.6).[37] The prokaryotic A-site represents a paradigm for understanding ligand–RNA interactions, and is characterized by its two dynamic A residues (1492 and 1493) which are involved in the decoding process of translation.[60–62] Analysis by both NMR[63–65] and X-ray crystallographic[66,67] structures of the entire 30S ribosomal subunit have revealed these conformationally flexible adenines are displaced from the RNA helix interior upon ligand binding.[64,68] The motility of these flexible adenosines in the A-site has recently been exploited by Hermann and Pilch to monitor ligand binding by replacing both residues with the fluorescent nucleobase analog 2-aminopurine.[69,70] Coincidentally in the HIV DIS, structural work has suggested purines A272 and A273 are displaced from the RNA helix upon small molecule binding.[50,58,59] We therefore hypothesized that replacing one or two of these purine bases with 2-aminopurine (2AP) might provide an effective reporting assay for DIS–ligand (Figure 4.7).

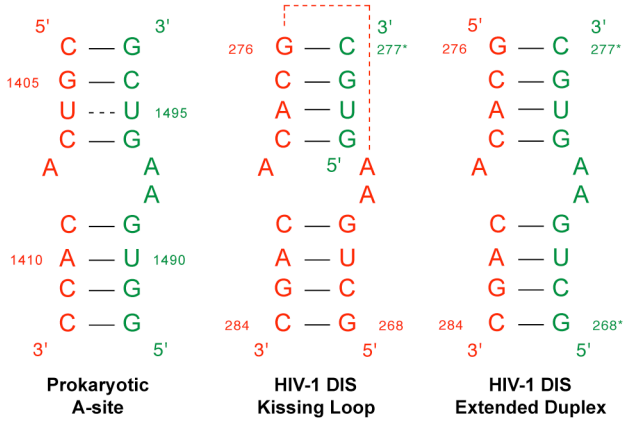


Figure 4.6. Similarity between the primary and secondary structures of the prokaryotic A-site, DIS kissing loop complex, and DIS extended duplex structure. Subtype F of the HIV DIS is shown above and is characterized by bases A272, A273 and A280 that flank the loop sequence 5′-GUGCAC-3′. Subtype A, however, contains an A273G mutation with the same loop sequence as subtype F. Subtype B also contains flanking bases A272, A273 and A280 but is instead characterized by the loop sequence 5′-GCGCGC-3′.

To mimic the extended duplex structure of HIV-1 DIS (subtype F), two minimized 23mer RNA constructs were designed containing the fluorescent nucleobase analog 2-aminopurine (Figure 4.8).[71] The constructs employed in our fluorescent assay, DIS272(2AP) and DIS273(2AP), were based on previously reported short RNA constructs that were shown to mimic the kissing loop and extended duplex structure found in viral particles.[37] As a control, the wild type DIS oligonucleotide was also synthesized (Dharmacon) along with a non-dimerizing construct. The major concern was the ability of the modified truncated DIS oligonucleotide to mimic the native viral RNA's three dimensional structure, and maintain the structural properties necessary for rearrangement to the extended duplex. Also of concern was the ability of the modified oligonucleotide's rearrangement to be controlled under conditions reported for larger RNA constructs.[72]

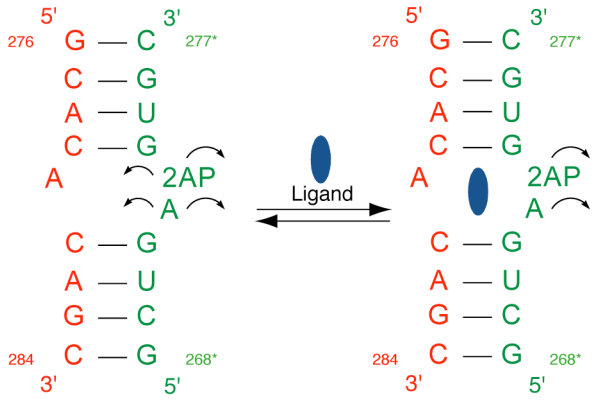


Figure 4.7. Putative base-flipping dynamics of purines 272 and 273 in the DIS extended duplex in analogy to the A-site. In the unbound state, purines 272 and 273 are presumed to freely rotate in and out of the helical domain. Upon ligand binding, purines 272 and 273 are displaced and forced out to the exterior of the DIS RNA helix.

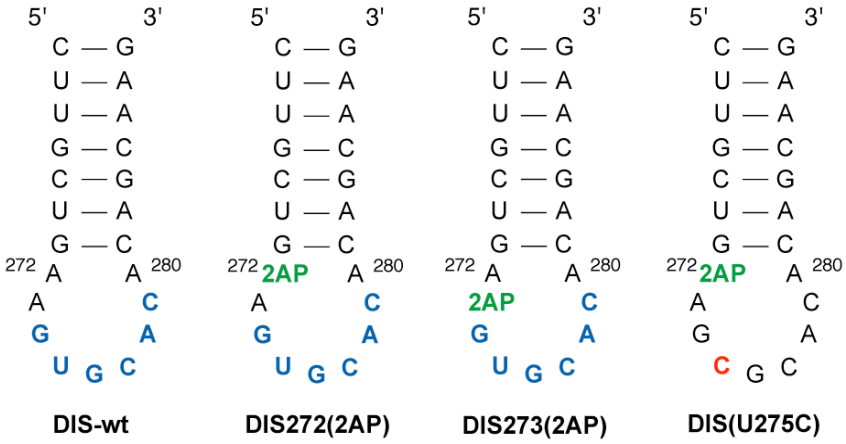


Figure 4.8. Sequence and secondary structure of 23 mer RNA DIS constructs (subtype F). Constructs DIS272(2AP) and DIS273(2AP) are the fluorescently-labeled RNA constructs, while DIS-wt is the wildtype sequence and DIS(U275C) is a non-dimerizing control sequence. Explicitly shown are the bases involved in the kissing-loop complex interaction (blue), 2-aminopurine (2AP, green), and the U275C mutation (red) that prevents kissing-loop complex formation.

4.6 CONTROLLING KISSING LOOP AND EXTENDED DUPLEX FORMATION

In viral particles, isomerization of the kissing loop complex to an extended duplex is mediated by encapsidation protein NCp7. Numerous groups have reported that isomerization can also be controlled in vitro under certain buffer and thermal conditions. Furthermore, other groups have investigated the isomerization process between these two structures on both full, modified, and short oligonucleotides.[51,71,73–75] However, the slight modification of replacing purines with 2AP could potentially have affected the stability of both structures since the flanking purines of the kissing loop are responsible for complex stability.[51]

Monitoring and determining the specific state of the RNA oligonucleotides (i.e., single stranded, kissing loop or extended duplex) was critical for assay development. Native polyacrylamide gel electrophoresis (PAGE) was utilized to monitor the association and isomerization processes of the DIS constructs (Figure 4.9). Several controls were first examined. DIS(U275C), was selected as a non-dimerizing RNA negative control sequence (Figure 4.8), for marking the relative migration of a monomeric RNA hairpin (Figure 4.9, lane 1). Denaturation and renaturation of this construct resulted in the formation of an extended duplex (with traces of a monomeric oligonucleotide) as shown in lane 2 (Figure 4.9). No kissing-loop was ever observed with this mutant construct. As a positive control, the fluorescent construct DIS272(2AP) was selected. This construct was denatured and renatured in the presence of Mg^{2+}, followed by incubation for 2 h at 55 °C, which are the same

conditions used for solution titrations, to generate the extended duplex (lane 3). To avoid extra-stabilization of the kissing loop complex, the association was carefully monitored as a function of time in a similar buffer without Mg^{2+}. Construct DIS272(2AP) was denatured at 90 °C and slowly renatured by incubation at 55 °C for 0, 15, 30, 60, 90 and 120 min (Figure 4.9, lanes 4–9, respectively). As the native gel revealed, the 2AP-modified RNA construct DIS272(2AP) spontaneously forms a kissing loop upon heat denaturation, then thermally rearranges into the extended duplex by incubation at 55 °C for an extended period of time (Figure 4.9). All binding experiments were conducted after 2 h of incubation due to the metastable nature of the kissing loop complex,[55,76] and to ensure a majority of extended duplex structure in solution.

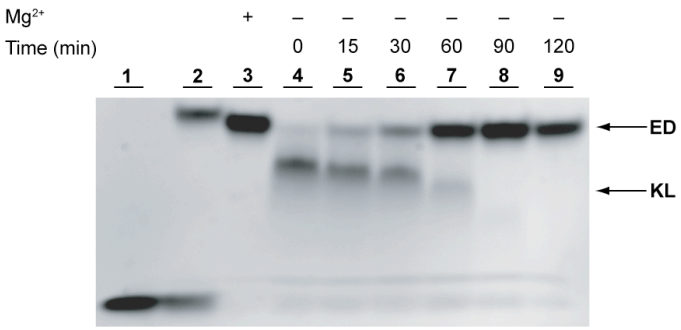


Figure 4.9. Native PAGE used to monitor the association of the RNA hairpins to a kissing loop (KL) and rearrangement to the extended duplex (ED). Lane 1: DIS(U275C) is a monomeric RNA hairpin control. Lane 2: DIS(U275C) was denatured and renatured to form an extended duplex. Lane 3: DIS272(2AP) was incubated for 2 h at 55 °C in the titration buffer to form an extended duplex structure. Lanes 4-9: DIS272(2AP) was denatured and incubated at 55 °C for 0, 15, 30, 60, 90, and 120 min, respectively, illustrating the thermal isomerization of the kissing loop complex to extended duplex over time. See Materials and Methods for gel and buffer conditions.

4.7 THERMAL DENATURATION STUDIES

To determine the impact of 2AP substitution on extended duplex stability, thermal denaturation experiments were employed. Melting curves of modified constructs DIS272(2AP) and DIS273(2AP) were compared to the unmodified DIS-wt (Figure 4.10). Nearly identical melting temperatures were observed for all three constructs, suggesting minimal perturbation to the extended duplex structure (T_m = 72 °C). These results are consistent with previous observations demonstrating that the incorporation of 2AP into nucleic acids minimally disturbs double helical structures.[70,77,78] Moreover, a thermal melting transition (inflection point) was not observed for a kissing loop structure, providing further evidence that formation of the extended duplex was complete under stated incubation and folding assay conditions.

