NUCLEIC ACID ASSEMBLY, POLYMERIZATION, AND LIGAND BINDING

A Dissertation
Presented to
The Academic Faculty

by

Aaron E. Engelhart

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
December 2010

NUCLEIC ACID ASSEMBLY, POLYMERIZATION,

AND LIGAND BINDING

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Stefan France
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. David G. Lynn
Department of Chemistry
Emory University

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Date Approved: August 24, 2010

ACKNOWLEDGEMENTS

I would like to thank my advisor, Nicholas Hud. I could not have asked for a better mentor professionally and scientifically, and he has been a joy to interact with personally. His patience, insightfulness, and ability to develop a model to explain data and an experiment to confirm it are traits I strive to emulate. He has made me a better scientist. In addition, his boundless scientific curiosity has resulted in countless enjoyable conversations about new ideas to examine in the lab – which, in turn, has helped me learn how to figure out which ideas are worth chasing after.

My committee has provided valuable insight and advice throughout my time here. Loren Williams, Stefan France, David Lynn, and Roger Wartell have frequently, upon seeing my data, suggested another aspect of chemistry to which it might apply (and which I had not considered). Each of these people has, at one time or another, made a suggestion that helped me solve a tough problem in the lab. They are also some of my favorite people to discuss ideas with; I have greatly enjoyed our scientific conversations.

I would like to thank my collaborators Tom Morton and Michael Lieber. I thank Ram Krishnamurthy for many enjoyable discussions and insights. I would like to thank Joshy Joseph and Mitchell Maddox for their help with oligonucleotide and phosphoramidite synthesis.

I would like to thank my labmates. Eric Horowitz, Heather Bean, Özgül Persil Çetinkol, Swapan Jain, Cathy Santai, Ragan Buckley, Brian Cafferty, Tim Lenz, Irena Mamajanov, Seth Lilavivat, Tumpa Sarkar, Denise Enekwa, Shreyas Athavale, Sara Walker, and Jim Collins have all created an enjoyable place to work, and they have been good friends and colleagues to me throughout my time here. I also would like to thank

my downstairs neighbors Chiaolong Hsiao, Derrick Watkins, T Maehigashi, Srividya Mohan, and Taylor Updegrove.

Georgia Tech has some amazing undergraduate students, and I have been lucky enough to work with a few. Michael Smith, Michael Chen, Kenny Taylor, Kaycee Quarles, and Brenda Morales-Pico have been a real help to me in my research, and they have also taught me about how to be a better teacher.

Finally, I thank my parents, brothers, and friends. They have been a real help to me in staying sane when things didn't work in the lab – often repeatedly. I also especially thank my parents for their love and support in encouraging me to pursue my interests.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

A	Adenine
AMBER	Assisted Model Building with Energy Refinement
APBS	Adaptive Poisson-Boltzmann Solver
aza3	Azacyanine 3
bp	Base Pair
C	Cytosine
CD	Circular Dichroism
d	deoxyribo-
DES	Deep Eutectic Solvent
DNA	Deoxyribonucleic Acid
D_2O	Deuterium Oxide
EtBr	Ethidium Bromide
G	Guanine
HPLC	High-Performance Liquid Chromatography
ImCN	N-cyanoimidazole
ITC	Isothermal Titration Calorimetry
LBHB	Low-Barrier Hydrogen Bond
M	Molar
m	Molal
MD	Molecular Dynamics
mdeg	Millidegrees
MES	2-Morpholinoethanesulfonic Acid

NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
nt	Nucleotide
p	phosphoryl-
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Databank
ppm	Parts Per Million
r	ribo-
RNA	Ribonucleic Acid
RTIL	Room-Temperature Ionic Liquid
SPR	Surface Plasmon Resonance
T	Thymine
T	Temperature
T_M	Melting Temperature
TEA	Triethylamine
TEAA	Triethylammonium acetate
UV-Vis	UV-Visible

SUMMARY

In the past 30 years, the discovery of capabilities of nucleic acids far beyond their well-known information-bearing capacity has profoundly influenced our understanding of these polymers. The discovery by the Cech and Altman labs that nucleic acids could perform catalytic functions, coupled with the Gold and Szostak groups' demonstration of the *de novo* evolution of nucleic acids that bind arbitrary ligands, has resulted in a proliferation of newfound roles for these molecules. Nucleic acids have found utility in both engineered systems, such as aptamer therapeutics, as well as in newly appreciated roles in extant organisms, such as riboswitches. As a result of these discoveries, many have pondered the potential importance of the dual (catalytic and informational) roles of nucleic acids in early evolution. A high-yielding synthetic route for the nonenzymatic polymerization of nucleic acids, based on the aqueous self-assembly of their components, would provide a powerful tool in nucleic acid chemistry, with potential utility in prebiotic and contemporary nucleic acid systems alike – however, such a route remains elusive.

In this thesis, I describe several steps towards such a synthetic route. In these systems, a nucleic-acid binding ligand drives the assembly of short DNA and RNA duplexes, promoting the production of long nucleic acid polymers, while suppressing the production of short, cyclic species. Additionally, the use of a reversible covalent linkage allows for the production of long polymers, as well as the incorporation of previously cyclized products into such polymers. I also report several explorations of novel base pairings, nucleic acid-ligand interactions, and nucleic acid-ion interactions that have informed our studies of self-assembling nucleic acid systems.

CHAPTER 1

INTRODUCTION[a]

1.1. Overview

In contemporary life, nucleic acids provide the amino acid sequence information required for protein synthesis, while protein enzymes carry out the catalysis required for nucleic acid synthesis. This mutual dependence has been described as a 'chicken-or-the-egg' dilemma concerning which came first. However, requiring that these biopolymers appeared strictly sequentially may be an overly restrictive preconception – nucleic acids and *noncoded* peptides may have arisen independently and later become dependent on each other. Nevertheless, the requirements for the chemical emergence of life would be simplified if one polymer was initially able to store and transfer information (i.e., genetics) as well as perform selective chemical catalysis (i.e., enzymatic activity) – two essential features of life.

The discovery of catalytic RNA molecules in the early 1980s [1, 2] created widespread interest in an earlier proposal [3-5] that nucleic acids were the first biopolymers of life, as nucleic acids transmit genetic information and could have once been responsible for catalyzing a wide range of reactions. The ever-increasing list of processes that involve RNA in contemporary life continues to strengthen this view.[6, 7] Furthermore, the rule-based one-to-one pairing of complementary bases in a Watson-

---

[a] This chapter was adapted from previously published work. (Engelhart, AE; Hud, NV. "Primitive Genetic Polymers." Cold Spring Harb. Perspect. Biol. (2010) DOI: 10.1101/cshperspect.a002196.). It is reproduced with permission. © Cold Spring Harbor Laboratory Press.

Crick duplex provides a robust mechanism for information transfer during replication that could have been operative from the advent of oligonucleotides. In contrast, there is no obvious and general mechanism by which the amino acid sequence of a polypeptide can be transferred to a new polypeptide as part of a replication process.

If we accept that nucleic acids must have appeared without the aid of coded proteins, we are still faced with the question of how the first nucleic acid molecules came to be. Synthesizing linear phosphate polyesters *de novo* in aqueous solution is an inherently difficult proposition. The formation of a phosphodiester bond requires dehydration – a difficult task in 55 M water. This has lead many investigators to employ monomer species with activated phosphate groups. Activated phosphate esters come with another set of problems – these species, when hydrolyzed or incorporated into polymer, form very kinetically stable products. While this is certainly desirable in contemporary life, it can be problematic in chemical polymerization. In the case of monomer polymerization, Ferris and coworkers have demonstrated that activated imidazolide derivatives of adenosine nucleotides will polymerize on mineral surfaces, but a large proportion of the monomer is irreversibly incorporated into cyclic dimers.[8] In attempting to polymerize tiled half-complementary hexanucleotides into higher oligo- and polymers by carbodiimide activation, Kawamura and Okamoto also observed substantial amounts of cyclized starting material.[9] In both of these examples, a significant amount of starting material is irreversibly incorporated into an undesired side product. In addition to being a nuisance in the lab, these side products would have amounted to a fatal and committed step in the synthesis of a nascent proto-RNA. This problem illustrates a difficulty in nonenzymatic polymerization that must be taken into

account when considering how the nature of the synthetic routes to and structural identities of early genetic polymers: irreversible linkages are adaptive for an informational polymer *only when mechanisms exist to make them conditionally reversible* (so as to allow proofreading).

The difficulties associated with irreversible linkage formation confound the efficient prebiotic chemical synthesis of phosphate esters. The polymerization of activated substrates yields mixed regioisomers and poor replication fidelity due to strand cyclization, base misincorporation, and premature product chain termination.[8-10] Work by Usher and colleagues has provided encouraging evidence that the 2′,5′ linkage is more labile than the contemporary 3′,5′ regioisomer, suggesting that thermodynamic selection may be helpful in addressing a subset of the aforementioned problems with phosphodiester linkages.[11] However, all these difficulties can be circumvented by the use of a linkage chemistry that initially forms a low-energy, reversible bond, allowing for selection of the thermodynamic product (i.e., in the case of RNA, the Watson-Crick base-paired product).

It is possible that an RNA-like polymer (or proto-RNA) containing such linkages was used by the earliest forms of life. Just as the deoxyribose sugar of DNA was likely the product of Darwinian evolution (selected for the hydrolytic stability it provides this long-lived biopolymer), so, too, might the RNA backbone have been refined by evolution. In this scenario, a proto-RNA is more likely to have spontaneously formed than RNA, because a proto-RNA could have had more favorable *chemical* characteristics (e.g., greater availability of precursors and ease of assembly), but such a polymer was eventually replaced, through evolution, by RNA (potentially after several incremental

changes), based on *functional* characteristics (e.g., nucleoside stability, versatility in forming catalytic structures). Thus, contemporary RNA may possess chemical traits that, while optimally suited for contemporary life, may have been ill-suited for the earliest biopolymers, with the converse being true for proto-RNA.

One remarkable example of the power of reversible linkages was provided by Lynn and coworkers, who demonstrated that 5′-deoxy,5′-amino, 3′-deoxy, 3′-formylmethyl-dT could polymerize via reductive amination in aqueous solution.[12] Polymerization occurred only in the presence of a d(A_8) template; without a template, no oligomer product was detected. Interestingly, only linear polymers were observed. Here, polymerization decreases the entropic cost of hybridization, driving base pairing (which is disfavored at the mononucleotide level in water [13]). In turn, hybridization increases the local concentration of amines and aldehydes, driving the formation of imine linkages (which are otherwise disfavored in 55 M water). Additionally, the presence of an amination-active imine linkage only when the reactants were hydribrized on the template selected for a linear polymer topology. In this system, the subtle interplay between base pairing and linkage formation drives the simultaneous formation of reversible covalent and non-covalent interactions which are otherwise disfavored. Several other intriguing examples of self-assembling reversible polymers have been demonstrated, particularly by the Lehn laboratory.[14-16]

These examples illustrate the great potential utility of (periodic) reversibility of the covalent bonds that connected a proto-RNA. In the early stages of life, it would have been highly advantageous for chemical building blocks to be available for reuse. Mechanisms for recycling are easy to imagine, if the covalent bonds that joined the

molecular components of proto-biopolymers could have been formed, broken and re-formed repeatedly. Such a process would have allowed proto-biopolymers to be created that were thermodynamically favored structures [12, 17, 18], and for "errors" in synthesis to be corrected (e.g., replacement of non- or mispaired nucleobases with pairing ones). Proto-RNA monomers might have even been repeatedly recycled into polymers with different nucleotide sequences (and corresponding functions) as survival pressures changed.[18] Furthermore, without some regular cycle by which RNA or a proto-RNA could have been depolymerized, any monomer incorporated into a polymer that was not functionally useful would be wasted – a condition under which there simply may have not been enough material to get life started.

An example is illustrative: If the earliest effective ribozyme was 50nt in length (similar to the hammerhead ribozyme), and the earliest genetic code contained C, U, A, and G , this ribozyme would be one of $4^{50} \approx 10^{30}$ possible sequences. A collection of one molecule of every possible 50mer would require a carbon mass of around 10^7 kg. Assuming each polymer was produced in equal yield, and by an irreversible process, production of 1 picomole of the ribozyme would require 10^{19} kg of carbon (roughly the total mass of carbon in the Earth's crust).

Any nucleic acid polymerization system – chemical or biological – should afford high-fidelity sequence generation. Additionally, a means by which to salvage monomers from erroneous products would be highly advantageous – if not absolutely necessary. In this thesis, I describe several systems that provide potential means to address these problems, as well as several nucleic acids systems of particular relevance to contemporary life. In Chapter 2, I demonstrate how a nucleic-acid binding ligand can

drive the assembly of short DNA and RNA duplexes, promoting the production of long nucleic acid polymers, while suppressing the production of undesired short, cyclic species. In Chapter 3, I demonstrate the utility of ligand binding in assembling a novel backbone containing a reversible covalent linkage. This backbone, furthermore, allows for the production of long polymers, as well as the incorporation of (previously cyclized) products into long polymers. In Chapters 4 and 5, I report the binding of a novel nucleic acid ligand to a G-quadruplex nucleic acid structure, and I describe the assembly of another non-Watson-Crick structure. Finally, in Chapters 6 and 7, I describe studies of nucleic acid secondary structures in high-salt, low-water activity conditions and consider the applicability of these systems to nucleic acid polymerization.

1.2. References

1. Kruger, K., et al., *Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**: p. 147-157.
2. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**: p. 849-857.
3. Woese, C., *The evolution of the genetic code.*, in *The genetic code*. 1967, Harper & Row: New York. p. 179-195.
4. Crick, F.H.C., *The origin of the genetic code.* J. Mol. Biol., 1968. **38**: p. 367-379.
5. Orgel, L.E., *Evolution of the genetic apparatus.* J. Mol. Biol., 1968. **38**: p. 381-393.
6. Gesteland, R. and J.F. Atkins, eds. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World.* 3rd ed. 2006, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 768.
7. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nature Reviews Molecular Cell Biology, 2004. **5**(6): p. 451-463.
8. Miyakawa, S., et al., *Studies in the mineral and salt-catalyzed formation of RNA oligomers.* Origins of Life and Evolution of Biospheres, 2006. **36**(4): p. 343-361.

9. Kawamura, K. and F. Okamoto, *Cyclization and dimerization of hexanucleotides containing guanine and cytosine with water-soluble carbodiimide.* Viva Origino, 2001. **29**(4): p. 162-167.

10. Joyce, G.F. and L.E. Orgel, *Prospects for understanding the origin of the RNA world*, in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. p. 49-77.

11. Usher, D.A. and A.H. McHale, *Hydrolytic stability of helical RNA - Selective advantage for natural 3',5'-bond.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(4): p. 1149-1153.

12. Li, X., et al., *DNA-catalyzed polymerization.* Journal of the American Chemical Society, 2002. **124**(5): p. 746-747.

13. Yakovchuk, P., E. Protozanova, and M.D. Frank-Kamenetskii, *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix.* Nucleic Acids Research, 2006. **34**(2): p. 564-574.

14. Sreenivasachary, N. and J.M. Lehn, *Gelation-driven component selection in the generation of constitutional dynamic hydrogels based on guanine-quartet formation.* Proc. Natl. Acad. Sci. USA, 2005. **102**(17): p. 5938-5943.

15. Sreenivasachary, N., et al., *DyNAs: Constitutional dynamic nucleic acid analogues.* Chemistry-a European Journal, 2006. **12**(33): p. 8581-8588.

16. Sreenivasachary, N. and J.M. Lehn, *Structural selection in G-quartet-based hydrogels and controlled release of bioactive molecules.* Chemistry-an Asian Journal, 2008. **3**: p. 134-139.

17. Hud, N.V., et al., *Addressing the problems of base pairing and strand cyclization in template-directed synthesis - A case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA.* Chemistry & Biodiversity, 2007. **4**(4): p. 768-783.

18. Ura, Y., et al., *Self-assembling sequence-adaptive peptide nucleic acids.* Science, 2009. **325**(5936): p. 73-77.

CHAPTER 2

LIGAND-DRIVEN NUCLEIC ACID POLYMERIZATION[b]

2.1. Introduction

Over the past two decades, significant evidence has been presented in support of the RNA world hypothesis, which proposes that RNA polymers predated coded proteins in early life. [1, 2] Current support for this hypothesis includes the fact that contemporary life still uses RNA as an informational polymer and in chemical catalysis.[3] The ability of RNA to catalyze reactions is exemplified by natural and artificial ribozymes that promote a wide variety of chemical reactions [4], as well as the observation that the catalytic core of the ribosome is comprised of RNA.[5] Despite the attractiveness of the RNA world as a hypothetical stage of early life, it remains unclear how RNA (or a predecessor of RNA [6-11]) would initially have been synthesized without the aid of protein enzyme catalysis.

Several distinct proposals have been presented for the abiotic origin of the first RNA polymers.[10, 12-17] Perhaps the most notable is that of Ferris and coworkers, in which mineral surfaces are used to locally concentrate and promote the polymerization of ligation-activated mononucleotides, an approach that allows formation of single-stranded RNA strands up to *ca.* 50 nucleotides in length. [18] Additionally, Sawai *et al.* demonstrated the use of UO_2^{2+} in the solution-state polymerization of activated

---

[b] This chapter was adapted from previously published work. (Horowitz, EH; Engelhart, AE; Quarles, KA; Smith, MW; Chen, MC; Lynn, DG. "Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world." Proc. Nat. Acad. Sci. USA (2010) 107:5288-5293.) It is reproduced with permission.

monomers into linear and cyclic oligomers up to decanucleotides [19, 20], and Steltsov *et al.* have shown that the tripeptide trivaline facilitated the ligation of d(pGTT). [21] However, it is not clear how any of these mechanisms alone would have selectively produced nucleic acids containing only bases capable of forming Watson–Crick base pairs from a prebiotic chemical inventory that likely contained a complex mixture of molecules.[22] Earlier work by Orgel and coworkers demonstrated that oligonucleotides (of rather restricted nucleotide sequence) can serve as templates for the polymerization of activated mononucleotides in solution, resulting in duplexes with Watson–Crick base pairs.[23-25] However, it remains an open question precisely how the first polymers of RNA (or a chemically related proto-RNA) could have assembled from mononucleotides, or even short oligonucleotides, without pre-existing templates.

Strand cyclization is a formidable problem facing most proposed systems for the prebiotic synthesis of nucleic acid polymers. Specifically, short ligation-activated oligonucleotides (i.e., di- to octanucleotides) undergo efficient intramolecular ligation (i.e., cyclization), limiting their potential for polymerization.[26] It is a general principle in polymer chemistry that the length of polymers formed by irreversible reactions can be greatly limited by cyclization.[27] When a nascent polymer becomes sufficiently long to sample conformations that allow for intramolecular bond formation, cyclization occurs, and polymer growth ceases. A polymer's persistence length (which relates to its rigidity) largely determines the length at which a growing polymer favors cyclization over continued growth. The persistence length of single stranded nucleic acids is only 3 to 6 nucleotides [28], whereas the persistence length of a base-paired duplex is 150 to 300 base pairs.[29] In order to form appreciable equilibrium amounts of base-paired duplexes,

oligonucleotides must be *ca*. five or more residues in length [30], even at temperatures near the freezing point of water and at moderately high oligonucleotide concentrations (10^{-4}-10^{-3} M). Thus, given a pool of chemically activated di-, tri- and tetranucleotides, one would predict that such short oligonucleotides would cyclize efficiently. This prediction is borne out by experiment, including model prebiotic reactions in which activated mononucleotides are condensed on a mineral surface, where cyclic products can have yields that are still comparable to those of linear products.[12] Even activated hexanucleotides of 100% GC content, designed to assemble into polymeric structures, have been shown to primarily form cyclic products.[26] Clearly, strand cyclization would have inhibited the prebiotic production of nucleic acid polymers, unless a mechanism existed to increase the persistence length of oligonucleotides during polymerization.

Here, we report that molecules that intercalate the base pairs of nucleic acid duplexes can circumvent the problem of oligonucleotide cyclization. Specifically, we demonstrate that certain intercalators (or "molecular midwives" [13]) promote the coupling of activated tetranucleotides into long duplex polymers, while in the absence of intercalators, only short cyclic oligonucleotides are formed. Further, we show that intercalator-mediated polymer formation is ligand and base-pair specific; size matching is required between the intercalator and the base pairs of a duplex. These data lend support to the hypothesis that the original structure of nucleic acids, including that of the base pairs, could have been templated by intercalators.[13, 31]

2.2. Experimental Procedures

2.2.1. Materials

Oligonucleotides were purchased from Oligos, Etc. (Wilsonville, OR) or IDT (Coralville, IA) and HPLC purified before use. Ethidium bromide (Fisher Scientific), proflavine hemisulfate (Sigma), and coralyne chloride (Sigma) were used as received. aza3 was synthesized as previously reported.[32] *N*-cyanoimidazaole was purchased from Toronto Research Chemicals (North York, ON, Canada).

2.2.2. Watson–Crick Paired Chemical Ligations

Ligation reactions were 200 μM in oligonucleotide strand, 10 mM triethylammonium MES (pH 6), 5 mM $MnCl_2$. Ligation reactions in the presence of various divalent ions demonstrated that $MnCl_2$ resulted in the highest yield of condensation products (Figure 2.1).

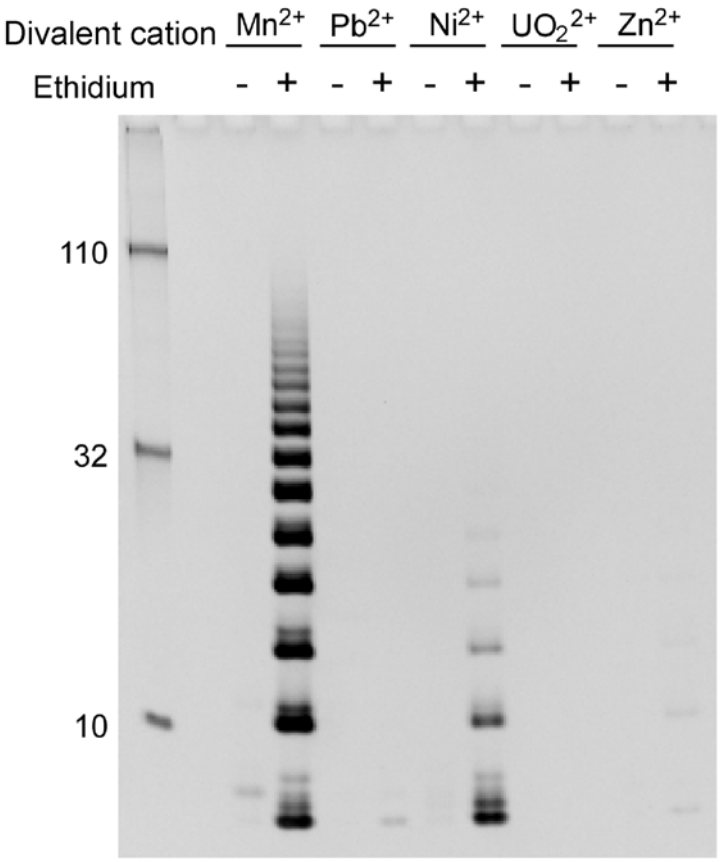


Figure 2.1. Divalent metal ion dependence of ethidium-mediated and *N*-cyanoimidazole (250 mM)-activated polymerization of d(pCGTA). Divalent metal ions were added at 5 mM. Tetranucleotides (200 μM) were condensed in the absence and presence of ethidium (600 μM) for 24 h at 4 ˚C.

For reactions containing ethidium, the intercalator was added as indicated in figure legends. After 15 min. equilibration at 4 °C, N-cyanoimidazole (from an H_2O stock) was added to 250 mM. Ligation reactions were incubated for 72 h at 4 ºC, unless otherwise stated. The reaction products were then ethanol precipitated and resuspended in PAGE loading buffer. In reactions containing <200 μM oligonucleotide, linear polyacrylamide was used as a carrier. Reaction products were separated using denaturing (7 M urea) PAGE (19% acrylamide:1% N,N′-methylenebisacrylamide) in TBE running buffer. The gels were then stained with SYBR Gold (Invitrogen) for band visualization.

2.2.3. A•A Paired Chemical Ligations

Ligation reactions were 500 μM in d(pA_6), 10 mM triethylammonium MES (pH 6), 5 mM $MnCl_2$. Reactions with intercalators were 750 μM in intercalator. Ligation reactions were incubated for 24 h at 4 ºC. Reactions products were separated using denaturing (7 M urea) PAGE (19% acrylamide:1% N,N′-methylenebisacrylamide) with TBE running buffer. Gels were then stained with SYBR Gold for band visualization.

2.2.4. HPLC Separation of Reaction Products

Reactions were 200 μM in tetranucleotide, 5 mM $MnCl_2$, 10 mM triethylammonium MES (pH 6) and 600 μM in ethidium (when present). N-cyanoimidazole was present at 25 mM, and the reactions were incubated at 4 °C. At each time point, a 10 μL aliquot was removed and diluted in 90 μL 22 mM EDTA to quench the reaction. The aliquot was then immediately chromatographed (Agilent 1100, 4.6 mm x 250 mm Phenomenex Luna C_{18}, ambient temperature). Gradient: Solvent A=100 mM

triethylammonium acetate, pH 7. Solvent B=Acetonitrile. 0-12 min, 7.5% B. 12-20 min, 7.5-20% B. 20-25 min, 20-70% B. 25-30 min, 70% B.

2.3. Intercalators Prevent Oligonucleotide Cyclization via Duplex Stabilization

We have previously demonstrated that the free energy associated with the intercalation of nucleic acid duplexes by small, planar molecules can promote the chemical ligation of oligonucleotides.[31, 33] However, our previous experimental systems were limited to the coupling of two monofunctional oligonucleotides, which were incapable of multiple couplings (i.e., polymerization), and, therefore, they were not subject to the problem of oligonucleotide cyclization. To determine if intercalation-mediated assembly can also circumvent the strand cyclization problem (Figure 2.2A), we investigated the chemical ligation of the tetranucleotide d(pCGTA) as a model system. The sequence of this oligonucleotide is such that, when base paired, it is capable of forming long, extensively nicked concatemeric (or tiling) Watson–Crick duplexes, with the 5′ and 3′ ends of the nick sites in close proximity for chemical ligation (Figure 2.2B).

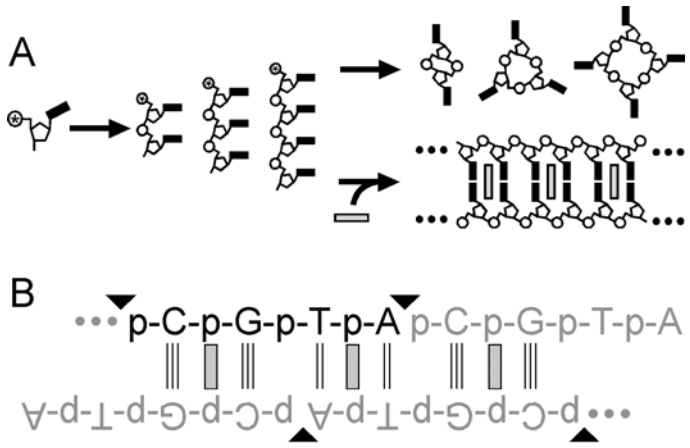


Figure 2.2. Illustration of the strand cyclization problem and intercalation-promoted assembly. (*A*) Schematic illustration of the strand cyclization problem (*top route*) and the circumvention of this problem by intercalation (*bottom route*). In the absence of intercalators, chemically activated oligonucleotides cyclize; in the presence of intercalators (shown as grey rectangles), they polymerize. (*B*) The nicked duplex resulting from intercalator-mediated assembly of d(pCGTA), used in the experimental studies presented in the text, is also illustrated. Triangles indicate nicks, which can be sealed by condensation after chemical activation of the terminal phosphate group.