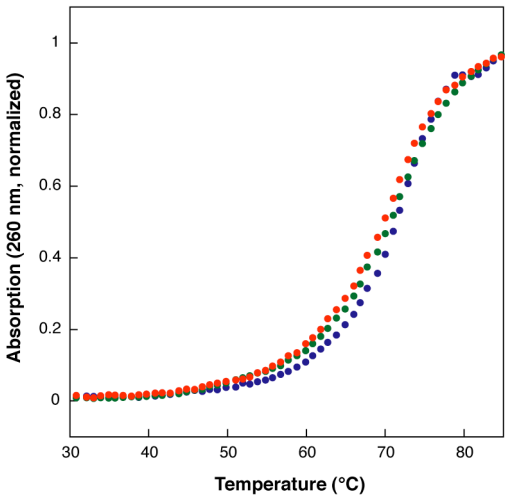


Figure 4.10. Normalized thermal melting profile of RNA constructs DIS-wt (blue, T_m = 72 ± 0.6 °C), DIS272(2AP) (green, T_m = 72 ± 0.6 °C), and DIS273(2AP) (red, T_m = 70 ± 2.1 °C). Refer to Chapter 4.13 for experimental details.

4.8 MONITORING SMALL MOLECULE BINDING WITH DIS272(2AP)

To test the viability and fluorescence response of the DIS272(2AP) minimized construct for ligand binding, different aminoglycosides were screened (Figure 4.11a). Initially, paromomycin, previously reported to bind the DIS by footprinting studies,[37] was titrated into a solution of the RNA extended duplex. An increase in fluorescence

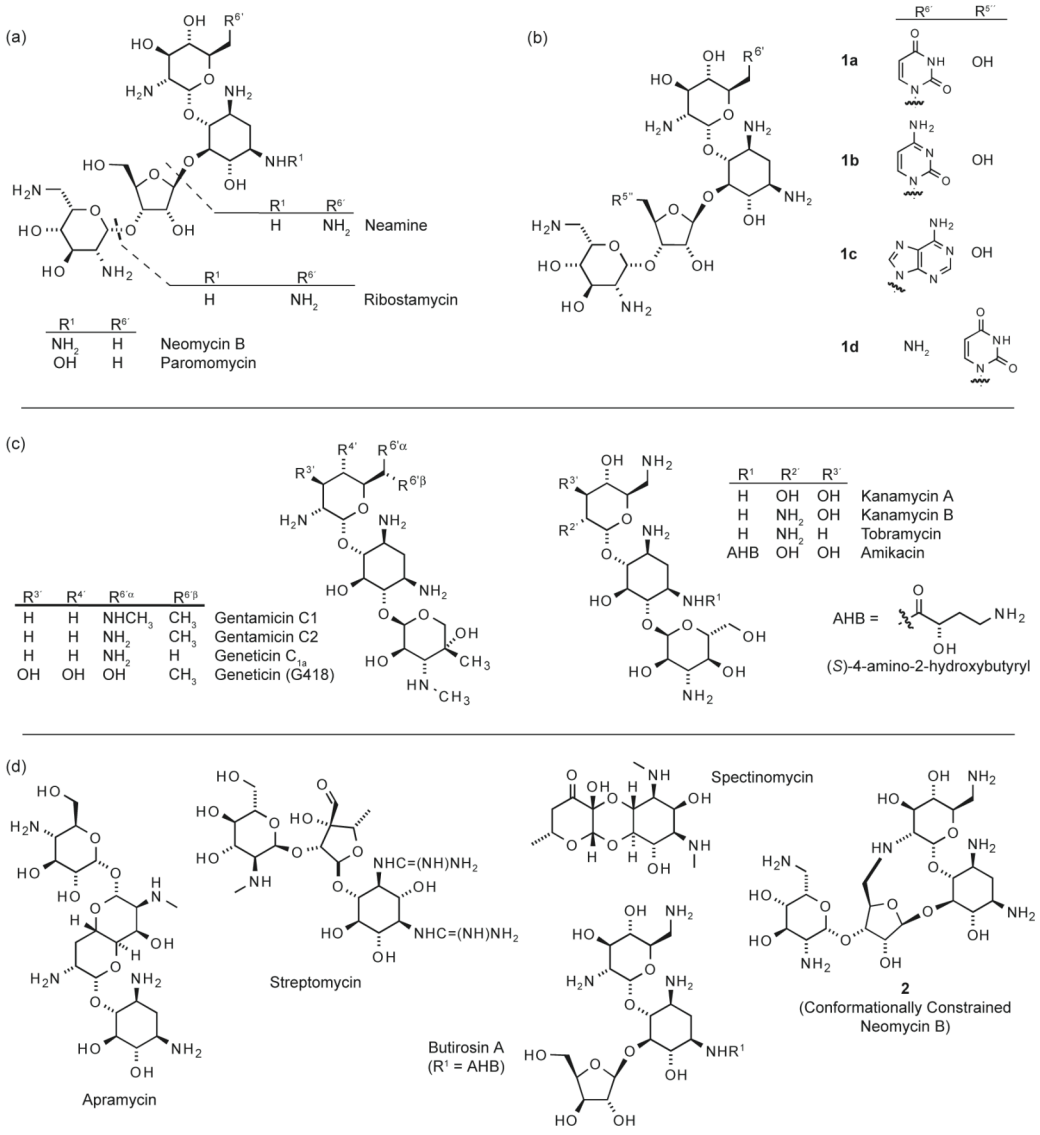


Figure 4.11. Structures of the different classes of ligands tested. a) 4,5-disubstituted 2-DOS aminoglycosides. b) Semi-synthetic nucleobase-neomycin conjugates. c) 4,6-disubstituted 2-DOS aminoglycosides. d) Other related aminoglycoside antibiotics.

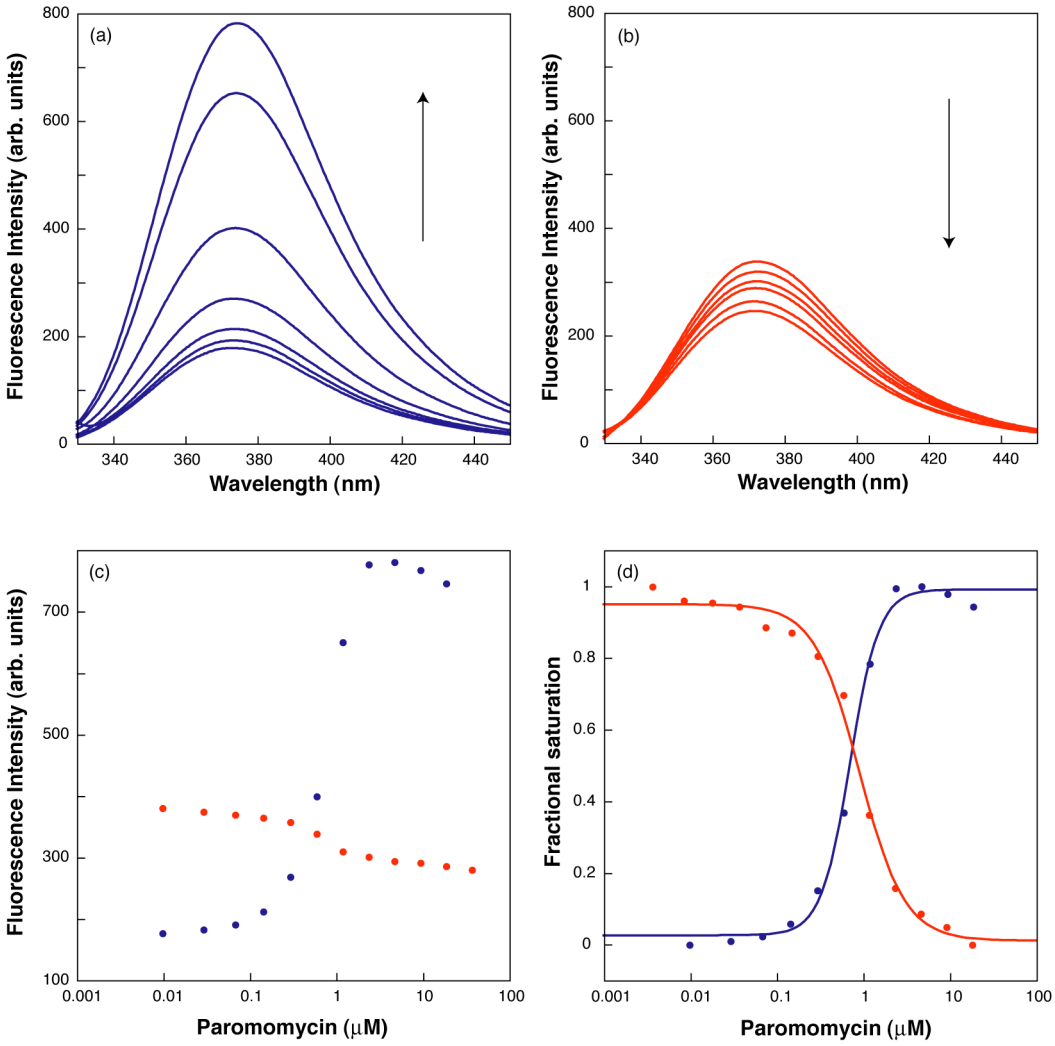


Figure 4.12. Fluorescence titrations of both 2AP-labeled DIS constructs with paromomycin. a) Representative spectra of fluorescence intensity enhancement for DIS272(2AP) with increasing concentrations of paromomycin. b) Representative spectra of fluorescence quenching for DIS273(2AP) with increasing concentrations of paromomycin. Note: some intermediate spectra were omitted for clarity in both (a) and (b). c) Comparison of absolute fluorescence intensity values of DIS272(2AP) (blue) and DIS273(2AP) (red) in the presence of increasing concentrations paromomycin. d) Comparison of normalized fluorescence intensity change for DIS272(2AP) (blue) and DIS273(2AP) (red) for paromomycin titrations. Note the similar EC_{50} values observed with both fluorescently-labeled RNA constructs.

intensity at 370 nm was observed as the concentration of paromomycin was increased (Figure 4.12a). At the end of the titration, the fluorescence intensity of 2AP increased over 300% (Figure 4.12c). Throughout the titration, shifts in the maximum emission wavelength were not observed. The fluorescence intensity was normalized and plotted against increasing concentrations of paromomycin resulting in a binding isotherm (Figure 4.12d) The data points were then curve fit with a dose-response algorithm to yield an EC_{50} value (the effective concentration at which 50% of the RNA is ligand bound) of 0.5 ± 0.2 μM.

Neomycin and paromomycin are closely related aminoglycosides, which differ by replacement of the 6′ hydroxyl with an amino functionality (Figure 4.11a). A similar titration was performed by addition of increasing concentrations of neomycin to a solution of the folded RNA. Upon the addition of neomycin, the fluorescence intensity was enhanced by 150% in a concentration dependent manner (Figure 4.13a). A dose-response curve fit of the normalized fluorescence data resulted in an EC_{50} value of 0.6 ± 0.2 μM (Figure 4.13c).