To test the ability of intercalators to promote oligonucleotide polymerization and prevent cyclization, d(pCGTA) was activated for ligation with *N*-cyanoimidazole in the presence and absence of ethidium, a well-known intercalator of Watson–Crick duplexes. The early products of the ligation reaction (cyclic tetranucleotide, linear octanucleotide and cyclic octanucleotide) were monitored by HPLC (Figure 2.3A). Example chromatograms and kinetic traces derived from analysis of these reactions are shown in Figure 2.3. In the absence of ethidium, the major product is the cyclic tetranucleotide (Figure 2.3B), consistent with the very short persistence length of single-stranded oligonucleotides. About one-fourth as much cyclic octanucleotide and only trace linear octanucleotide were detected. The lack of appreciable linear octanucleotide formation indicates that the rate of cyclization is faster than the rate of formation for the linear octanucleotide. In contrast, in the presence of ethidium, linear octanucleotide formation is greatly favored over tetranucleotide cyclization (Figure 2.3C). The rate of octanucleotide formation in the presence of ethidium is slower than the rate of tetranucleotide cyclization in the absence of ethidium. Thus, ethidium-mediated assembly of d(pCGTA) inhibits formation of the cyclic tetranucleotide, in addition to promoting linear octanucleotide formation (Figure 2.2B).

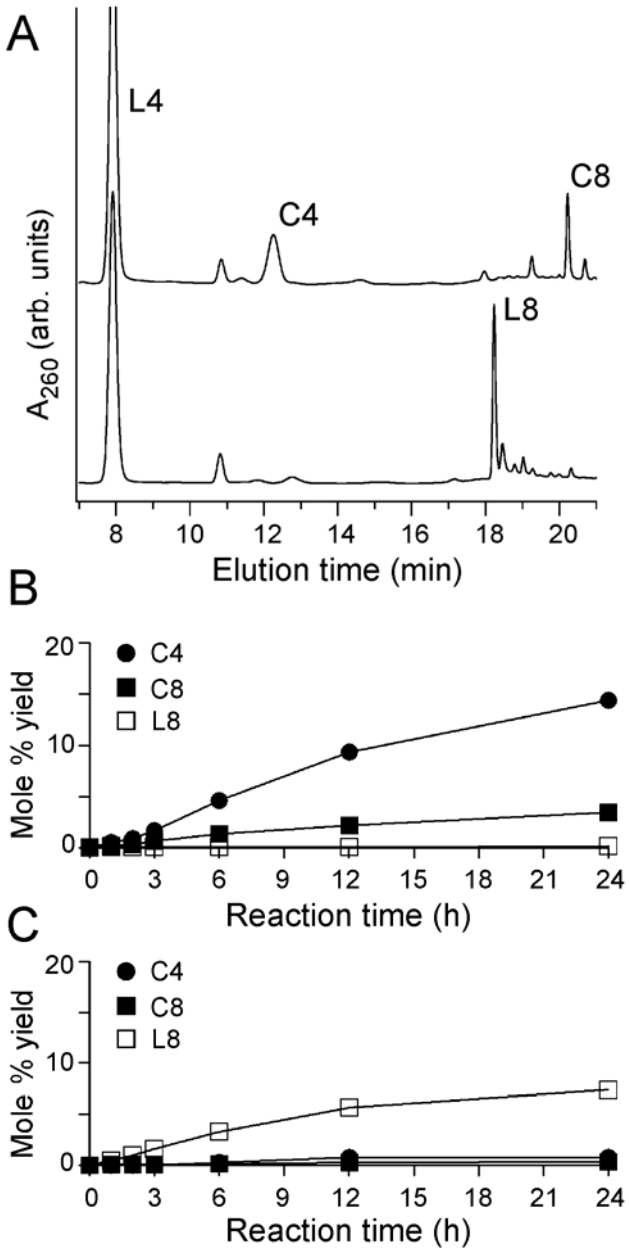


Figure 2.3. Chromatographic analyses of the effect of ethidium on tetranucleotide chemical condensation. (*A*) HPLC chromatogram of d(pCGTA) (200 μM) activated with *N*-cyanoimidazole (25 mM) and incubated at 4 °C for 24 h. The reaction analyzed in the upper chromatograph contained no ethidium; the reaction analyzed in the lower chromatograph contained 600 μM ethidium. Labels: L4, linear tetranucleotide (starting material); C4, cyclic tetranucleotide; L8, linear octanucleotide; and C8, cyclic octanucleotide. Products were identified using pure standards and phosphatase shift experiments (*B*) Kinetic analysis of the reaction described in *A* without ethidium. (*C*) Kinetic analysis of the reaction described in *B* in the presence of 600 μM ethidium.

Ethidium also suppresses strand cyclization in the absence of duplex formation. For example, the initial rate of cyclization of the tetranucleotide d(pCCTA), which cannot tile, is depressed twofold when ethidium is present (Figure 2.4). In contrast, the diminution of this rate in the case of d(pCGTA), which does tile, is nearly 2000-fold, indicating that assembly of oligonucleotides into Watson–Crick-paired duplexes is the dominant mechanism for suppression of cyclization. As we discuss below, this ability for intercalation to selectively protect base-pairing oligonucleotides from cyclization (and promote their polymerization) could have helped select the first informational polymers of life.

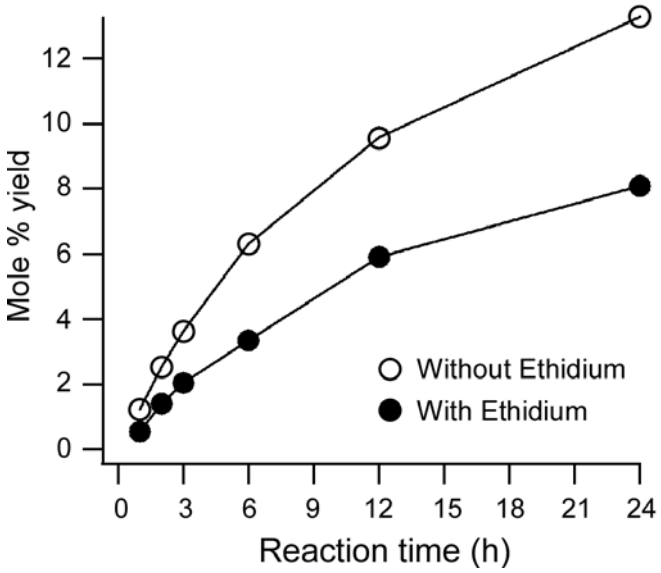


Figure 2.4. Cyclization kinetics of d(pCCTA) (200 μM) in the presence and absence of ethidium (600 μM). Both reactions contained 25 mM *N*-cyanoimidazole and were maintained at 4 °C. At each time interval, a 10 μL aliquot was removed and diluted in 90 μL 22 mM EDTA to quench the reaction. Aliquots were then immediately analyzed by HPLC.

2.4. Intercalators Promote Polymer Formation

The oligonucleotide d(pCGTA) does not ligate appreciably in the absence of an intercalator (Figure 2.5). In contrast, when ethidium is added to the same reactions, polymers of up to 100 nucleotides in length are observed (i.e., 24 linear couplings) (Figure 2.5). This analysis also illustrates that the ratio of linear to cyclic products increases in the presence of ethidium for all polymer lengths (Figure 2.5). For example, relatively low amounts of linear octa- and dodecanucleotide are detected compared to cyclic octa- and dodecanucleotide when ethidium is absent (lane 1, Figure 2.5), while approximately equal amounts of linear and cyclic octa- and dodecanucleotide products are observed when ethidium is present at a stoichiometry of one ethidium per tetranucleotide (lane 2, Figure 2.5).

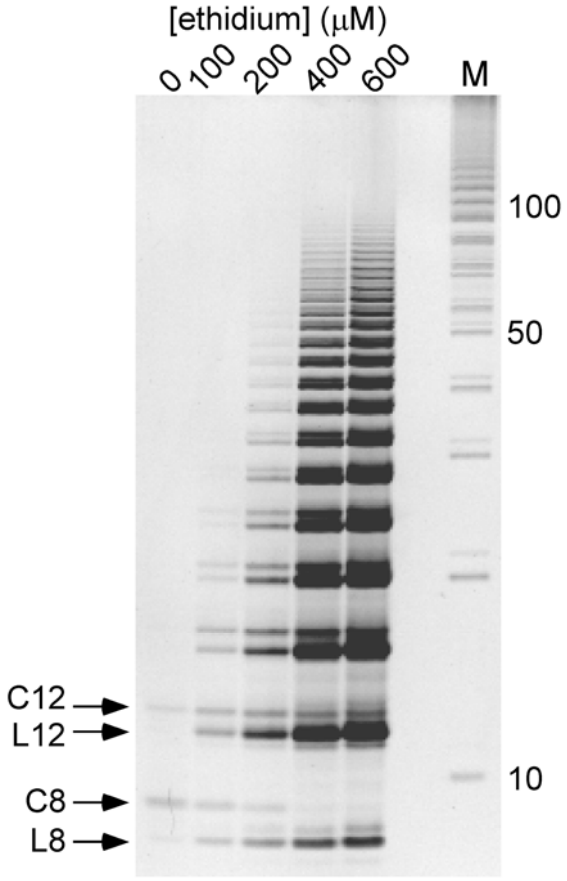


Figure 2.5. Polyacrylamide gel electrophoresis analysis of reaction products of activated d(pCGTA) in the presence of varying amounts of ethidium. Ethidium concentrations are given above lanes 1-5. All reactions contained 200 μM d(pCGTA), condensed with 250 mM *N*-cyanoimidazole at 4 ºC for 72 h. With no intercalator present (lane 1), only small amounts of cyclic octanucleotide (C8) and cyclic dodecanucleotide (C12) are produced, and almost no linear octanucleotide (L8) or linear dodecanucleotide (L12) is produced. Ethidium (lanes 2-5) dramatically promotes linear polymerization, to *ca.* 100mer products. The tetranucleotide starting material and cyclic tetranucleotide products observed in HPLC analyses did not stain efficiently; band intensities for ≥8mers are *ca.* linear with total nucleic acid. The gel lane marked M contains marker bands generated by the enzymatic ligation of a tiling decanucleotide.

For higher ethidium to tetranucleotide stoichiometries, the relative amounts of cyclic products are far lower than linear products of the same length (lanes 3-5, Figure 2.5). An exploration of the range of tetranucleotide concentrations over which ethidium promotes polymerization revealed polymerization at oligonucleotide concentrations as low as 5 μM, whereas, in the absence of ethidium, polymerization is not observed even at a tetranucleotide concentration of 60 mM (Figure 2.6). Thus, ethidium increases the concentration range over which d(pCGTA) can be polymerized by at least a factor of 10^4.

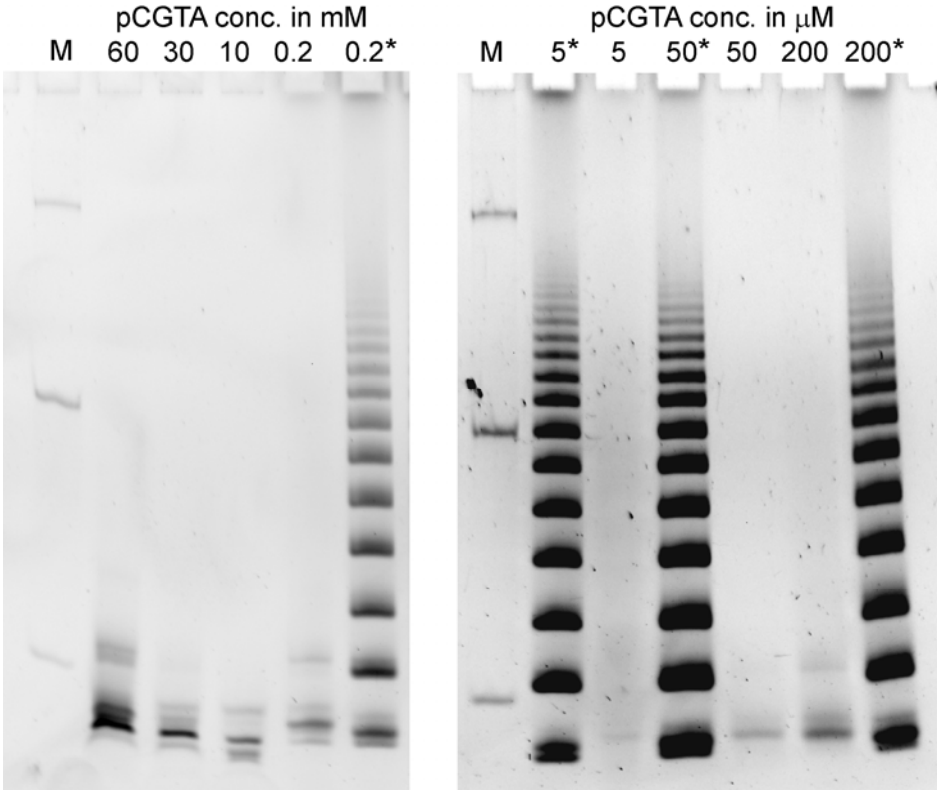


Figure 2.6. Oligonucleotide concentration and ethidium dependence of d(pCGTA) polymerization. Reactions contained 200 μM d(pCGTA), 5 mM $MnCl_2$, 10 mM triethylammonium MES (pH 6), 600 μM ethidium (noted with an asterisk when present), and 250 mM N-cyanoimidazole. All reactions were incubated for 72 h at 4 ˚C. Reactions ≤200 μM in tetranucleotide were ethanol precipitated with linear polyacrylamide carrier before loading to facilitate nucleic acid recovery. The molecular weight marker, labeled as lane M, contains DNA oligonucleotides of length 10, 32 and 110 nt.

2.5. Ethidium-Promoted Ligation Requires Watson–Crick Base Pairing

In order to demonstrate that the ethidium-promoted ligation of oligonucleotides requires Watson–Crick base pairing (and that nonspecific, e.g., hydrophobic or electrostatic, interactions are insufficient), we examined the ligation of two other oligonucleotides, d(pCCTA) and d(pGGTA). Individually, neither tetranucleotide can form a fully Watson–Crick base-paired tiled duplex, whereas a 1:1 mixture of the two strands can do so. Neither tetranucleotide alone polymerizes appreciably when activated, irrespective of the presence of ethidium (Figure 2.7). By contrast, polymerization occurs only when both activated tetranucleotides and ethidium are present (Figure 2.7), demonstrating the importance of Watson–Crick base pairing in ethidium-mediated polymerization.