Other aminoglycosides within the 4,5-disubstituted 2-deoxystreptamine family, including ribostamycin, butirosin A and neamine, were similarly evaluated (Figure 4.11a,d). A summary of EC_{50} values and percent change in 2AP emission intensity upon ligand binding are listed in Table 4.1. Neomycin B and paromomycin have similar binding affinities but differ in fluorescence enhancement upon binding nearly two-fold. Ribostamycin, a smaller three-ring aminoglycoside, exhibited a fluorescence intensity enhancement of approximately 100% and produced an EC_{50}

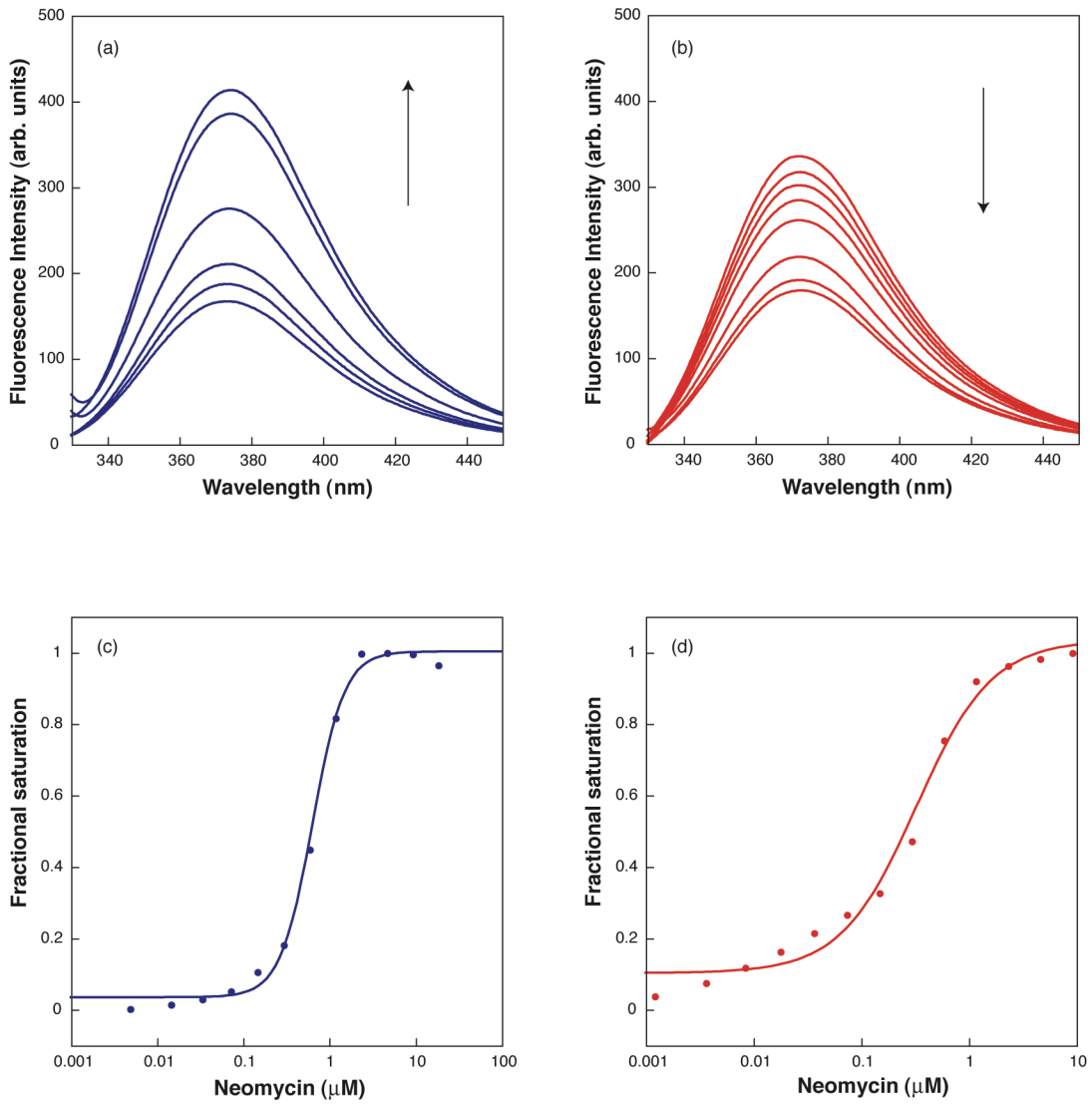


Figure 4.13. (a) Fluorescence emission curves of DIS272(2AP) and (b) DIS273(2AP) with increasing concentrations of neomycin B. (c) Normalized binding isotherms of DIS272(2AP) and (d) DIS273(2AP) with increasing concentrations of neomycin B.

value of 16.0 ± 4.0 μM (Figure 4.14a and 4.14c). Butirosin A, a derivative of ribostamycin that incorporates an aminohydroxybutyric acid moiety (AHB) on the N1 of 2-DOS, induced a fluorescence intensity enhancement of 100% and yielded EC_{50} value of 5.5 ± 0.6 μM (Figure 4.15a and 4.15c). This data indicates butirosin A binds

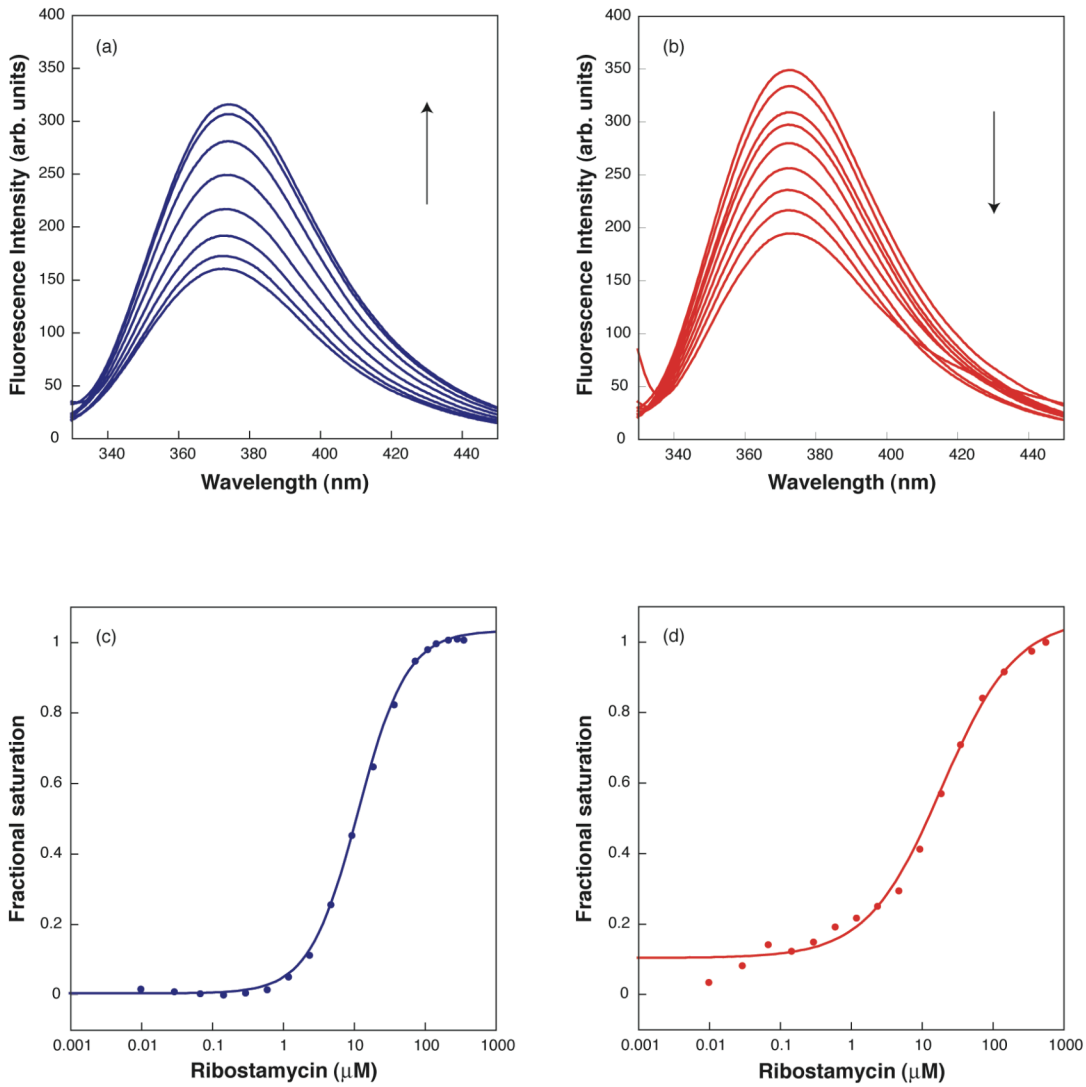


Figure 4.14. (a) Fluorescence emission curves of DIS272(2AP) and (b) DIS273(2AP) with increasing concentrations of ribostamycin. (c) Normalized binding isotherms of DIS272(2AP) and (d) DIS273(2AP) with increasing concentrations of ribostamycin.

more tightly than the parent ribstomycin. Neamine, consisting of rings I and II, has been recognized as the minimal pharmacore necessary for binding to the prokaryotic A-site and was also screened for DIS binding. This aminoglycoside induced a 75% increase in fluorescence intensity and possesses an EC_{50} value of 21.0 ± 1.2 μM (Figure 4.16a and 4.16c) Consequently, neamine was defined as the "fluorescence

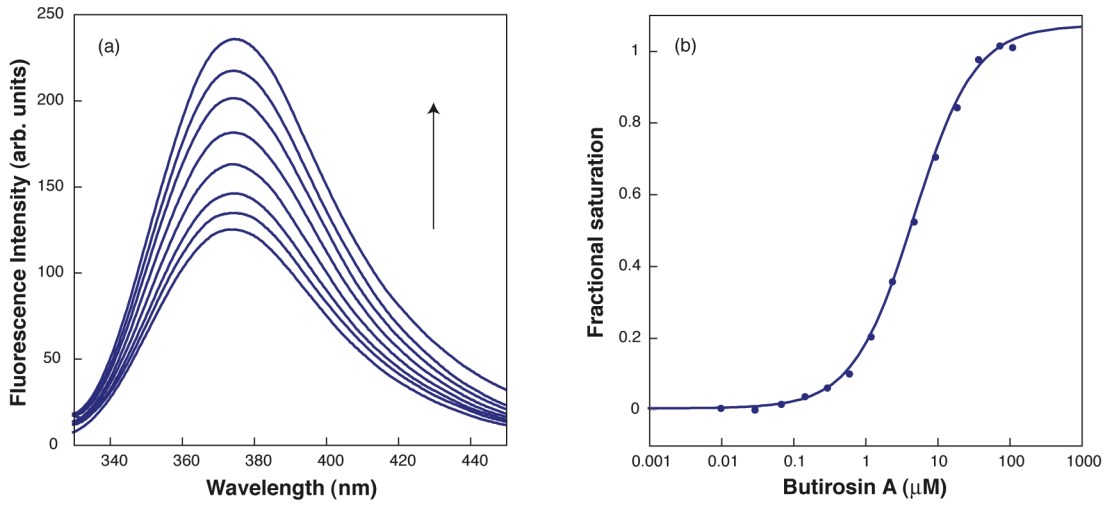


Figure 4.15. (a) Fluorescence emission curves of DIS272(2AP) with increasing concentrations of butirosin A. (b) Normalized binding isotherm of DIS272(2AP) with increasing concentrations of butirosin A.

change threshold" for DIS272(2AP) to differentiate between specific and non-specific binding. Ligands that induced a smaller overall change in fluorescence intensity were considered to be poor DIS binders and not further tested.