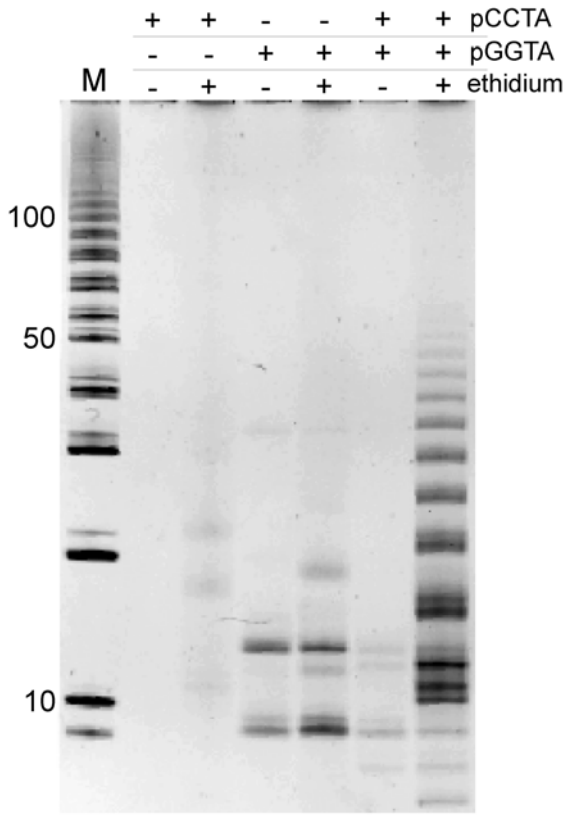


Figure 2.7. Condensation reactions demonstrating that ethidium-mediated oligonucleotide polymerization requires Watson-Crick base pairing. Reactions loaded in lanes 2 through 7 contained various compositions of d(pCCTA) and d(pGGTA) (all reactions 200 μM in tetranucleotide) and ethidium (600 μM, when present). Reactions were incubated at 4 ºC for 72 h with 250 mM *N*-cyanoimidazole.

As another means of examining the base-pairing specificity of ethidium-promoted ligation, we examined reactions containing d(TTTT), a tetranucleotide that can neither form a Watson–Crick base-paired tiled duplex nor polymerize on its own (due to its lack of a terminal phosphate), does not interfere with the intercalation-mediated polymerization of d(pCGTA), even when it comprises >99% of the total oligonucleotide present (Figure 2.8). Thus, ethidium selectively promotes the polymerization of Watson–Crick pairing oligonucleotides, even in the presence of a substantial excess of non-pairing oligonucleotides.

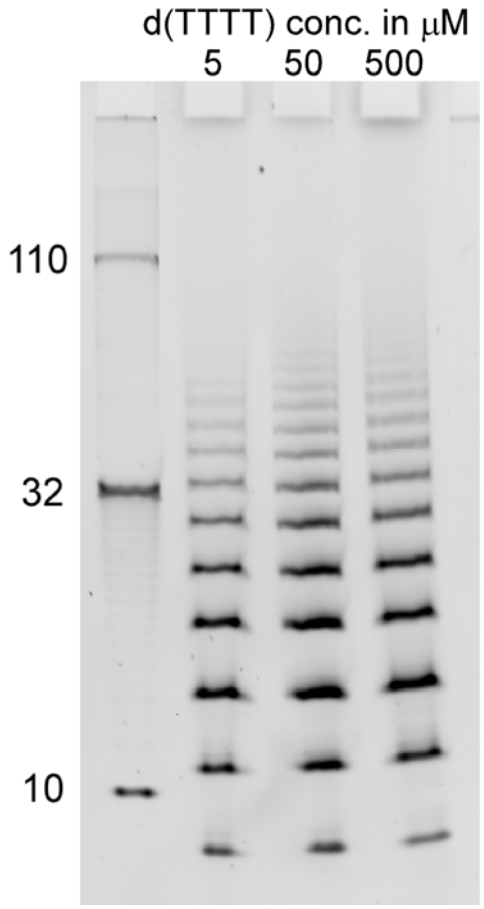


Figure 2.8. Tolerance of ethidium-mediated d(pCGTA) polymerization to the presence of non-pairing, non-polymerizable oligonucleotides. Reactions contained 5 μM d(pCGTA), 5 mM $MnCl_2$, 10 mM triethylammonium MES (pH 6), 600 μM ethidium (when present, noted with an asterisk above gel lanes), 250 mM N-cyanoimidazole, and the indicated concentration of d(TTTT). All reactions were incubated for 72 h at 4 ˚C. Reactions ≤200 μM in tetranucleotide were ethanol precipitated with linear polyacrylamide carrier before loading to facilitate nucleic acid recovery. The molecular weight marker contains DNA oligonucleotides of length 10, 32 and 110 nt. Reactions were ethanol precipitated with linear polyacrylamide carrier before loading.

2.6. Intercalators can Promote the Polymerization of Duplexes Containing Non-Watson–Crick Base Pairs

Intercalators can drive the assembly of a variety of base-pairing structures. Previously, we have demonstrated that a ligand for pyrimidine triplexes promotes the ligation of two Hoogsteen-paired strands in a triplex.[31] With this principle in mind, we sought to achieve ligand-dependent polymerization of nucleic acids by using a system comprised of ligand-dependent non-Watson–Crick base pairs. The recent discovery that aza3 (Figure 2.9) promotes the formation of duplex nucleic acids with A·A base pairs [34-39] suggested to us that homo-dA oligonucleotides might be suitable for a ligation system promoted by these molecules. Indeed, we found that aza3 promotes the assembly and ligation of d(pA_6). In contrast, ethidium, which promotes the ligation of oligonucleotides that form Watson–Crick base pairs, does not promote the ligation of d(pA_6) (Figure 2.8). Similarly, aza3, which has low affinity for Watson–Crick duplexes [40], does not promote the ligation of d(pCGTA) (Figure 2.9).

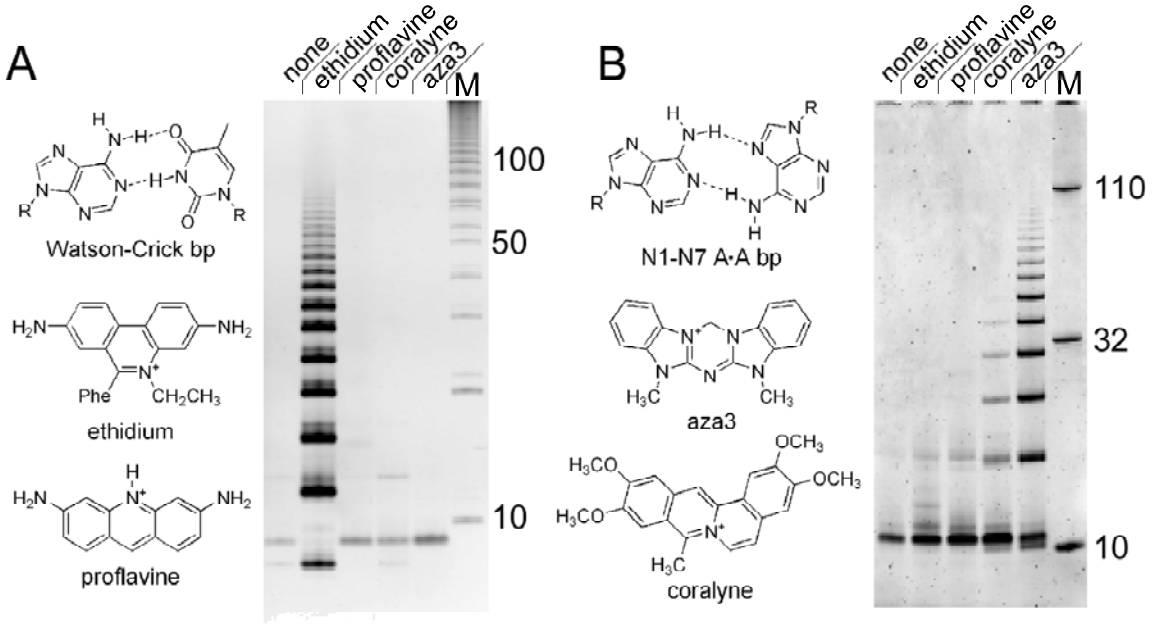


Figure 2.9. Ligand-base pair recognition is necessary, but not sufficient, for intercalation mediated ligation. *(A)* PAGE analysis of products from the condensation of d(pCGTA) (200 μM) in the absence and the presence of various small molecules (600 μM when present). Relative to the ligand-free reaction (lane 1), ethidium (lane 2) promotes the polymerization of d(pCGTA), whereas proflavine (lane 3), another Watson–Crick intercalator, does not. Neither coralyne (lane 4) nor aza3 (lane 5), which binds homo-A duplexes, promotes d(pCGTA) ligation. *(B)* PAGE analysis of products formed by the condensation of d(pA_6) (500 μM) in the absence and the presence of various small molecules (750 μM when present, lanes as in *A*). Relative to the ligand-free reaction (lane 1), neither Watson–Crick intercalator promotes the ligation of d(pA_6) (lanes 2 and 3). In contrast, aza3 (lane 5) promotes the ligation of d(pA_6). Although coralyne also binds homo-A duplexes, it promotes the ligation of d(pA_6) only weakly (lane 4). The N1-N7 base pair shown for homo-adenine duplexes has been shown to be consistent with molecular modeling and chemical probing studies. Reactions were incubated at 4 ºC for 24 h with 250 mM *N*-cyanoimidazole.

Although coralyne binds duplexes with A·A base pairs with an even greater affinity compared to aza3 [35], coralyne promotes d(pA_6) ligation only weakly. Similarly, proflavine, which binds Watson–Crick DNA with an affinity that is comparable to that of ethidium, does not promote ligation (Figure 2.9). Taken together, these data illustrate that molecular recognition between an intercalator and base pair may be necessary for the intercalation-mediated stabilization of a duplex structure, but it is clearly not sufficient for the promotion of intercalation-mediated polymerization.

2.7. Conclusion

It is widely assumed that the abiotic synthesis and template-directed replication of nucleic acid polymers (or their predecessors) was an early and critical step in the origin of life.[41-44] Strand cyclization is an inherent and general obstacle to the growth of polymers from bifunctional mono- and oligomers, and it would have therefore thwarted the abiotic synthesis of early phosphodiester-linked RNA or RNA-like polymers, unless a mechanism existed to organize mono- and short oligonucleotides prior to the bond formation step of polymerization. Previous chemical polymerization studies required moderate to high concentrations of di- and tetranucleotide substrates (10^{-3}-10^{-1} M) [45, 46], or the use of monofunctional substrates in order to prevent strand cyclization.[42] Here, we have shown that intercalator binding can circumvent strand cyclization and enable polymerization, even at low (10^{-6} M) substrate concentrations. Additionally, intercalators can drive the polymerization of base-pairing oligonucleotides, even when they are trace components (<1%) of a high-concentration pool of a non-base pairing, non-polymerizable oligonucleotide. While we do not propose that the intercalators used in the present study are prebiotic, it is possible that the reactions which gave rise to the first

nucleobases produced other heterocycles, which might have functioned as molecular midwives for the base pair selection, assembly, and polymerization of the earliest proto-RNA polymers.

Here, we have presented results from experiments involving the ligation of DNA oligonucleotides, which provide an experimentally practical model system for exploring intercalation-promoted polymerization. Nevertheless, these results are also relevant to the nonenzymatic polymerization of RNA, which has a duplex persistence length somewhat greater than that of DNA [29], but a single-stranded persistence length that still renders RNA oligonucleotides susceptible to cyclization.[12] Indeed, activated r(pCGUA) exhibits enhanced ligation in the presence of ethidium, albeit to a much lesser extent than observed for d(pCGTA) (Figure 2.10). The diminished performance of RNA in this system might result from the activation chemistry and intercalators tested thus far being less compatible with RNA oligonucleotides or the intercalated A-form helix. As noted above, not all intercalators tested were found to facilitate DNA ligation, even though they have binding affinities comparable to those that do support ligation. Current efforts in our laboratory include the search for an intercalator that more efficiently promotes RNA ligation.

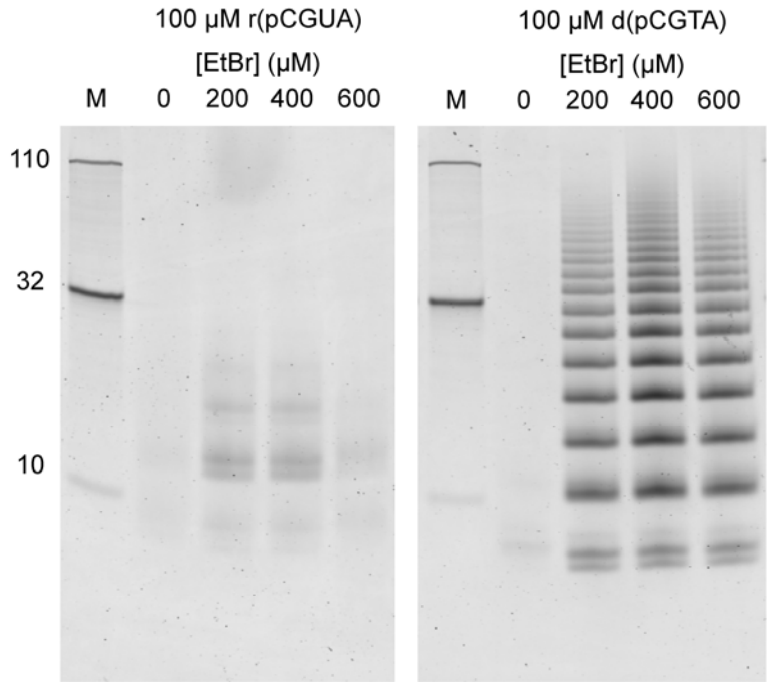


Figure 2.10. Comparison of ethidium-mediated polymerization of RNA and DNA tetranucleotides. Reactions contained 200 μM d(pCGTA) or r(pCGUA), 5 mM $MnCl_2$, 10 mM triethylammonium MES (pH 6), 600 μM ethidium (when present, noted with an asterisk above gel lanes), and 250 mM N-cyanoimidazole. All were incubated for 72 h at 4 ˚C. The molecular weight marker, labeled as lane M, contains DNA oligonucleotides of length 10, 32 and 110 nt.

Consistent with our previous work, the present ligation system exhibits sensitivity to the particular intercalator used to promote assembly, even among intercalators that exhibit similar affinity for the duplex to be ligated. In particular, proflavine did not promote ligation in our Watson-Crick system, despite the fact that proflavine and ethidium have similar association constants for Watson-Crick duplex DNA. Interestingly, in a previous study of a ligation system using a different activation chemistry, we found that proflavine, but not ethidium, promotes oligonucleotide ligation.[33] Thus, the efficiency of an intercalator for promoting a given ligation reaction appears not to be fully expressed by its association constant for a nucleic acid assembly. In retrospect, this

observation is not surprising; covalent bond formation requires an energetically accessible reactant complex with appropriate bond geometries, and individual intercalators make structure-specific helical contacts, which could modify the conformational landscape of the polymer backbone in an intercalator-specific manner. This feature could prove useful for selecting specific backbone linkages in intercalation-mediated reactions.

We have used tetranucleotides as a model system to illustrate the nonenzymatic production of polymers from short oligonucleotides. With regard to a more complete model for the origin of RNA-like polymers, it is certainly desirable to demonstrate polymerization with even shorter oligonucleotides and, ultimately, mononucleotides. Thus far, our attempts to use intercalators to drive the polymerization of mononucleotides and dinucleotides in the absence of a pre-existing template strand have proven unsuccessful. In the case of dinucleotides, we surmise that the nearest neighbor exclusion principle [47-49], which states that intercalators can bind at most between every other base pair, could potentially inhibit the coupling of dinucleotides, as dinucleotides are known to crystallize in the presence of intercalators with ligands bound both within and outside each minihelix [50], an arrangement that could block backbone coupling between dinucleotides. To find a possible means by which to circumvent this problem, we have initiated studies of intercalation of nucleic acids with alternative backbone linkages.[51, 52] Given the wide variety of RNA-like polymers with nucleoside elements other than ribofuranosides that also form duplexes [11, 53, 54], it is possible that the backbone of the first informational polymers may not have utilized ribofuranosides. If this was the case, the original pre-RNA backbone might have allowed intercalation between every

dinucleotide step. With regard to mononucleotide polymerization, it appears that the poor stacking of the pyrimidine nucleotides precludes assembly of mononucleotides by intercalating molecular midwives. The poor stacking of pyrimidine bases has prompted our laboratory, and several others [55-57], to reconsider an early proposal by Crick that the first RNA-like polymers might have used purine-purine base pairs.[58] Our demonstration in the present work that intercalators can promote the polymerization of homo-adenine oligonucleotides represents an early step towards the realization of intercalation-mediated mononucleotide polymerization in a homo-purine system.

Recently, Szostak and coworkers demonstrated the non-enzymatic template-directed synthesis of RNA inside model protocells.[59] These model protocells exhibited high permeability for small, minimally charged species. Therefore, before the advent of protein enzymes, molecules similar to the intercalator midwives discussed in the present work, which contain only a single positive charge, could have permeated protocells along with nucleotides, thus assisting in the assembly of encapsulated nucleic acid polymers.

Finally, it was recently proposed that the tendency for Watson–Crick DNA duplexes to form liquid crystals could reflect the process by which prebiotic nucleic acids were organized (and partitioned) prior to polymerization.[60] The experimental results presented here demonstrate that intercalators can promote the assembly of oligonucleotides of comparable length at concentrations $>10^4$-fold lower than those required for liquid crystal formation.[60]

2.8. References

1. Joyce, G.F., *RNA evolution and the origins of life.* Nature, 1989. **338**: p. 217-224.

2. Gesteland, R.F., T.R. Cech, and J.F. Atkins, *The RNA World*. 3rd ed. 2006, Cold Spring harbor, NY: Cold Spring Harbor Laboratory Press. 768.

3. Fedor, M.J. and J.R. Williamson, *The catalytic diversity of RNAs.* Nature Reviews Molecular Cell Biology, 2005. **6**(5): p. 399-412.

4. Ellington, A.D., et al., *Evolutionary origins and directed evolution of RNA.* International Journal of Biochemistry & Cell Biology, 2009. **41**(2): p. 254-265.

5. Cech, T.R., *Structural biology - The ribosome is a ribozyme.* Science, 2000. **289**(5481): p. 878-879.

6. Heuberger, B.D. and C. Switzer, *A pre-RNA candidate revisited: both enantiomers of flexible nucleoside triphosphates are DNA polymerase substrates.* Journal of the American Chemical Society, 2008. **130**(2): p. 412-413.

7. Mittapalli, G.K., et al., *Mapping the landscape of potentially primordial informational oligomers: oligodipeptides tagged with 2,4-disubstituted 5-aminopyrimidines as recognition elements.* Angewandte Chemie-International Edition in English, 2007. **46**(14): p. 2478-2484.

8. Bean, H.D., et al., *Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA.* Origins of Life and Evolution of the Biosphere, 2006. **36**(1): p. 39-63.

9. Benner, S.A., A. Ricardo, and M.A. Carrigan, *Is there a common chemical model for life in the universe?* Curr. Opin. Chem. Biol., 2004. **8**(6): p. 672-689.

10. Benner, S.A., *Understanding nucleic acids using synthetic chemistry.* Accounts of Chemical Research, 2004. **37**: p. 784-797.

11. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 1999. **284**(5423): p. 2118-2124.

12. Ertem, G. and J.P. Ferris, *Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds.* Journal of the American Chemical Society, 1997. **119**: p. 7197 - 7201.

13. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* Journal of Theoretical Biology, 2000. **205**(4): p. 543-562.

14. Leitzel, J.C. and D.G. Lynn, *Template-directed ligation: from DNA towards more versatile templates. The Chemical Record*, 2001. **1**: p. 53-62.

15. Costanzo, G., et al., *Nucleoside phosphorylation by phosphate minerals.* Journal of Biological Chemistry, 2007. **282**(23): p. 16729-16735.

16. Anastasi, C., et al., *RNA: Prebiotic product, or biotic invention?* Chemistry & Biodiversity, 2007. **4**(4): p. 721-739.

17. Rajamani, S., et al., *Lipid-assisted synthesis of RNA-like polymers from mononucleotides.* Origins of Life and Evolution of the Biosphere, 2008. **38**: p. 57–74.

18. Huang, W. and J.P. Ferris, *One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis.* J. Am. Chem. Soc., 2006. **128**: p. 8914-8919.

19. Sawai, H., et al., *Non-enzymatic, template-directed ligation of 2 '-5 ' oligoribonucleotides. Joining of a template and a ligator strand.* Nucleic Acids Research, 1998. **26**(12): p. 2995-3000.

20. Sawai, H., K. Kuroda, and T. Hojo, *Uranyl-ion as a highly effective catalyst for internucleotide bond formation.* Bulletin of the Chemical Society of Japan, 1989. **62**(6): p. 2018-2023.

21. Streltsov, S.A., et al., *Trivaline catalyzes 5'-pd(GTT) oligomerization in solution.* Febs Letters, 1992. **298**(1): p. 57-60.

22. Saladino, R., et al., *Formamide chemistry and the origin of informational polymers.* Chem. Biodiv., 2007. **4**: p. 694-720.

23. Joyce, G., et al., *Chiral selection in poly(C)-directed synthesis of oligo(G).* Nature, 1984. **310**: p. 602-604.

24. Joyce, G.F., T. Inoue, and L.E. Orgel, *Non-enzymatic template-directed synthesis on RNA random copolymers - poly(C,U) templates.* Journal of Molecular Biology, 1984. **176**(2): p. 279-306.

25. Weimann, B.J., et al., *Template-directed synthesis with adenosine-5'-phosphorimidazolide.* Science, 1968. **161**(3839): p. 387.

26. Kawamura, K. and F. Okamoto, *Cyclization and dimerization of hexanucleotides containing guanine and cytosine with water-soluble carbodiimide.* Viva Origino, 2001. **29**(4): p. 162-167.

27. Eichinger, B.E., *Cyclization in reversible and irreversible step-growth polymerizations.* Computational and Theoretical Polymer Science, 2000. **10**(1-2): p. 83-88.

28. Mills, J.B., E. Vacano, and P.J. Hagerman, *Flexibility of single-stranded DNA: Use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA).* Journal of Molecular Biology, 1999. **285**(1): p. 245-257.

29. Kebbekus, P., D.E. Draper, and P. Hagerman, *Persistence length of RNA.* Biochemistry, 1995. **34**(13): p. 4354-4357.

30. Porschke, D., *Elementary steps of base recognition and helix-coil transitions in nucleic acids.* Mol Biol Biochem Biophys, 1977. **24**: p. 191-218.

31. Hud, N.V., et al., *Addressing the problems of base pairing and strand cyclization in template-directed synthesis - A case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA.* Chemistry & Biodiversity, 2007. **4**(4): p. 768-783.

32. Huang, K.S., et al., *Synthesis and reactions of some heterocyclic azacyanines.* Journal of Organic Chemistry, 2001. **66**(4): p. 1310-1315.

33. Jain, S.S., et al., *Enzymatic behavior by intercalating molecules in a template-directed ligation reaction.* Angewandte Chemie, International Edition in English, 2004. **43**: p. 2004-2008.

34. Joung, I.S., et al., *Molecular dynamics simulations and coupled nucleotide substitution experiments indicate the nature of A center dot A base pairing and a putative structure of the coralyne-induced homo-adenine duplex.* Nucleic Acids Research, 2009. **37**(22): p. 7715-7727.

35. Çetinkol, Ö.P. and N.V. Hud, *Molecular recognition of poly(A) by small ligands: an alternative method of analysis reveals nanomolar, cooperative and shape-selective binding.* Nucleic Acids Research, 2009. **37**: p. 611-621.

36. Persil, Ö., et al., *Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding.* Journal of the American Chemical Society, 2004. **126**(28): p. 8644-8645.

37. Xing, F., et al., *Molecular recognition of nucleic acids: coralyne binds strongly to poly(A).* Febs Letters, 2005. **579**(22): p. 5035-5039.

38. Jain, S.S., M. Polak, and N.V. Hud, *Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne-driven duplex disproportionation.* Nucleic Acids Research, 2003. **31**(15): p. 4608-4615.

39. Polak, M. and N.V. Hud, *Complete disproportionation of duplex poly(dT).poly(dA) into triplex poly(dT).poly(dA).poly(dT) and poly(dA) by coralyne.* Nuc. Acid Res., 2002. **30**: p. 983-992.

40. Çetinkol, Ö.P., et al., *Submicromolar, selective G-quadruplex ligands from one pot: thermodynamic and structural studies of human telomeric DNA binding by azacyanines.* ChemBioChem, 2008. **9**(12): p. 1889-1892.

41. James, K.D. and A.D. Ellington, *Surprising fidelity of template-directed chemical ligation of oligonucleotides.* Chemistry & Biology, 1997. **4**(8): p. 595-605.

42. Sievers, D. and G. von Kiedrowski, *Self replication of complementary nucleotide-based oligomers.* Nature, 1994. **369**(6477): p. 221-224.

43. von Kiedrowski, G., *A self-replicating hexadeoxynucleotide.* Angewandte Chemie-International Edition in English, 1986. **25**(10): p. 932-935.

44. Orgel, L.E. and R. Lohrmann, *Prebiotic chemistry and nucleic acid replication.* Accounts of Chemical Research, 1974. **7**(11): p. 368-377.

45. Zielinski, W.S. and L.E. Orgel, *Oligoaminonucleoside phosphoramidates. Oligomerization of dimers of 3'-amino-3'-deoxy-nucleotides (GC and CG) in aqueous-solution.* Nucleic Acids Research, 1987. **15**(4): p. 1699-1715.

46. Bolli, M., R. Micura, and A. Eschenmoser, *Pyranosyl-RNA: chiroselective self-assembly of base sequences by ligative oligomerization of tetra nucleotide-2',3'-cyclophosphates (with a commentary concerning the origin of biomolecular homochirality).* Chemistry & Biology, 1997. **4**(4): p. 309-320.

47. Arnott, S., P.J. Bond, and R. Chandrasekaran, *Visualization of an unwound DNA duplex.* Nature, 1980. **287**: p. 561-563.

48. Bond, P.J., et al., *X-Ray fiber diffraction evidence for the neighbor exclusion binding of a platinum metallointercalation reagent to DNA.* Proc. Nat. Acad. Sci. USA, 1975. **72**: p. 4825-4829.

49. Crothers, D.M., *Calculation of binding isotherms for heterogeneous polymers.* Biopolymers, 1968. **6**(4): p. 575-584.

50. Shieh, H.-S., et al., *The Structure of a Drug-Deoxydinucleoside Phosphate Complex; Generalized Conformational Behavior of Intercalation Complexes with RNA and DNA Fragments. Nucleic Acids Res.*, 1980. **8**(1): p. 85-97.

51. Horowitz, E.D. and N.V. Hud, *Ethidium and proflavine binding to a 2',5'-linked RNA duplex.* Journal of the American Chemical Society, 2006. **128**(48): p. 15380-15381.

52. Horowitz, E.D., et al., *Solution structure and thermodynamics of 2',5' RNA intercalation.* Journal of the American Chemical Society, 2009. **in press**.

53. Schöning, K.U., et al., *Chemical etiology of nucleic acid structure: The alpha -threofuranosyl-(3'->2') oligonucleotide system.* Science, 2000. **290**: p. 1347-1351.

54. Eschenmoser, A., *Searching for nucleic acid alternatives.* Chimia, 2005. **59**(11): p. 836-850.

55. Groebke, K., et al., *Why pentose- and not hexose-nucleic acids? Purine-purine pairing in homo-DNA: guanine, isoguanine, 2,6-diaminopurine, and xanthine.* Helv. Chim. Acta, 1998. **81**: p. 375-474.

56. Battersby, T.R., M. Albalos, and M.J. Friesenhahn, *An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids.* Chemistry & Biology, 2007. **14**(5): p. 525-531.

57. Heuberger, B.D. and C. Switzer, *An alternative nucleobase code: characterization of purine-purine DNA double helices bearing guanine-isoguanine and diaminopurine 7-deaza-xanthine base pairs.* ChemBioChem, 2008. **9**: p. 2779-2783.

58. Crick, F.H.C., *The origin of the genetic code.* J. Mol. Biol., 1968. **38**: p. 367-379.

59. Mansy, S.S., et al., *Template-directed synthesis of a genetic polymer in a model protocell.* Nature, 2008. **454**(7200): p. 122-125.