To further assess the utility of our fluorescence assay, three neomycin B analogs were screened. Previously reported by Blount,[79] the uracil-conjugated neomycin derivative **1a** (Figure 4.11b) induced an intensity increase of approximately 175% upon binding to DIS272(2AP) with an EC_{50} value of 2.6 ± 1.1 μM (Figure 4.17a and 4.17b). While conjugate **1d** with a uracil moiety at the 5″-position produced an increase in intensity of over 375% upon binding to DIS272(2AP), it exhibited similar binding affinity (2.2 ± 0.1 μM) as compared to **1a** (Figure 4.18a and 4.18c). Other compounds, where adenine and cytosine moieties were conjugated at the 6′ or 5′ positions, were also tested for binding (Figure 4.11b); however, none elicited a

fluorescence change greater than 75%, therefore, binding could not be confirmed through fluorescence. Finally, a conformationally-constrained neomycin B analog **2**[80] (Figure 4.11d) was tested against DIS272(2AP) which produced an intensity increase of greater than 175%, and an approximately 3-fold lower binding affinity (1.9 ± 0.2 μM) as compared to neomycin or paromomycin (Figure 4.19a,c).

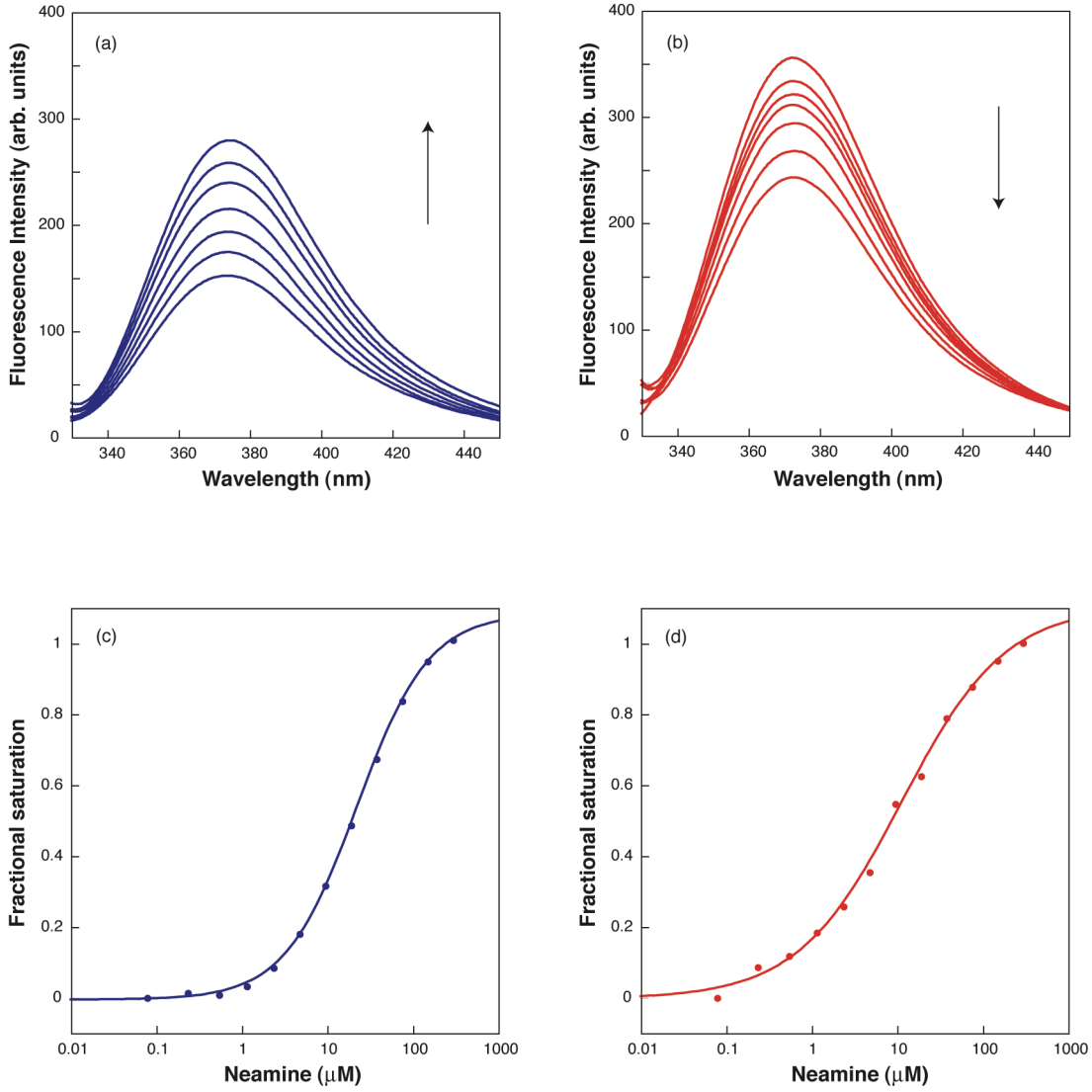


Figure 4.16. (a) Fluorescence emission curves of DIS272(2AP) and (b) DIS273(2AP) with increasing concentrations of neamine. (c) Normalized binding isotherms of DIS272(2AP) and (d) DIS273(2AP) with increasing concentrations of neamine.

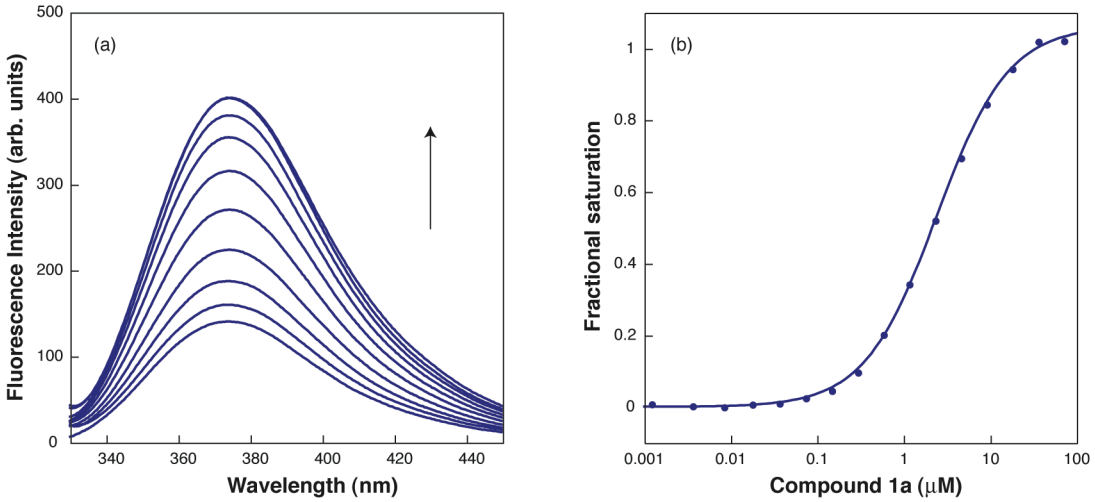


Figure 4.17. (a) Fluorescence emission curves of DIS272(2AP) with increasing concentrations of **1a**. (b) Normalized binding isotherm of DIS272(2AP) with increasing concentrations of **1a**.

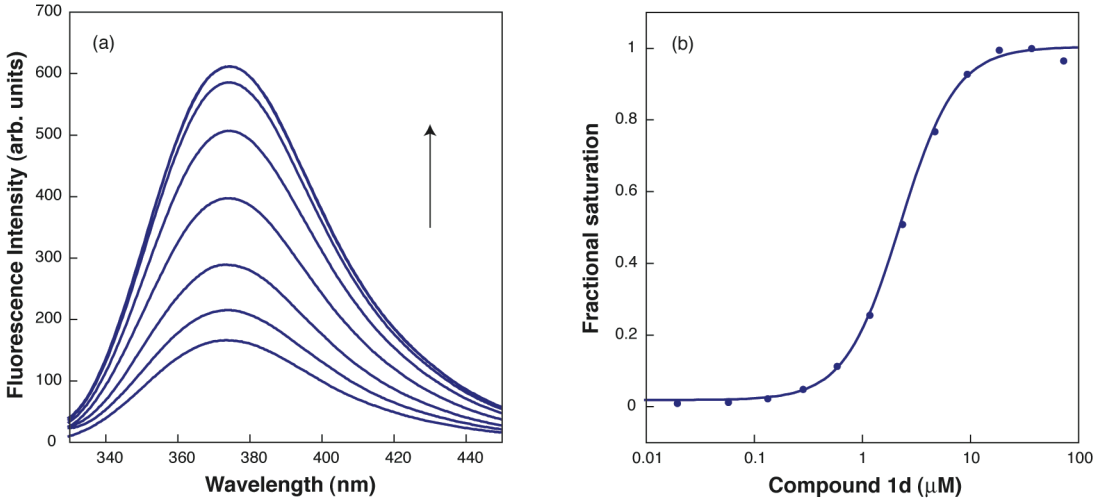


Figure 4.18. (a) Fluorescence emission curves of DIS272(2AP) with increasing concentrations of **1d**. (b) Normalized binding isotherm of DIS272(2AP) with increasing concentrations of **1d**.