60. Nakata, M., et al., *End-to-end stacking and liquid crystal condensation of 6-to 20-base pair DNA duplexes.* Science, 2007. **318**(5854): p. 1276-1279.

CHAPTER 3

OLIGONUCLEOTIDE POLYMERIZATION VIA LIGAND-PROMOTED REVERSIBLE COVALENT BOND FORMATION

3.1. Introduction

In the past three decades, nucleic acids have demonstrated a capacity for numerous roles beyond their principal information-carrying function. These polymers can perform catalysis and bind diverse, even arbitrarily selected, ligands.[1-3] Thanks to this suite of capabilities, nucleic acids have found utility in noncoding applications, in both *in vitro* and previously undiscovered *in vivo* roles, in applications as varied as sensors, drugs, gene regulation, and imaging technologies.[4-7] The wide-ranging utility of nucleic acids has caused many to consider a dual (catalytic and informational) role for these polymers in prebiotic chemistry.[8, 9] A high-yielding and general synthetic route for the nonenzymatic polymerization of nucleic acids based on the aqueous self-assembly of their monomer components is an enticing prospect for both biotechnology and prebiotic chemistry, yet such a route remains elusive.

Considerable inroads towards this aim have been made towards this goal.[10-16] The work of Ferris, in which mono- and oligonucleotides are allowed to react in the presence of templates and catalysts such as clays, is particularly notable. These reactions have been shown to produce oligonucleotides as long as 50 residues in length. However, they also produce considerable amounts of cyclic products.[17, 18] Due to the kinetic stability of the phosphodiester linkage, these side reactions irreversibly consume monomer, impairing the production of linear polymer. One strategy that has been employed to prevent cyclization in polymerization reactions is the use of chain

terminators. Such terminators impair cyclization and result in the production of longer polymers, at the expense of allowing elongation in only one direction.[19]

One route around the problem of strand cyclization is the use of a reversible step in polymerization.[20, 21] By using a polymerization scheme involving a reversible step, followed by trapping of the product in an irreversible linkage, we can effectively recapitulate the "proofreading" capability observed in some contemporary polymerases. Towards this end, the Lynn laboratory has demonstrated that the use of a polymerization reaction that incorporates a reversible step can promote the production of linear polymer: 5′-amino-5′-deoxy, 3′-formylmethyl-3′-deoxythymidine polymerizes via imine formation and reductive amination, but it does so only if the appropriate d(A_8) template is present.[22] Liu and colleagues have extended this reversible ligation system to PNA oligomers.[23, 24]

One route to an aldehydic functionality in a nucleic acid polymer is the reaction of a 3′-ribonucleotide residue with periodate, which produces a 2′,3′ dialdehyde (Scheme 1b). This dialdehyde can reversibly react with an alkylamine, and, upon reduction, yield a hydrolytically stable morpholine linkage.[25] This scheme has been explored previously in nucleic acid chemistry. Micklefield recently demonstrated the DNA template-directed extension of a PNA oligomer with a single RNA nucleoside via reductive amination.[26] Similarly, Torrence has demonstrated the untemplated extension of PNA and amino-terminated phosphorodiamidate morpholino oligomers using periodate-oxidized RNA.[27, 28] Wincott and Komatsu have each used this chemistry to ligate bimolecular duplexes, thereby forming hairpins.[29, 30]

While it is clear such linkages are capable of spanning a loop or extending a modified oligonucleotide, it is unclear how well these linkages are tolerated within a double helix. Here, we demonstrate that a linkage of this type is well-accommodated among phosphodiester linkages within a double helix. We then use this ligation chemistry to demonstrate that an oligonucleotide with 5′ amine and 3′ dialdehyde functionalities can be quantitatively ligated into polymers without the need for enzyme-controlled catalysis, and that the reversibility of this linkage affords a means by which to circumvent the deleterious effects of cyclization and chain terminators in polymerization.

3.2. Experimental Procedures

3.2.1. Materials

The sequences prepared were as follows: Ald5, d(GAG TrC); Am5, d(H_2N-TAA GC); Temp10, d(GCT TAG ACT C); Bifunc12, d(H_2N-TTC CTG ACA TCrG); Bifunc12-5pDefect, d(TTC CTG ACA TCrG); Bifunc12-3pDefect, d(H_2N-TTC CTG ACA TCG); Bifunc12-3p5pDefect, d(TTC CTG ACA TCrG); Temp12, d(CAG GAA CGA TGT). Oligonucleotides containing 5′-amino-dT substitutions were prepared in-house using an Expedite 8909 (ABI), 5′-amino-dT phosphoramidite (Glen Research), standard phosphoramidites, reagents, and supports (ChemGenes). These oligonucleotides were prepared MMT-on and deprotected by treatment with 10% diethylamine in acetonitrile for 10 minutes to remove cyanoethyl protecting groups, followed by treatment with aqueous ammonia at room temperature for 48 H. They were then purified by RP-HPLC, desalted on a C18 Sep-Pak Plus (Waters), lyophilized, and subsequently treated with 20% aqueous acetic acid at room temperature for one hour to remove the

MMT group. This solution was neutralized with triethylamine, desalted on a C18 Sep-Pak Plus, lyophilized, and resuspended in water. Oligonucleotides not containing this substitution were purchased from IDT, purified by IE-HPLC, desalted on a C18 Sep-Pak Plus, lyophilized, and resuspended in water. The identity of oligonucleotides was confirmed by MALDI using an ABI 4700 (Negative ion, THAP/Ammonium Citrate matrix). Oligonucleotides were stored at -20 °C in 1mM-5mM stock solutions.

Sodium borate buffer, pH 8.5 was prepared by sodium hydroxide titration of boric acid, sterile filtered, and stored at 4 °C. Sodium periodate and sodium cyanoborohydride solutions were prepared weekly and stored at 0 °C in the dark. Ethidium solutions were stored in the dark at room temperature.

3.2.2. Ligation Reactions

Ligation reactions were prepared in a PCR tube by the sequential addition of 5× or 10× solutions of the following components in the following order: buffer (100 mM final concentration), oligonucleotides (10 or 100 μM final concentration), ethidium (60 or 600 μM final concentration), sodium periodate (2.5 mM final concentration), and water (q.s.). After equilibration for one day, dry sodium cyanoborohydride (prepared by evaporation of an aqueous stock, 25 mM final concentration) was added to the reaction. Temperature control was provided by either an ice bath or a Bio-Rad MyCycler.

3.2.3. Thermal Denaturation

Ligation reactions, 1 mM each in Temp10, Ald5, and Am5 (5 mM ethidium bromide, when present) were held at 0 °C overnight and diluted to 2 μM in each strand in 100 mM sodium borate, pH 8.5, 25 mM NaCl. These samples were subjected to two

cycles of heating and cooling between 5 °C and 50 °C, with spectra taken at 1 °C intervals. Each spectrum was fit to a weighted average of the 5 °C and 50 °C spectra of the second heating trace with a constant offset. The T_M was determined as the midpoint of a sigmoidal fit of the resulting transition.

3.3. 2′,3′-Dialdehyde and 5′-Amino-dT Functionalities Comprise an Efficient Template- and Ligand-Promoted Ligation System

In our ligation system, we sought to employ the reductive amination of ribose-derived dialdehydes to connect two strands annealed to a third template strand. To this end, we incorporated the 5′-amino,5′-deoxy-dT substitution Lynn and colleagues used previously.[22, 31] Our ligation system consists of a template strand Temp10, dialdehyde precursor Ald5, and a 5′-amino,5′-deoxy-dT-substituted strand Am5 (Figure 3.1b). In the absence of a template, only a small amount of background ligation occurs between Ald5 and Am5; overnight reduction of a reaction containing 10 μM of each strand resulted in only 9.1% conversion of Ald5 and Am5 to the morpholine-linked decamer product (Table 3.1). The ligation yield is substantially enhanced in the presence of Temp10; in the presence of this template, 68% of Ald5 and Am5 were ligated. In previous work with two other ligation systems, we have had success in using ligand binding to drive the assembly of short oligonucleotides, thus enhancing ligation yield.[32, 33] We found this strategy to be effective in the present system as well; the addition of nearest-neighbor ethidium bromide to the reaction mixture resulted in essentially quantitative (94%) ligation of Ald5 and Am5.

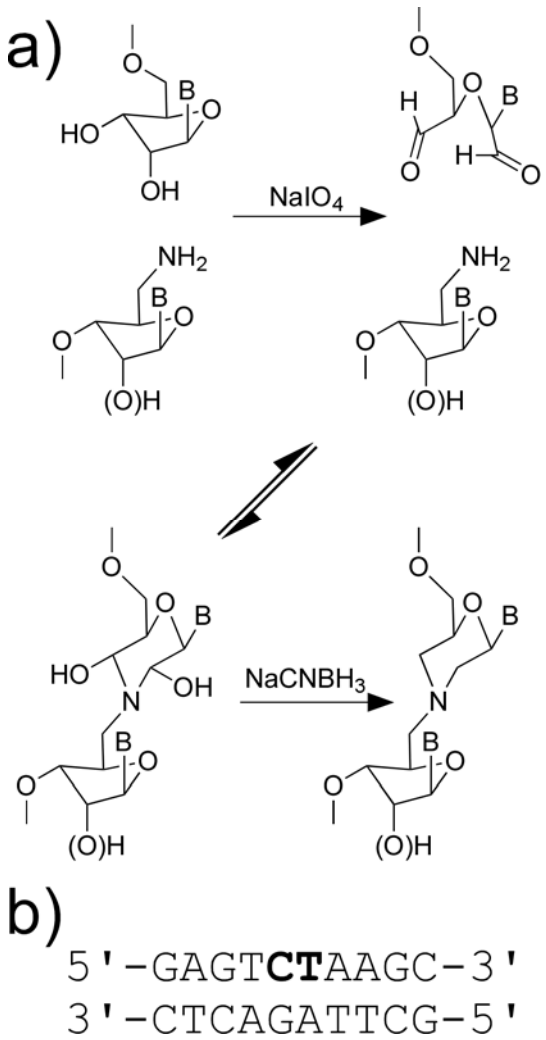


Figure 3.1. Reversible ligation system. a) An oligonucleotide with a 3′-ribonucleotide terminus, when treated with periodate, yields a 2′,3′-dialdehyde. This species can reversibly couple with a strand containing a 5′-amino,5′-deoxy residue. Upon reduction, a hydrolytically stable morpholine linkage is formed. b) Template-directed ligation system. Ald5 (d(GAGTrC)) and Am5 (d(H_2N-TAAGC)) anneal to Temp10 (d(GCTTAGACTC)), forming a morpholine linkage between the bold residues.

Table 3.1. Template and ligand dependence of yields of morpholine-linked decamer. Yields are following overnight reduction at 0 °C.

[Ald5]=[Am5]	[Temp10]	[EtBr]	Yield
10 μM	0 μM	0 μM	9.1%
	10 μM	0 μM	68%
	10 μM	50 μM	94%
100 μM	0 μM	0 μM	22%
	100 μM	0 μM	90%
	100 μM	500 μM	97%
1 mM	0 mM	0 mM	64%
	1 mM	0 mM	95%
	1 mM	5 mM	98%

We unexpectedly observed a small, but measurable, amount of background ligation, even in the absence of a template. In reductive aminations, care is often taken to exclude water by performing reactions in organic solvents and employing drying agents.[34] In previous studies of monoaldehydes, Liu and we both observed a rigorous template dependence.[22, 24] At higher concentrations, still higher template-independent ligation yields result, with a remarkable 64% overnight ligation yield of Ald5 and Am5 observed at 1 mM strand concentration. At these concentrations, the same trend is observed – the inclusion of a template, as well as the inclusion of ethidium, enhanced ligation yields.

In retrospect, it is perhaps not entirely surprising that untemplated ligation occurs in this system, particularly at high concentrations. Dialdehydes generated by the periodate treatment of the ribose ring have previously been used to functionalize the 3′ end of RNAs – another non-template directed reaction with an amine. What appears to occur in this system is a chelate effect: addition of the nitrogen atom to the second aldehyde equivalent in the dialdehyde occurs at drastically diminished entropic cost, owing to six-membered ring formation – an intramolecular reaction.[35] A similar

situation exists for sugars, which exist in furanose (five-membered cyclic hemiacetal) and pyranose (six-membered cyclic hemiacetal) forms, despite the low equilibrium concentration of hemiacetal and acetals in dilute solutions of alcohols and aldehydes in 55 M water.[36, 37] In the present system, while background ligation occurs, it is concentration-dependent, and it can be ameliorated by performing ligations at low concentrations.

3.4. The 4-atom Morpholine Linker is Well-Accommodated in the 6-Atom Phosphodiester-Linked Helix

It was not initially self-evident that our linkage would be tolerated within in the DNA helix. In our previous work from the Lynn laboratory with the ligation system comprised of 5′-amino-5′-deoxy-dT and 3′-formylmethyl-3′-deoxy-dT substitutions, it was found that the ethylamine linkage destabilized the DNA helix.[38] Based on this work, we reasoned that a long alkylamino functionality, which was well-tolerated in ligation across a deleted tetraloop in the work of Wincott and colleagues, might not be as well-tolerated within a double helix, as was required in our ligation system. Thus, we reasoned that simply connecting the dialdehyde to 5′-amino-5′-deoxy-dT might be the best ligation strategy.

Subtle changes to the nucleic acid backbone can have dramatic effects – for example, RNA containing six-membered pyranosyl rings forms more stable duplexes than that containing five-membered furanosyl rings, while certain acyclic analogues of RNA can exhibit elevated or diminished duplex stability.[39-42] It is clear that the morpholine ring itself does not preclude stable helix formation, since the well-known phosphoramidate morpholino oligonucleotides (PMOs) form very stable duplexes.[43]

However, PMOs retain the six-bond periodicity found in DNA and RNA. The morpholine linkage we sought to use contains only four bonds and is, thus, substantially shorter.

Table 3.2. Melting temperatures of duplexes comprised of Temp10 and d(GAGT**CT**AAGC), with a varied C5-T6 linkage. T_Ms for unreduced reversible linkages are given as lower bounds, due to concomitant strand dissociation and linkage hydrolysis.