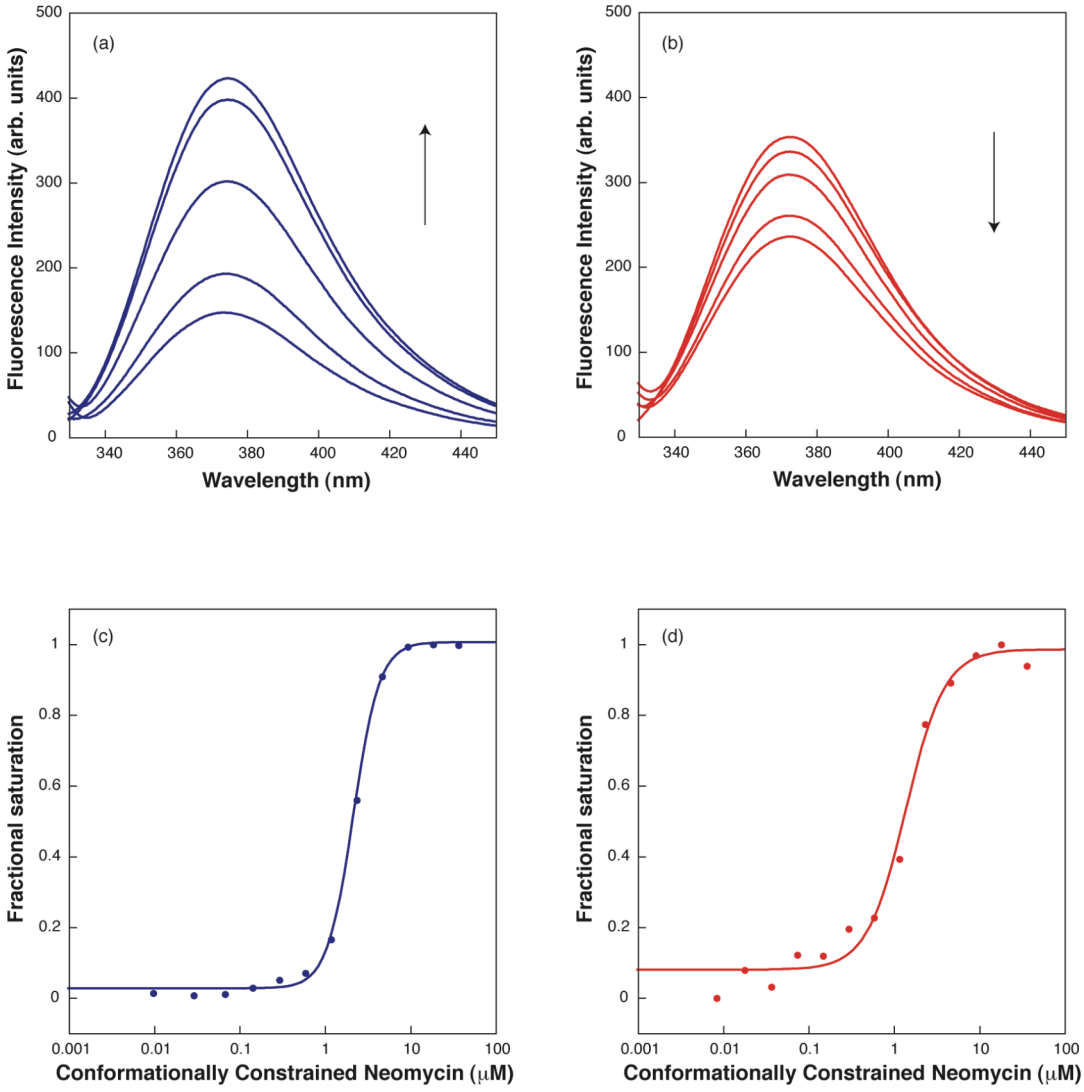


Figure 4.19. (a) Fluorescence emission curves of DIS272(2AP) and (b) DIS273(2AP) with increasing concentrations of **2**. (c) Normalized binding isotherms of DIS272(2AP) and (d) DIS273(2AP) with increasing concentrations of **2**.

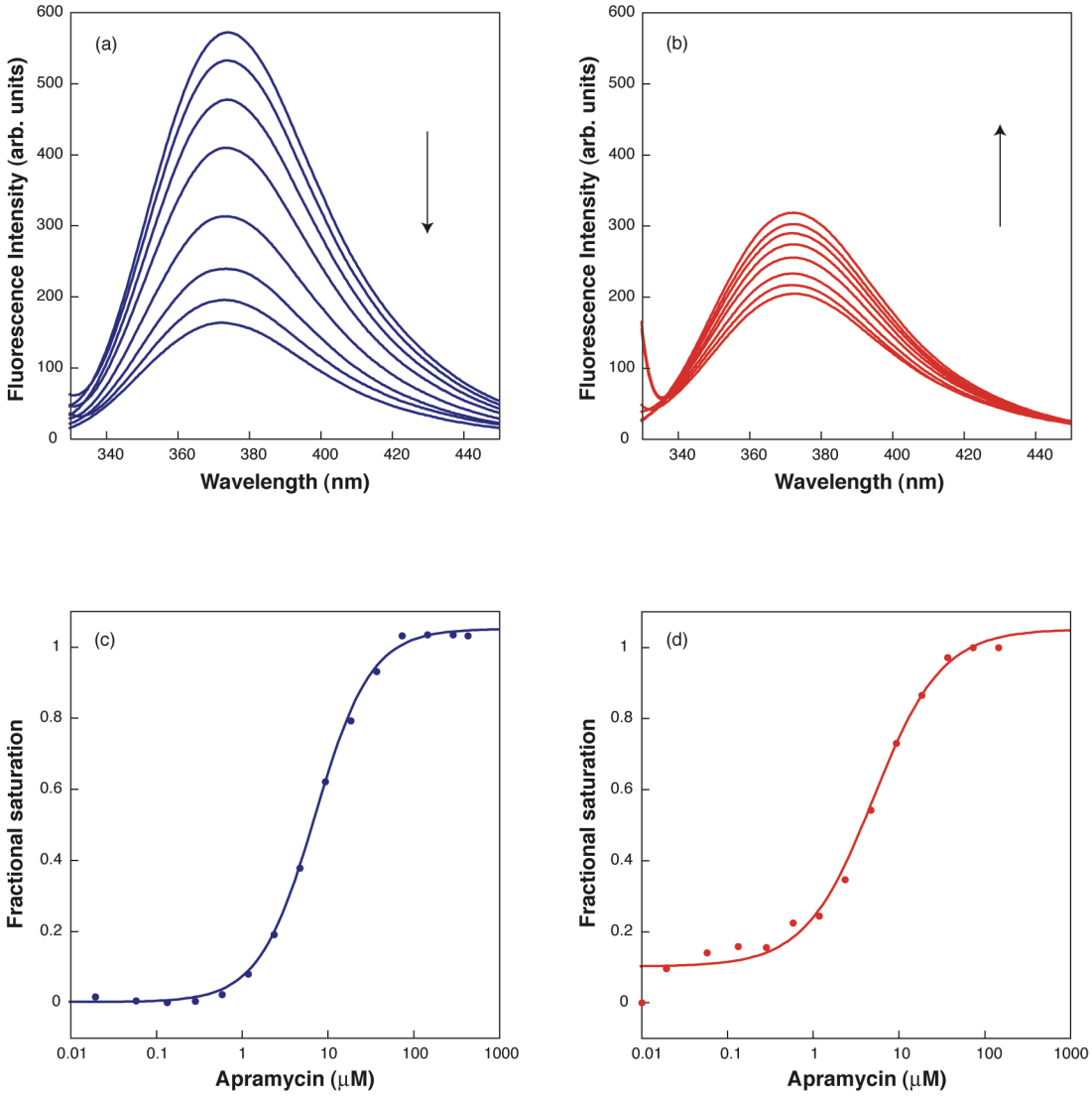


Figure 4.20. (a) Fluorescence emission curves of DIS272(2AP) and (b) DIS273(2AP) with increasing concentrations of apramycin. (c) Normalized binding isotherms of DIS272(2AP) and (d) DIS273(2AP) with increasing concentrations of apramycin.

Other aminoglycosides were tested for DIS binding, including the 4,6-disubstituted 2-DOS family, streptomycin and spectinomycin (Figure 4.11d). None elicited a change in fluorescence intensity. Notwithstanding, apramycin, a structurally unrelated aminoglycoside to the neomycin family, was discovered to possess affinity for the DIS. Upon binding, the antibiotic caused a fluorescence enhancement of 200% with an EC_{50} value of 7.6 ± 1.0 μM (Figure 4.20a,c).

4.9 MONITORING SMALL MOLECULE BINDING WITH DIS273(2AP)

Structural information on the DIS kissing loop and the extended duplex show adjacent adenine residue 273 rotates out of the RNA helix upon ligand binding; therefore, an additional fluorescent construct was designed to determine if modification of this position could be utilized to monitor ligand binding. Construct DIS273(2AP), where 2AP was placed at position 273 instead of 272 was screened with the same ligands as above. In contrast to DIS272(2AP), a decrease in fluorescence intensity was observed for DIS273(2AP) upon ligand binding. Addition of paromomycin resulted in a 33% intensity decrease as compared to the unbound RNA (Figure 4.12b). A curve fit of the binding isotherm determined the EC_{50} value of paromomycin to be 0.9 ± 0.3 μM, similar to the affinity determined utilizing construct DIS272(2AP) (Figure 4.12d). A subsequent titration with neomycin resulted in a decrease in fluorescence of over 50% and produced an affinity of 0.4 ± 0.03 μM (Figure 4.13b and 4.13d). The EC_{50} value determined by DIS273(2AP) was within experimental error of the affinity obtained by DIS272(2AP). Similarly, titrations with

ribostamycin elicited a fluorescence intensity decrease of 50% compared to unbound DIS273(2AP) (Figure 4.14b). A moderate affinity of 17.2 ± 2.6 μM was determined from the dose-response curve fit (Figure 4.14d). The binding affinity value obtained was comparable with DIS272(2AP). Finally, neamine was titrated with the RNA construct and a similar decrease in emission (ca. 33%) was observed upon ligand binding, yielding an affinity of 12.3 ± 1.1 μM (Figure 4.16b and 4.16d)

Interestingly, butirosin A and the nucleobase conjugates **1a–d** did not elicit any changes in fluorescence when titrated with DIS273(2AP). In contrast, both apramycin and the conformationally constrained neomycin analog **2** induced moderate changes in fluorescence when titrated into DIS273(2AP). Apramycin and **2** elicited a 33% and 40% decrease in emission while yielding EC_{50} values of 6.1 ± 0.9 and 1.5 ± 0.1 μM, respectively (Figures 4.19b,d and 4.20b,d). These affinities correlate well with the values obtained using DIS272(2AP).