Linkage	**T_M (no ethidium)**	**T_M (NN ethidium)**
Phosphate	26 °C	34 °C
Reduced	24 °C	33 °C
Unreduced	≥20 °C	≥28 °C

We characterized the compatibility of the linkage with a DNA double helix experimentally by determining the thermal stability of DNA duplexes containing the linkage. We compared the T_Ms of the decameric DNA duplexes shown in Figure 3.1, varying the linkage across the two bold bases between the standard phosphodiester linkage, the reduced morpholine ring, and the reversible reaction intermediate. The duplex containing a single morpholine linkage exhibited thermal stability close to that of the all-phosphodiester-linked duplex, with only a 2 °C depression in T_M associated with a single morpholine linkage (Table 3.2). Additionally, the incorporation of the morpholine linkage did not diminish the thermal stabilization associated with ethidium binding; nearest neighbor ethidium stabilized the duplex containing all phosphodiester linkages by +8 °C, and it stabilized the duplex containing a single morpholine linkage by +9 °C.

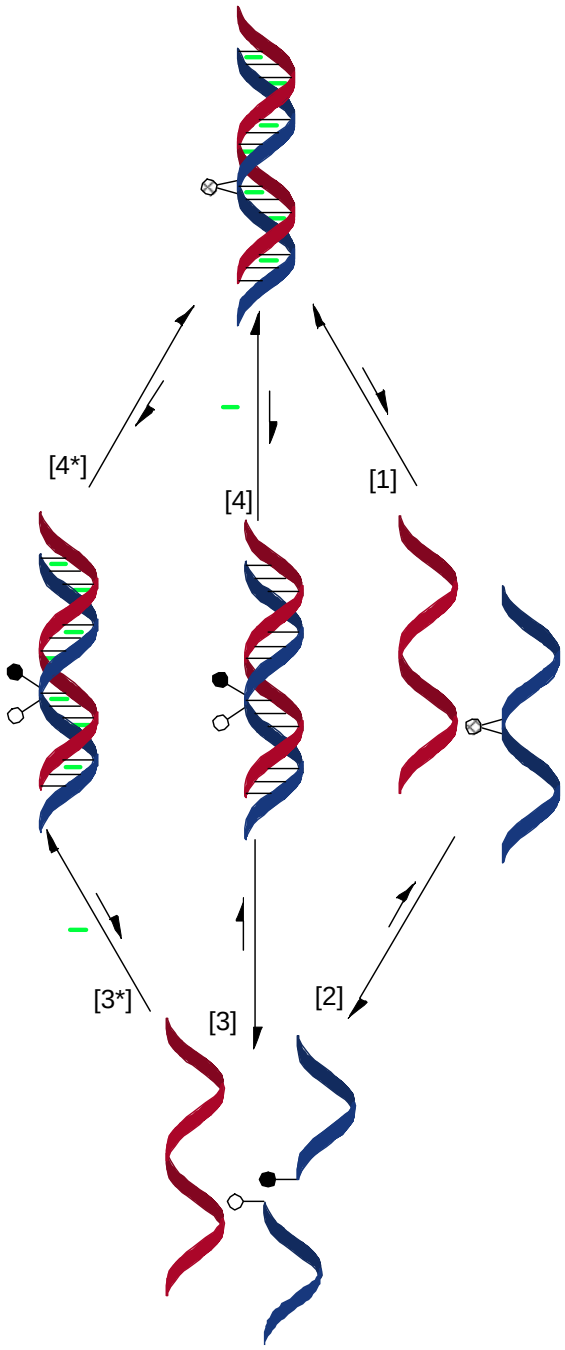


Figure 3.2. Schematic diagram of strand melting/annealing and hydrolysis/reformation of reversible linkage. When a duplex containing the reversible linkage (hatched circle) is heated, it dissociates into two strands (1). The reversibly linked strand (long blue strand, hatched circle) hydrolyzes to Am5 (short blue strand, open circle) and Ald5 (short blue strand, closed circle) (2). The short strands have low affinity for Temp10 (red) and do not reanneal (3); the reversible linkage is not reformed (4). However, in the presence of ethidium, both Am5-Temp10 and Ald5-Temp10 interactions are enhanced, and duplex formation occurs (3*), allowing reformation of the reversible linkage (4*).

The pre-reduction system illustrated in Scheme 1 exhibits complex thermal denaturation behavior (Figure 3.2). At high concentration (1 mM), Ald5 and Am5 anneal to Temp10, yielding the pre-reduction ligation product. When diluted to the concentration used in thermal denaturation experiments (2 μM), the linkage remains stable when the duplex remains annealed (i.e., below the melting temperature). However, when heated, the duplex dissociates, and the reversible linkage between Ald5 and Am5 hydrolyzes (Steps 1 and 2, Figure 3.2). Ald5 and Am5 are below the dissociation constant for untemplated linkage formation, as well as the dissociation constant for duplex formation with Temp10; therefore, the linkages does not reform (either through the agency of the pentanucleotides, as in Steps 3 and 4, Figure 3.2, or via untemplated reformation of the linkage and subsequent annealing, as in Steps 1 and 2, Figure 3.2). However, the free energy of ethidium binding drives duplex formation; in the presence of ethidium, Ald5 and Am5 anneal to Temp10 at low temperatures (Step 3*, Figure 3.2). When Ald5 and Am5 are annealed to Temp10, they are preorganized at a high local concentration, and the linkage forms again (Step 4*, Figure 3.2). At this concentration, (re)formation of the linkage is both template- and ligand-dependent.

3.5. Oligonucleotide Polymerization

Having ascertained the compatibility of the morpholine linkage within a double helix, we sought to employ this ligation system in a DNA polymerization system (Figure 3.3). An oligonucleotide containing both 5′-amine and 3′-dialdehyde functional groups, such as Bifunc12, should exhibit Temp12-dependent polymer formation. Indeed, at room temperature, a modest amount of template-dependent polymerization occurs (Figure 3.4). Absent cyanoborohydride, polymerization products are hydrolytically unstable, and only

a trace of Bifunc12 dimer is observed on the gel. Absent periodate, Bifunc12 is not converted to a polymerizable species. Absent Temp12, Bifunc12 is quantitatively converted to a cyclic product. When nearest-neighbor (600 μM) ethidium is introduced, the assembly of the concatemeric Temp12-Bifunc12 is favored, resulting in quantitative conversion of Bifunc12 into long polymers.

Figure 3.3. Bifunc12-Temp12 polymerization system. Bifunc12, in the presence of Temp12 and treated under the ligation conditions described in Subsection 3.2.1, polymerizes. The resulting Bifunc12-derived polynucleotide contains a morpholine linkage every twelve bases and is continuously base-paired with Temp12 oligonucleotides, as shown.

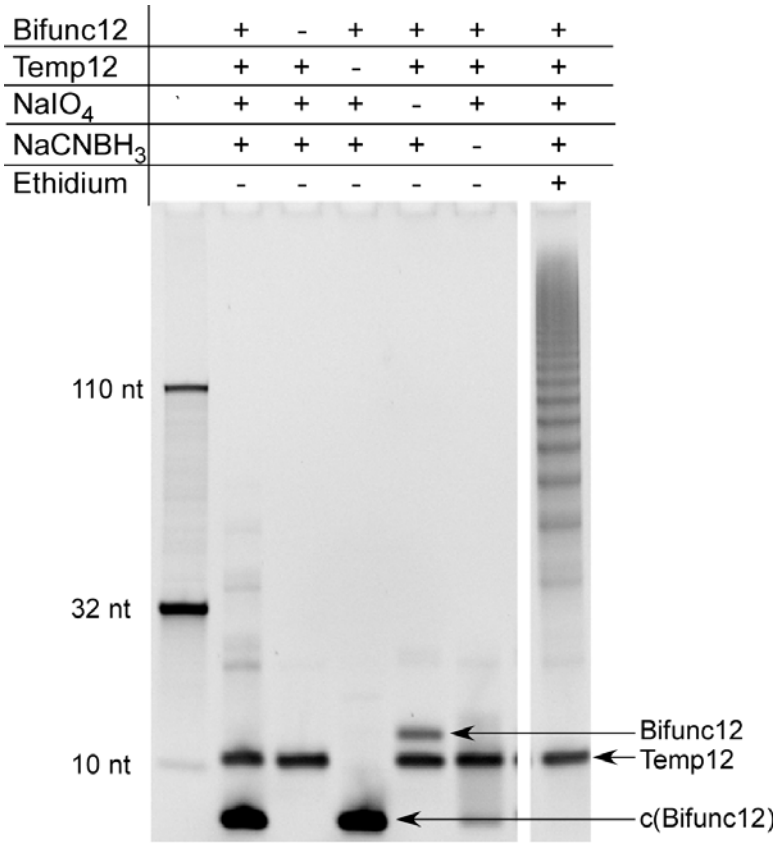


Figure 3.4. Polymerization reactions of Bifunc12. Each reaction contained Bifunc12, Temp12 (100 μM each strand), except for control reactions, in which the specified reagents or oligonucleotides were omitted. Reactions were equilibrated at room temperature without cyanoborohydride for 24H, then cyanoborohydride was added. Electrophoresis was performed after 4 days reduction at room temperature. c(Bifunc12) denotes cyclized Bifunc12.

3.6. Reversibility, Defective Substrates, and Cyclization

A Bifunc12 analogue with one end that cannot couple (i.e., a chain terminator), once incorporated into a nascent polymer, will allow polymerization in only one direction. Furthermore, such a strand will anneal to Temp12, amounting to a steric polymerization inhibitor. Indeed, Bifunc12-5pDefect, which lacks a 5′-amine, and Bifunc12-3pDefect, which lacks a 3′-vicinal diol, both depress the length of polymer produced in polymerization reactions (Figure 3.5). However, Bifunc12-3p5pDefect, which lacks both these functional groups, does not significantly affect the length of polymer produced, indicating that the steric effect of a strand with the correct sequence but no polymerization-competent ends is minimal.

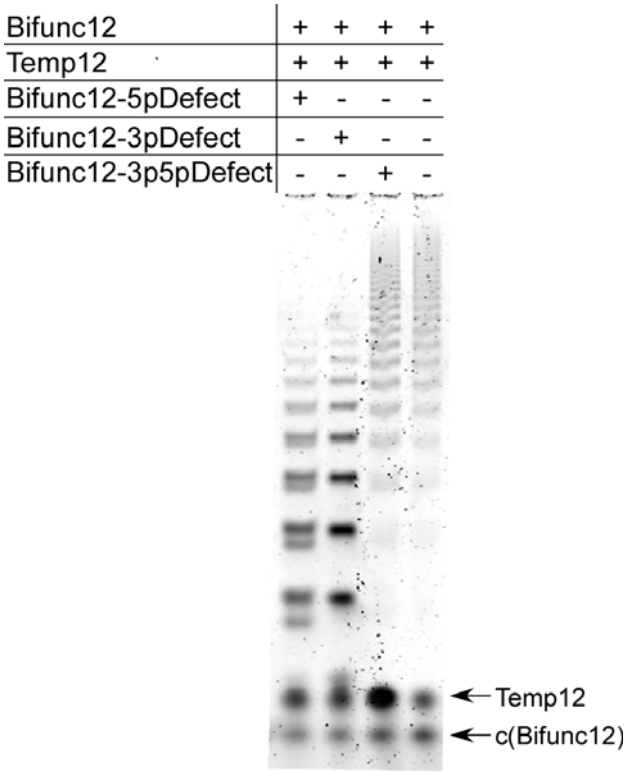


Figure 3.5. Effects of interfering strands on the ethidium-promoted, Temp12-templated polymerization of Bifunc12. Where indicated, reactions contained a full molar equivalent (with respect to Bifunc12) of Bifunc12-5pDefect, Bifunc12-3pDefect, or Bifunc12-3p5pDefect. c(Bifunc12) denotes cyclized Bifunc12. Bifunc12-3p5pDefect co-elecrophoreses with Temp12.

The polymerization of Bifunc12 is template-dependent (Figure 3.4). However, the ligation reaction occurs via an initial reversible step, and a hydrolytically stable linkage is formed only after treatment with cyanoborohydride. By delaying the introduction of cyanoborohydride, the reversible linkage can be exploited to recover cyclic material. When Temp12 and Bifunc12 are both present, polymerization occurs, even if cyanoborohydride is initially present (Figure 3.6). If Temp12 is initially absent, cyanoborohydride efficiently and irreversibly traps c(Bifunc12); the introduction of Temp12 after one day of reduction does not result in the production of linear polymer. If, however, Bifunc12 is equilibrated without Temp12 or cyanoborohydride overnight, the introduction of Temp12 on the second day can recover some reversibly cyclized Bifunc12, resulting in the restoration of polymer formation.

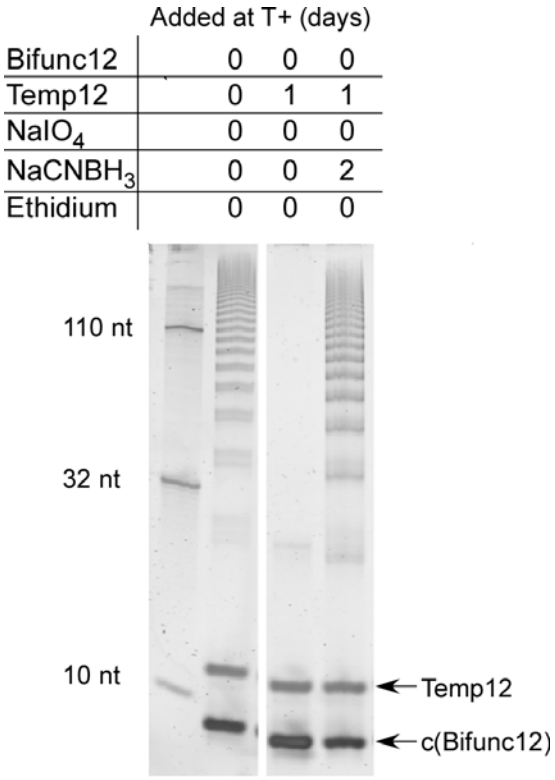


Figure 3.6. Salvage of cyclized Bifunc12 into linear polymer. The reagents were added in the order indicated, then the reactions were reduced for 2 days at room temperature. c(Bifunc12) denotes cyclized Bifunc12.