Table 4.1. Table of EC_{50} Values (μM) for DIS Ligand Binding.[a]

	DIS272(2AP)		DIS273(2AP)	
Ligand	**EC_{50}[b]**	**% Intensity Change**	**EC_{50}**	**% Intensity Change**
Neamine	21.0 ± 1.2	+ 75%	12.3 ± 1.1	– 30%
Neomycin B	0.6 ± 0.2[c]	+ 150%	0.4 ± 0.03	– 50%
Paromomycin	0.5 ± 0.2[c]	+ 300%	0.9 ± 0.3	– 25%
Ribstomycin	16.0 ± 4.0	+ 100%	17.2 ± 2.6	– 45%
Apramycin	7.6 ± 1.0	+ 235%	6.1 ± 0.9	– 35%
Butirosin A	5.5 ± 0.6	+ 100%	n.a.[e]	n.a.
1a[d]	2.6 ± 1.1	+ 170%	n.a.	n.a.
1d[d]	2.2 ± 0.1	+ 265%	n.a.	n.a.
2[d]	1.9 ± 0.2[c]	+ 175%	1.5 ± 0.1	– 35%

[a] See Section 4.13 for experimental details
[b] EC_{50} values determined by a dose-response curve fit with $R > 0.99$.
[c] Hill coefficients, n~2, where for all other binders, n~1.
[d] Trifluoroacetate salt used instead of free amine.
[e] not applicable (n.a.); binding is not observed with DIS273(AP) construct.

4.10 FLUORESCENT DIS CONSTRUCTS DETECT AND QUANTIFY LIGAND BINDING

Inspired from ligand binding studies with fluorescent prokaryotic A-site oligonucleotides,[70] DIS272(2AP) and DIS273(2AP) were designed to contain both the necessary motif for DIS–small molecule recognition and extended duplex formation, while retaining the conformationally flexibile adenine residues 272 and 273 (Figure 4.8). In both constructs, either adenine is replaced by 2AP, a fluorescent nucleobase analog that is highly sensitive to changes in its environment. Binding was anticipated to displace the 2AP fluorophore and induce fluorescence intensity changes due to the different solvent environment surrounding 2AP. Initially, for a standard titration, a low fluorescence signal was observed for unbound DIS272(2AP), suggesting 2AP at position 272 is significantly shielded within DIS extended duplex structure. A dramatic increase in intensity is observed upon ligand binding, possibly due to unstacking of 2AP and rotation toward the solvent-exposed exterior. The opposite is observed for construct DIS273(AP). In its unbound state, a significant fluorescence increase is observed suggesting an initial unstacked orientation for 2AP. Upon ligand binding, fluorescence intensity is moderately quenched caused by the partial stacking of adjacent A272 as it is rotated out of the duplex interior. Although two different photophysical changes are observed, similar behavior is observed for both constructs suggesting the same binding event is monitored.

X-ray crystallography studies have shown a strong structural resemblance between the prokaryotic A-site and HIV DIS.[37] Interestingly, our data show strong

similarity in solution behavior between the two RNA structures. Specifically, the bound and unbound states of both constructs exhibit similar fluorescence characteristics suggesting A272/A273 in the DIS and A1492/A1493 in the A-site, respectively, are dynamically similar. In an unbound state, bases A1492 and A272 are partially buried within the asymmetric bulge of their respective RNA structures, while in a fully bound state, the nucleotides are displaced from the helix and become more solvent exposed. Both oligonucleotides exhibit a fluorescence enhancement upon helical displacement. While A1493 and A273 are more solvent exposed in the absence of ligand, stacking interactions with adjacent purines result in decreased emission upon ligand binding. Since the prokaryotic A-site has been a paradigm structure for understanding RNA–ligand interactions, related structural principles may be applied to the HIV DIS system.

Although both fluorescent RNA constructs have been proven to be effective at monitoring ligand binding, for practical purposes, DIS272(2AP) is more advantageous since binding elicits a fluorescence enhancement and the construct detects a larger fluorescence intensity difference between bound and unbound states. To differentiate between specific and non-specific binding, a minimal fluorescence enhancement threshold was determined. The fluorescence enhancement elicited by neamine has been chosen as the minimal increase threshold for DIS binding (ca. 75%). Neamine is recognized as the minimal pharmacore of most aminoglycosides necessary for specific binding to the prokaryotic A-site and is the smallest known binder with reasonable affinity. Although paromomycin and neomycin are the highest affinity binders,

paromomycin induced a fluorescent enhancement of over 300% while neomycin with a similar EC_{50} values only elicited a 150% increase. Ribostamycin, a three-ring aminoglycoside, caused a 2-fold fluorescence enhancement upon DIS272(2AP) binding similar to neomycin but has nearly a 15-fold lower affinity for the HIV DIS. Comparison of EC_{50} values and per cent intensity changes, as summarized in Table 4.1, revealed that there is no correlation between affinity and fluorescence intensity change. This is not unprecedented, as similar characteristics have been observed with 2AP-labeled A-site oligonucleotides where neomycin elicits a minimal fluorescence response while being one of the highest affinity binders to that RNA.

A specific binding event was typically observed for the majority of molecules screened with DIS272(2AP) and DIS273(2AP); however, butirosin A, and compounds **1a** and **1d** only elicited a definitive binding event with DIS272(2AP). Without structural data of the RNA-ligand complex, it would be difficult to speculate on why DIS273(2AP) does not adequately monitor binding of these compounds. It is interesting to note that this anomaly occurs with compounds that have structural moieties conjugated to the core aminoglycoside scaffold. Butirosin A contains a 2-aminohydroxybutryryl side chain on N1 of 2-deoxystreptamine (2-DOS) while conjugates **1a** and **1d** possess a uracil residue at the 6′ and 5″ positions, respectively. It is plausible that these protrusions may abolish the environmental variations between the free and bound states of 2AP at position 273, resulting in minimal intensity changes for the otherwise environmentally sensitive fluorophore.

4.11 LIGAND SCREENING, QUANTIFICATION AND DISCOVERY

At the time Ennifar and colleagues reported the structural similarity between the A-site and HIV-1 DIS,[37] gel shift assays and footprinting experiments showed DIS RNA to be the target of neomycin-family aminoglycosides; however, the affinities of these antibiotics to the DIS had not been reported. Additionally, Dabrowiak *et. al.* have also shown aminoglycoside binding to DIS in footprinting studies for the entire 5′-end of the HIV genome. Although binding constants are reported, affinity values are preliminary due to competition with other binding sites.[81,82] A quick and efficient assay for screening ligands and quantifying binding has been lacking. From our binding assay, ligands derived from neomycin (both synthetic and semi-synthetic) have been quantified for binding by successfully eliciting fluorescence intensity changes. The striking similarity between the A-site and DIS allows comparison of binding affinities using the corresponding 2AP-based assay which have shown excellent correlation. Paromomycin, one of the best DIS and A-site binders has EC_{50} values of 0.5 ± 0.2 and 0.75 ± 0.05 μM , respectively. This suggests related binding modes in solution and corroborates recent structural analyses that show similar modes of binding for aminoglycosides of the neomycin/paromomycin family.

Despite the structural homology between the A-site and DIS, kanamycin and tobramycin family aminoglycosides screened did not elicit a definitive binding event. These results substantiate previous reports of limited binding in DIS footprinting studies. In the prokaryotic A-site, however, these classes of aminoglycosides bind with micromolar affinity.[83] The A-site is characterized by a non-canonical U1406-

U1495 base pair (Figure 4.6), known to be essential for binding of 4,6-disubstituted 2-DOS aminoglycosides. Substitution of U1406 for adenine has been shown to prevent binding of kanamycin and tobramycin to the A-site, while retaining affinity for the neomycin-family of aminoglycosides.[84] In contrast, the HIV DIS contains a Watson-Crick base pair at the same position (A282-U270*). This single difference in sequence composition allows for unprecedented binding selectivity for 4,5- over 4,6-disubstituted 2-DOS aminoglycosides (Figure 4.11a,c, respectively).

Screening of additional natural aminoglycosides led to the discovery of apramycin, a structurally unique and unrelated aminoglycoside, as a HIV-1 DIS binder. Binding of apramycin is most likely facilitated again by the Watson-Crick A282-U270* base pair that is absent in the prokaryotic A-site. In the A-site, a U1406A mutation that replaces the non-canonical U-U base pair actually confers a 4-fold hypersensitivity to apramycin, while being detrimental for 4,6-disubstituted 2-DOS aminoglycoside binding.[85] Through crystallographic and molecular modeling experiments,[86] apramycin binds the A-site in a different orientation compared to the neomycin-family of aminoglycosides, suggesting the U1406A mutation to be better accommodating for an A-site bound apramycin. These observations assist in explaining apramycin's moderate affinity for the HIV-1 DIS.

4.12 CONCLUSIONS

While the similarity between the HIV-1 DIS and the bacterial A-site in sequence, as well as in secondary and tertiary structure, has previously been

recognized, our solution studies shed light on the conformation and flexibility of the purine-rich bulge that represent the main ligand binding site. In particular, the parallel in relative fluorescence intensities and ligand induced changes in the emission pattern of 2AP272/2AP273 in DIS with respect to 2AP1492/2AP1493, the analogous positions in the A-site, suggest highly similar conformational behavior in solution. This intriguing observation may be pertinent to future attempts to design DIS-selective binders. On a positive note, the A-site has been recognized to be an encapsulating and therefore reasonably selective RNA target,[80] and attempts to design A-site selective ligands have met with some success.[87,88] This may translate to similar selectivity traits in the analogous HIV sequences, which is highly encouraging. A major challenge that remains to be addressed, however, is the inherently different pharmacokinetic properties that distinguish a bacterial A-site binder from a potential antiviral. In particular, aminoglycoside based drugs are known to have favorable uptake properties into bacterial cells, while displaying rather poor uptake into mammalian cells.[89] The opposite behavior might need to be established to ensure selectivity and proper uptake into HIV infected mammalian cells.

4.13 ACKNOWLEDGEMENT

Portions of Chapter 4 have been previously published in the article "Fluorescent HIV-1 Dimerization Initiation Site: Design, Properties, and Use for Ligand Discovery," *J. Am. Chem. Soc.* **2007**, *129*, 3257–3266 by V. Tam, D. Kwong

and Y. Tor. The dissertation author was the primary investigator and author of this paper.

4.14 EXPERIMENTAL

Native gel electrophoresis. Formation of the kissing loop and extended duplex conformations was assayed using native PAGE on a non-denaturing 15% polyacrylamide gel (19:1 acrylamide:bisacrylamide) and run in 0.25 × TBM buffer (2.25×10^{-2} M Tris·HCl, $\times 10^{-2}$ M boric acid, 1.0×10^{-4} M $MgCl_2$, pH 8.0) at 4 °C at 5 V/cm. Samples (3.3×10^{-6} M) were folded in a cacodylate buffer (2.0×10^{-2} M sodium cacodylate, pH 7.0, 2.5×10^{-2} M KCl, 2.0×10^{-3} M $MgCl_2$) by heating to 90 °C for 2 min and then flash cooling on ice for 2 min. Incubated at indicated times for kissing loop or extended duplex structures (See Figure 5). RNA was visualized by ethidium bromide staining (0.5 μg/mL)

Thermal UV-Visible Denaturing Experiments. Thermal denaturing profiles were obtained using a Beckman Coulter DU® 640 Spectrophotometer with a high performance temperature controller and micro auto six holder. All hybridizations and UV melting experiments were carried out in 2.0×10^{-2} M sodium cacodylate (pH 7.0), 2.5×10^{-2} M KCl, and 2.0×10^{-3} M $MgCl_2$. RNA samples (1.0×10^{-6} M) were pre-formed by heating to 90 °C for 2 min then flash cooling on ice for 2 min. The extended duplex structure was selectively formed by incubating at 55 °C for 2 h then slowly cooled to RT over 30 min. At the beginning of each thermal denaturing experiment, all samples were held at 20 °C for 5 min. Samples were placed in a stoppered 1.0 cm path length cell and a background spectra (buffer) was subtracted from each sample. All melting curves were obtained upon increasing the temperature from 20 °C to 90 °C at a rate of 1.0 °C per min and absorbance measurements at 260

nm were performed every minute. Beckman Coulter software (provided with T_m Analysis Accessory for DU® Series 600 Spectrometers) determined the extended duplex melting temperatures by calculating the first derivative of the melting profile.

Drug Molecules. Aminoglycosides were purchased as the sulfate salt from the following: neomycin B sulfate (MP Biomedicals), paromomycin sulfate (Sigma-Aldrich), apramycin sulfate (Research Products International Corp.), and ribstomycin sulfate (Sigma-Aldrich), except for the free base form of butirosin A (Sigma-Aldrich). All semi-synthetic aminoglycosides such as the constricted neomycin[61] and nucleoside-aminoglycoside conjugates[62] were studied as their TFA salt form. Sodium cacodylate buffer and all inorganic salts were purchased from Fisher (enzyme grade).

Purification of aminoglycosides. All aminoglycosides were purified and isolated as the free base form using DOWEX MONOSPHERE 550A (OH) anion exchange resin. A general procedure involved loading ca. 5 mL of aqueous aminoglycoside sulfate (~50 mg/mL) onto the resin column that had been equilibrated with water. The aminoglycoside was then eluted with 1.0 M NH_4OH. Fractions containing the free base (identified by ninhydrin staining on thin layer chromatography plates) were combined and lyophilized.

Preparation of RNA oligonucleotides. All DIS 23mer oligonucleotides were purchased from Dharmacon RNA Technologies (Lafayette, CO), deprotected as per manufacturer protocol, and lyophilized. The deprotected oligonucleotides were resuspended in water and purified using 20% denaturing PAGE. Excised gel slices were eluted overnight into a buffer containing 0.5 M NH_4OAc followed by ethanol

precipitation. After a 70% ethanol:water wash, the RNA pellets were dried and resuspended in water, quantified by UV absorbance, and their identities confirmed by MALDI-TOF mass spectrometry.

Fluorescence Binding Assay. For binding experiments, DIS constructs were refolded by heating a 25 μM solution of the 2AP-labeled DIS in a cacodylate binding buffer (2.0×10^{-2} M sodium cacodylate, pH 7.0, 2.5×10^{-2} M KCl, 2.0×10^{-3} M $MgCl_2$) to 90 °C for 2 minutes, followed by flash cooling on ice for 2 minutes. The kissing loop conformation was favored under these conditions. DIS constructs were incubated for 2 h at 55 °C and slowly cooled to room temperature to obtain the extended duplex. RNA constructs were quantified in 50 mM Na_3PO_4 (pH 7.5) with the following extinction coefficient (260 nm) values [in units of cm^{-1}/M^{-1}]: DIS-wt (ε = 225,500), DIS(U275C) (ε = 212,900), DIS272(2AP) (ε = 216,200), DIS273(2AP) (ε = 216,400). For all fluorescence measurements, a Perkin-Elmer LS50B fluorimeter was used, with an excitation slit width of 15 nm and an emission slit width of 20 nm. Upon excitation at 310 nm, the spectrum between 315 and 500 nm (scan rate of 300 nm/min) was recorded.

In a typical binding experiment, the fluorescence spectrum of a 120 μL solution of the buffer in the absence of any RNA or aminoglycoside was recorded. This spectral blank, for which only Raman scatter was observed, was subtracted from all subsequent spectra within each binding experiment. Following determination of the buffer blank, 5 μL of a 25 μM solution of refolded, 2AP-labeled DIS was added (final concentration is 1 μM), the solution mixed, and the spectrum again recorded.

Subsequent aliquots of 1 μL of an aqueous aminoglycoside solution (increasing concentrations from 1 μM to 10 mM) and the fluorescence spectrum was recorded after each aliquot until the 2-aminopurine fluorescence reach saturation. Over the entire range of aminoglycoside concentrations, the emission maximum at 370 nm did not significantly vary (<1.0 nm). No 'background emission' was observed near 370 nm. As small, concentrated volumes aminoglycosides were titrated in, the total volume of the sample never changed more than 10%.

EC_{50} values were calculated through a four-parameter nonlinear curve fit using Kaleidagraph with a floating Hill slope coefficient (equation 1) where fractional change in emission intensity was taken as the fraction of aminoglycoside bound (χ). The calculated fractional saturation to derive the effective concentration at which 50% of aminoglycoside (AG) is bound (EC_{50}) to the DIS RNA:

$$\chi_{bound} = y_{\min} + \left(\frac{y_{\max}[AG]^n}{[EC_{50}]^n + [AG]^n} \right) \qquad \text{(Equation 1)}$$

where *n* is the Hill coefficient or degree of cooperativity associated with binding.

4.15 REFERENCES

1. Fan, H.; Conner, R. F.; Villarreal, L. P., *The biology of AIDS*. 4th ed.; Jones and Bartlett: Boston, 2000; p xvii, 202.

2. Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 868–871.

3. Popovic, M.; Sarngadharan, M. G.; Read, E.; Gallo, R. C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **1984**, *224*, 497–500.

4. Garcia-Lerma, J. G.; Heneine, W. Resistance of human immunodeficiency virus type 1 to reverse transcriptase and protease inhibitors: genotypic and phenotypic testing. *J. Clin. Virol.* **2001**, *21*, 197–212.

5. Durant, J.; Clevenbergh, P.; Halfon, P.; Delgiudice, P.; Porsin, S.; Simonet, P.; Montagne, N.; Boucher, C. A.; Schapiro, J. M.; Dellamonica, P. Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. *Lancet* **1999**, *353*, 2195–2199.

6. Loveday, C.; Devereux, H.; Huckett, L.; Johnson, M. High prevalence of multiple drug resistance mutations in a UK HIV/AIDS patient population. *AIDS* **1999**, *13*, 627–628.

7. Schapiro, J. M.; Lawrence, J.; Speck, R.; Winters, M. A.; Efron, B.; Coombs, R. W.; Collier, A. C.; Merigan, T. C. Resistance mutations to zidovudine and saquinavir in patients receiving zidovudine plus saquinavir or zidovudine and zalcitabine plus saquinavir in AIDS clinical trials group 229. *J. Infect. Dis.* **1999**, *179*, 249–253.

8. Wainberg, M. A. A new source of resistance to HIV drugs. *J. Am. Med. Assoc.* **1999**, *281*, 2169–2170.

9. Vandamme, A. M.; Van Laethem, K.; De Clercq, E. Managing resistance to anti-HIV drugs: an important consideration for effective disease management. *Drugs* **1999**, *57*, 337–361.

10. Sarafianos, S. G.; Das, K.; Ding, J.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chem. Biol.* **1999**, *6*, R137–R146.

11. De Clercq, E. HIV resistance to reverse transcriptase inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.

12. Macchi, B.; Mastino, A. Pharmacological and biological aspects of basic research on nucleoside-based reverse transcriptase inhibitors. *Pharmacol. Res.* **2002**, *46*, 473–482.

13. Claessens, Y. E.; Chiche, J. D.; Mira, J. P.; Cariou, A. Bench-to-bedside review: severe lactic acidosis in HIV patients treated with nucleoside analogue reverse transcriptase inhibitors. *Crit. Care.* **2003**, *7*, 226–232.

14. Jost, S.; Bernard, M. C.; Kaiser, L.; Yerly, S.; Hirschel, B.; Samri, A.; Autran, B.; Goh, L. E.; Perrin, L. A patient with HIV-1 superinfection. *N. Engl. J. Med.* **2002**, *347*, 731–736.

15. Goulder, P. J.; Walker, B. D. HIV-1 superinfection--a word of caution. *N. Engl. J. Med.* **2002**, *347*, 756–758.

16. *UNAIDS: 2006 Report on the global AIDS epidemic.*; Available at http://www.unaids.org.

17. Moore, R. D.; Chaisson, R. E. Natural history of HIV infection in the era of combination antiretroviral therapy. *AIDS* **1999**, *13*, 1933–1942.

18. Baliant, G. A. Antiretroviral therapeutic possibilities for human immunodeficiency virus/acquired immunodeficiency syndrome. *Pharm. Therap.* **2001**, *89*, 17–27.

19. Chen, L. F.; Hoy, J.; Lewin, S. R. Ten years of highly active antiretroviral therapy for HIV infection. *Med. J. Aust.* **2007**, *186*, 146–151.

20. Vaishnav, Y. N.; Wong-Staal, F. The biochemistry of AIDS. *Annu. Rev. Biochem.* **1991**, *60*, 577–630.

21. Emerman, M.; Malim, M. H. HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. *Science* **1998**, *280*, 1880–1884.

22. Frankel, A. D.; Young, J. A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **1998**, *67*, 1–25.

23. Luedtke, N. W.; Tor, Y. A Novel Solid-Phase Assembly for Identifying Potent and Selective RNA Ligands. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 1788–1790.

24. Luedtke, N. W.; Tor, Y. Fluorescence-based methods for evaluating the RNA affinity and specificity of HIV-1 Rev-RRE inhibitors. *Biopolymers* **2003**, *70*, 103–119.

25. Rosen, C. A.; Pavlakis, G. N. Tat and Rev: positive regulators of HIV gene expression. *AIDS* **1990**, *4*, 499–509.

26. Green, M. R., Molecular Mechanisms of Tat and Rev. In *AIDS Research Reviews III*, Koff, W. C., Wong-Staal, F., Kennedy, R. C., Ed. Marcel Dekker, Inc.: New York, 1993; pp 41–55.

27. Zapp, M. L.; Stern, S.; Green, M. R. Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **1993**, *74*, 969–978.

28. Nelson, D. L. C., Michael M., *Lehninger Principles of Biochemistry*. 3rd ed.; Worth Publishers: 2000; p 1152.

29. Pollard, V. W.; Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **1998**, *52*, 491–532.

30. Berkhout, B.; van Wamel, J. L. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J. Virol.* **1996**, *70*, 6723–6732.

31. Russell, R. S.; Liang, C.; Wainberg, M. A. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **2004**, *1*, 23–36.

32. Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5572–5577.

33. Fu, W.; Rein, A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **1993**, *67*, 5443–5449.

34. Paillart, J. C.; Dettenhofer, M.; Yu, X. F.; Ehresmann, C.; Ehresmann, B.; Marquet, R. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* **2004**, *279*, 48397–48403.

35. Opar, A. New HIV drug classes on the horizon. *Nat. Rev. Drug Discov.* **2007**, *6*, 258–259.

36. Lalezari, J. P., Henry, K., O'Hearn, M., Montaner, J. S. G., Piliero, P. J., Trottier, B., Walmsley, S., Cohen, C., Kuritzkes, D. R., Eron, J. J., Chung, J., DeMasi, R., Donattaci, L., Drobnes, C., Delehanty, J., Salgo, M. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. *N. Engl. J. Med.* **2003**, *348*, 2175–2185.

37. Ennifar, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C.; Dumas, P.; Walter, P. HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. *J. Biol. Chem.* **2003**, *278*, 2723–2730.

38. Robinson, W. S.; Robinson, H. L.; Duesberg, P. H. Tumor virus RNA's. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 825–834.

39. Kung, H. J.; Bailey, J. M.; Davidson, N.; Nicolson, M. O.; McAllister, R. M. Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol* **1975**, *16*, 397–411.

40. Khandjian, E. W., Meric, C. A procedure for Northern blot analysis of native RNA. *Anal. Biochem.* **1986**, *159*, 227–232.

41. Darlix, J. L., Gabus, C., Nugeyre, M. T., Clavel, F., Barre-Sinoussi, F. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **1990**, *216*, 689–699.

42. Laughrea, M., Jetté, L., Mak, J., Kleiman, L., Liang, C., Wainberg, M. A. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **1997**, *71*, 3397–3406.

43. Clever, J. L., Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization and encapsidation. *J. Virol.* **1997**, *71*, 3407–3414.

44. Russell, R. S., Roldan, A., Detorio, M., Hu, J., Wainberg, M. A., Liang, C. Effects of a single amino acid substitution within the p2 region of human immunodeficiency virus type 1 on packaging of spliced viral RNA. *J. Virol.* **2003**, *77*, 12986–12995.

45. Paillart, J. C.; Shehu-Xhilaga, M.; Marquet, R.; Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2004**, *2*, 461–472.

46. Hu, W. S.; Temin, H. M. Retroviral recombination and reverse transcription. *Science* **1990**, *250*, 1227–1233.

47. Hu, W. S.; Temin, H. M. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 1556–1560.

48. Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4945–4949.

49. Laughrea, M.; Jette, L. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **1994**, *33*, 13464–13474.

50. Ennifar, E.; Paillart, J. C.; Bernacchi, S.; Walter, P.; Pale, P.; Decout, J. L.; Marquet, R.; Dumas, P. A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. *Biochimie* **2007**, *in press*.

51. Mujeeb, A.; Clever, J. L.; Billeci, T. M.; James, T. L.; Parslow, T. G. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.* **1998**, *5*, 432–436.

52. Muriaux, D.; De Rocquigny, H.; Roques, B. P.; Paoletti, J. NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem* **1996**, *271*, 33686–33692.

53. Muriaux, D.; Girard, P. M.; Bonnet-Mathoniere, B.; Paoletti, J. Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J. Biol. Chem* **1995**, *270*, 8209–8216.

54. Muriaux, D.; Fosse, P.; Paoletti, J. A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* **1996**, *35*, 5075–5082.

55. Huthoff, H.; Berkhout, B. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry* **2002**, *41*, 10439–10445.

56. Laughrea, M.; Jette, L.; Mak, J.; Kleiman, L.; Liang, C.; Wainberg, M. A. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **1997**, *71*, 3397–3406.

57. Shen, N.; Jette, L.; Liang, C.; Wainberg, M. A.; Laughrea, M. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.* **2000**, *74*, 5729–5735.

58. Ennifar, E.; Paillart, J. C.; Bodlenner, A.; Walter, P.; Weibel, J. M.; Aubertin, A. M.; Pale, P.; Dumas, P.; Marquet, R. Targeting the dimerization initiation

site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res.* **2006**, *34*, 2328–2339.

59. Ennifar, E.; Dumas, P. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J. Mol. Biol.* **2006**, *356*, 771–782.

60. Yoshizawa, S.; Fourmy, D.; Puglisi, J. D. Recognition of the codon-anticodon helix by ribosomal RNA. *Science* **1999**, *285*, 1722–1725.

61. Kaul, M.; Barbieri, C. M.; Pilch, D. S. Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *J. Am. Chem. Soc.* **2006**, *128*, 1261–1271.

62. Ogle, J. M.; Brodersen, D. E.; Clemons, W. M., Jr.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001**, *292*, 897–902.

63. Yoshizawa, S.; Fourmy, D.; Puglisi, J. D. Structural origins of gentamicin antibiotic action. *EMBO J.* **1998**, *17*, 6437–6448.

64. Fourmy, D.; Yoshizawa, S.; Puglisi, J. D. Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. *J. Mol. Biol.* **1998**, *277*, 333–345.

65. Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996**, *274*, 1367–1371.

66. Brodersen, D. E.; Clemons, W. M., Jr.; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **2000**, *103*, 1143–1154.

67. Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340–348.

68. Vicens, Q.; Westhof, E. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **2001**, *9*, 647–658.

69. Kaul, M.; Barbieri, C. M.; Pilch, D. S. Fluorescence-based approach for detecting and characterizing antibiotic-induced conformational changes in ribosomal RNA: comparing aminoglycoside binding to prokaryotic and

eukaryotic ribosomal RNA sequences. *J. Am. Chem. Soc.* **2004**, *126*, 3447–3453.

70. Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. Monitoring molecular recognition of the ribosomal decoding site. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 3177–3182.

71. Rist, M. J.; Marino, J. P. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002**, *41*, 14762–14770.

72. Previous studies with other minimized RNA constructs have been utilized to mimic the kissing loop and extended duplex structure of the full length viral genome. A 19 nt RNA consisting of a self-complementary loop and its flanking stem was shown to conform to a kissing loop structure by NMR but not an extended duplex; see (a) Dardel, F.; Marquet, R.; Ehresmann, C.; Ehresmann, B.; Blanquet, S. *Nucleic Acids Res.* **1998**, *26*, 3567–3571; however, the addition of two base-pairs to the stem reportedly stabilized the 23 nt construct to allow formation of a kissing loop as observed by 2D NMR; (b) see Mujeeb, A.; Clever, J.; Billeci, T. M.; James, T. L.; Parslow, T. G. *Nat. Struct. Biol*. **1998**, *5*, 432–436. A 23 nt extended duplex was later reported by x-ray crystallography; see (c) Ennifar, E.; Yusupov, M.; Walter, P.; Marquet, R.; Ehresmann, B.; Ehresmann, C.; Dumas, P. *Structure* **1999**, *7*, 1439–1449. Both structures appear to mimic the DIS secondary structure found in viral particles.

73. Jossinet, F.; Paillart, J. C.; Westhof, E.; Hermann, T.; Skripkin, E.; Lodmell, J. S.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA* **1999**, *5*, 1222–1234.

74. Lodmell, J. S.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Structure and dimerization of HIV-1 kissing loop aptamers. *J. Mol. Biol.* **2001**, *311*, 475–490.

75. Mihailescu, M. R.; Marino, J. P. A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc. Natl. Acad. Sci. U.S.A*. **2004**, *101*, 1189–1194.

76. Thermoydynamic studies on the stability of DIS type RNA loop-loop interactions have been reported. See: Weixlbaumer, A.; Werner, A.; Flamm, C.; Westhof, E.; Schroeder, R. Nucleic Acids Res. 2004, 32, 5126–5133; Lorenz, C.; Piganeau, N.; Schroeder, R. Nucleic Acids Res. 2006, 34, 334–342.

77. Rist, M.; Marino, J. Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. *Nucleic Acids Res.* **2001**, *29*, 2401–2408.

78. Bradrick, T. D.; Marino, J. P. Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *RNA* **2004**, *10*, 1459–1468.

79. Blount, K. F., Tor, Y. Synthesis and RNA target selectivity of novel nucleobase–aminoglycoside conjugates. *ChemBioChem* **2006**, *7*, 1612–1621.

80. Blount, K. F.; Zhao, F.; Hermann, T.; Tor, Y. Conformational constraint as a means for understanding RNA-aminoglycoside specificity. *J. Am. Chem. Soc.* **2005**, *127*, 9818–9829.

81. McPike, M. P.; Sullivan, J. M.; Goodisman, J.; Dabrowiak, J. C. Footprinting, circular dichroism and UV melting studies on neomycin B binding to the packaging region of human immunodeficiency virus type-1 RNA. *Nucleic Acids Res.* **2002**, *30*, 2825–2831.

82. McPike, M. P.; Goodisman, J.; Dabrowiak, J. C. Footprinting and circular dichroism studies on paromomycin binding to the packaging region of human immunodeficiency virus type-1. *Bioorg. Med. Chem.* **2002**, *10*, 3663–3672.

83. Griffey, R. H.; Hofstadler, S. A.; Sannes-Lowery, K. A.; Ecker, D. J.; Crooke, S. T. Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10129–10133.

84. Pfister, P.; Hobbie, S.; Vicens, Q.; Bottger, E. C.; Westhof, E. The molecular basis for A-site mutations conferring aminoglycoside resistance: relationship between ribosomal susceptibility and X-ray crystal structures. *ChemBioChem* **2003**, *4*, 1078–1088.

85. Recht, M. I.; Puglisi, J. D. Aminoglycoside resistance with homogeneous and heterogeneous populations of antibiotic-resistant ribosomes. *Antimicrob. Agents Chemother.* **2001**, *45*, 2414–2419.

86. Han, Q.; Zhao, Q.; Fish, S.; Simonsen, K. B.; Vourloumis, D.; Froelich, J. M.; Wall, D.; Hermann, T. Molecular recognition by glycoside pseudo base pairs and triples in an apramycin–RNA complex. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 2694–2700.

87. Zhou, Y. F.; Gregor, V. E.; Sun, Z. X.; Ayida, B. K.; Winters, G. C.; Murphy, D.; Simonsen, K. B.; Vourloumis, D.; Fish, S.; Froelich, J. M.; Wall, D.; Hermann, T. Structure-guided discovery of novel aminoglycoside mimetics as

antibacterial translation inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49*, 4942–4949.

88. Hermann, T.; Tor, Y. RNA as a target for small-molecule therapeutics. *Exper. Opin. on Therapeutic Patents* **2005**, *15*, 49–62.

89. Luedtke, N. W.; Carmichael, P.; Tor, Y. Cellular uptake of aminoglycosides, guanidinoglycosides, and poly-arginine. *J. Am. Chem. Soc.* **2003**, *125*, 12374–12375.