3.7. Conclusion

The work presented here extends previous studies of reversible linkages in single ligation reactions to the production of long polymers. Such reactions could be of considerable utility in the formation of the first informational polymers.[20, 21] Additionally, we have shown that a small molecule can drive the assembly of these polymers, amounting to the promotion of reversible covalent bond formation by noncovalent ligand binding. This amounts to a synthetically facile dynamic covalent chemistry system, coupled with an informational polymer that binds well-known nucleic acid dyes. Given the previously demonstrated utility of dynamic covalent chemistry and nucleic acid dye binding in controlled drug delivery, as well as nano- and biotechnological applications, the present system provides a readily accessible chemical toolset of considerable potential utility.[44-47]

3.8. References

1. Cech, T.R. and B.L. Bass, *Biological catalysis by RNA.* Annual Review of Biochemistry, 1986. **55**: p. 599-629.

2. Ellington, A. and J. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**: p. 818-822.

3. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**: p. 505-510.

4. Potyrailo, R.A., et al., *Adapting selected nucleic acid ligands (aptamers) to biosensors.* Analytical Chemistry, 1998. **70**(16): p. 3419-3425.

5. Ng, E.W.M., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease.* Nature Reviews Drug Discovery, 2006. **5**(2): p. 123-132.

6. Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.* Nature, 2002. **419**(6910): p. 952-956.

7. Tavitian, B., *In vivo imaging with oligonucleotides for diagnosis and drug development.* Gut, 2003. **52**: p. 40-47.

8. Joyce, G.F., *RNA evolution and the origins of life.* Nature, 1989. **338**: p. 217-224.

9. Gesteland, R.F., T.R. Cech, and J.F. Atkins, *The RNA World*. 3rd ed. 2006, Cold Spring harbor, NY: Cold Spring Harbor Laboratory Press. 768.

10. Ertem, G. and J.P. Ferris, *Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds.* Journal of the American Chemical Society, 1997. **119**: p. 7197 - 7201.

11. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* Journal of Theoretical Biology, 2000. **205**(4): p. 543-562.

12. Leitzel, J.C. and D.G. Lynn, *Template-directed ligation: from DNA towards more versatile templates. The Chemical Record*, 2001. **1**: p. 53-62.

13. Benner, S.A., *Understanding nucleic acids using synthetic chemistry.* Accounts of Chemical Research, 2004. **37**: p. 784-797.

14. Costanzo, G., et al., *Nucleoside phosphorylation by phosphate minerals.* Journal of Biological Chemistry, 2007. **282**(23): p. 16729-16735.

15. Anastasi, C., et al., *RNA: Prebiotic product, or biotic invention?* Chemistry & Biodiversity, 2007. **4**(4): p. 721-739.

16. Rajamani, S., et al., *Lipid-assisted synthesis of RNA-like polymers from mononucleotides.* Origins of Life and Evolution of the Biosphere, 2008. **38**: p. 57–74.

17. Ferris, J.P., et al., *Synthesis of long prebiotic oligomers on mineral surfaces.* Nature, 1996. **381**(6577): p. 59-61.

18. Prabahar, K.J. and J.P. Ferris, *Adenine derivatives as phosphate-activating groups for the regioselective formation of 3',5'-linked oligoadenylates on montmorillonite: Possible phosphate-activating groups for the prebiotic synthesis of RNA.* Journal of the American Chemical Society, 1997. **119**(19): p. 4330-4337.

19. Wang, K.J. and J. Ferris, *Catalysis and selectivity in prebiotic synthesis: Initiation of the formation of oligo(U)s on montmorillonite clay by adenosine-5 '-methylphosphate.* Origins of Life and Evolution of the Biosphere, 2005. **35**(3): p. 187-212.

20. Engelhart, A.E. and N.V. Hud, *Primitive genetic polymers.* Cold Spring Harb. Perspect. Biol., 2010.

21. Hud, N.V., et al., *Addressing the problems of base pairing and strand cyclization in template-directed synthesis - A case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA.* Chemistry & Biodiversity, 2007. **4**(4): p. 768-783.

22. Li, X., et al., *DNA-catalyzed polymerization.* Journal of the American Chemical Society, 2002. **124**(5): p. 746-747.

23. Kleiner, R.E., et al., *DNA-templated polymerization of side-chain-functionalized peptide nucleic acid aldehydes.* Journal of the American Chemical Society, 2008. **130**(14): p. 4646-4659.

24. Rosenbaum, D.M. and D.R. Liu, *Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes.* Journal of the American Chemical Society, 2003. **125**(46): p. 13924-13925.

25. Ermolinsky, B.S. and S.N. Mikhailov, *Periodate oxidation in chemistry of nucleic acids: Dialdehyde derivatives of nucleosides, nucleotides, and oligonucleotides (review).* Bioorganicheskaya Khimiya, 2000. **26**(7): p. 483-504.

26. Bell, N.M., R. Wong, and J. Micklefield, *A Non-Enzymatic, DNA Template-Directed Morpholino Primer Extension Approach.* Chemistry-a European Journal, 2010. **16**(7): p. 2026-2030.

27. Zhou, L.H., et al., *Endowing RNase H-inactive antisense with catalytic activity: 2-5A-morphants.* Bioconjugate Chemistry, 2005. **16**(2): p. 383-390.

28. Wang, Z.F., et al., *Convergent synthesis of ribonuclease L-active 2 ',5 '-oligoadenylate-peptide nucleic acids.* Bioorganic & Medicinal Chemistry Letters, 2000. **10**(12): p. 1357-1360.

29. Bellon, L., et al., *Morpholino-linked ribozymes: A convergent synthetic approach.* Journal of the American Chemical Society, 1996. **118**(15): p. 3771-3772.

30. Kojima, N., et al., *Enhanced reactivity of amino-modified oligonucleotides by insertion of aromatic residue.* Bioorganic & Medicinal Chemistry Letters, 2006. **16**(19): p. 5118-5121.

31. Goodwin, J.T. and D.G. Lynn, *Template-directed synthesis: Use of a reversible reaction.* J. Am. Chem. Soc., 1992. **114**: p. 9197-9198.

32. Jain, S.S., et al., *Enzymatic behavior by intercalating molecules in a template-directed ligation reaction.* Angewandte Chemie, International Edition in English, 2004. **43**: p. 2004-2008.

33. Horowitz, E.D., et al., *Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(12): p. 5288-5293.

34. Bomann, M.D., I.C. Guch, and M. Dimare, *A mild, pyridine-borane-based reductive amination protocol.* Journal of Organic Chemistry, 1995. **60**(18): p. 5995-5996.

35. Page, M.I. and W.P. Jencks, *Entropic contributions to rate accelerations in enzymic and intramolecular reactions and chelate effect.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(8): p. 1678-&.

36. Guthrie, J.P., *Carbonyl addition-reactions - factors affecting hydrate-hemiacetal and hemiacetal-acetal equilibrium-constants.* Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1975. **53**(6): p. 898-906.

37. Angyal, S.J., *Composition and conformation of sugars in solution.* Angewandte Chemie-International Edition, 1969. **8**(3): p. 157-&.

38. Luo, P.Z., et al., *Analysis of the structure and stability of a backbone-modified oligonucleotide: Implications for avoiding product inhibition in catalytic*

template-directed synthesis. Journal of the American Chemical Society, 1998. **120**(13): p. 3019-3031.

39. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 1999. **284**(5423): p. 2118-2124.

40. Schlegel, M.K., et al., *Insight into the high duplex stability of the simplified nucleic acid GNA.* Angewandte Chemie-International Edition, 2009. **48**(5): p. 960-963.

41. Zhang, L.L., A. Peritz, and E. Meggers, *A simple glycol nucleic acid.* Journal of the American Chemical Society, 2005. **127**(12): p. 4174-4175.

42. Schneider, K.C. and S.A. Benner, *Oligonucleotides containing flexible nucleoside analogs.* Journal of the American Chemical Society, 1990. **112**(1): p. 453-455.

43. Summerton, J. and D. Weller, *Morpholino antisense oligomers: Design, preparation, and properties.* Antisense & Nucleic Acid Drug Development, 1997. **7**(3): p. 187-195.

44. Benvin, A.L., et al., *Fluorescent DNA nanotags: Supramolecular fluorescent labels based on intercalating dye arrays assembled on nanostructured DNA templates.* Journal of the American Chemical Society, 2007. **129**(7): p. 2025-2034.

45. Ozhalici-Unal, H. and B.A. Armitage, *Fluorescent DNA Nanotags Based on a Self-Assembled DNA Tetrahedron.* Acs Nano, 2009. **3**(2): p. 425-433.

46. Sreenivasachary, N. and J.M. Lehn, *Structural selection in G-quartet-based hydrogels and controlled release of bioactive molecules.* Chemistry-an Asian Journal, 2008. **3**: p. 134-139.

47. Godin, G., et al., *Reversible formation of aminals: a new strategy to control the release of bioactive volatiles from dynamic mixtures.* Chemical Communications, 2010. **46**(18): p. 3125-3127.

CHAPTER 4

A SELECTIVE, EASILY SYNTHESIZED G-QUADRUPLEX LIGAND[c]

4.1. Introduction

The discovery of G-quadruplex structures in nucleic acid sequences associated with cancers, such as the c-myc proto-oncogene and telomeric DNA,[1-6] has created intense interest in G-quadruplexes as potential drug targets. These four-stranded structures, with planar G-tetrads (Figure 4.1a), represent appealing DNA targets, as they are structurally distinct from the Watson-Crick duplex of most genomic DNA. Small molecules with high affinity and high selectivity for G-quadruplexes have even begun to show medicinal promise. For example, quarfloxin, a ligand targeting the G-quadruplex of the c-myc promoter, is presently in clinical trials.

---

[c] This chapter was adapted from the pre-peer reviewed version of previously published work. (Cetinkol, ÖP; Engelhart, AE; Nanjunda, RK; Wilson, WD; Hud, NV. "Submicromolar, selective G-quadruplex ligands from one pot: thermodynamic and structural studies of human telomeric DNA binding by azacyanines." *ChemBioChem* (2008). 9 (12) 1889-1892.). It is reproduced with permission.

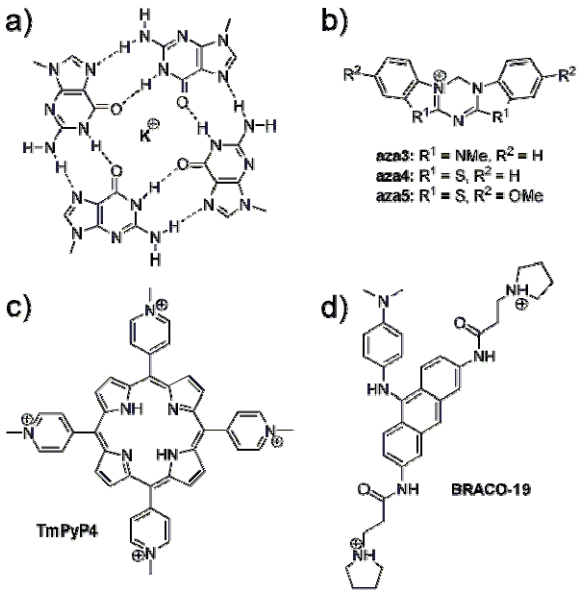


Figure 4.1. Structures of a G-tetrad and G-quadruplex ligands. a) a G-tetrad and the G-quadruplex ligands b) azacyanines, c) TmPyP4, and d) BRACO-19.

Most investigators seeking small molecules as ligands for G-quadruplex DNA have approached the problem of ligand design on two fronts: first, via the use of heterocycles with a relatively large and planar surface area, which maximizes stacking with the ca. 1 nm^2 surface of a G-tetrad (e.g., TmPyP4, Figure 4.1c).[7, 8] Second, many have used multiple charges to enhance the solubility of potential ligands (often necessary for large hydrophobic surfaces) and to increase electrostatic interactions with the high-charge density G-quadruplex (e.g., TmPyP4 and BRACO-19, Figure 4.1c, d).[9, 10] While these strategies have resulted in several high affinity ligands for G-quadruplexes, most ligands do not exhibit high selectivity over duplex DNA.

We have taken a different approach to targeting the G-quadruplex . Based on our previous discovery that a planar molecule larger than a typical DNA intercalator can

selectively bind purine-purine base pairs,[11-13] we hypothesized that planar, monocationic molecules that are marginally too large to intercalate a Watson-Crick duplex, such as a bispurine analogue, might selectively bind G-quadruplexes. We report a new class of selective, submicromolar quadruplex ligands with a facile synthetic route: the azacyanines (Figure 4.1b). The synthesis of azacyanines was previously reported by Kurth and coworkers; the route is one-pot and workup is by filtration.[14, 15] The route is general and succeeds for aminobenzimidazoles and aminobenzothiazoles. The synthetic ease makes the class extraordinarily amenable to library preparation and high-throughput screening.

4.2. Experimental Procedures

4.2.1. Materials

The oligonucleotide sequences d[TTGGG(TTAGGG)$_3$A] (tel24), d[AAAGGG(TTAGGG)$_3$AA] (tel26), d[CGCATATATGCG] (dd1) and d[CGCAAATTTGCG] (dd2) were purchased from Integrated DNA Technologies (Coralville, IA). Stock concentrations were calculated by using the extinction coefficients provided by the supplier. Calf thymus DNA was purchased from Sigma-Aldrich (Milwaukee, WI) and an A_{260} of 12824 M^{-1} cm^{-1} was used to calculate the stock concentration.[16] All samples were prepared in 25mM potassium phosphate buffer (pH 7.0) and 70 mM KCl, unless otherwise stated. Samples were annealed by heating to 95^{o}C in a water bath and then cooling overnight to room temperature in a Styrofoam box prior to each experiment.

aza3-5 were synthesized as described by Huang et al.[14, 15] and characterized by 1D (^{1}H, ^{13}C) and 2D (HSQC, HMBC) NMR spectroscopy, accurate mass spectrometry,

and elemental analysis. Aza3 was also characterized by X-Ray crystallography (crystallization was effected from a saturated methanol-water solution at RT). Extinction coefficients were determined by UV-vis spectrophotometry, using at least 5 different samples with different concentrations of the compound.

4.2.2. Circular Dichroism and UV-Vis Spectroscopy

CD spectra were acquired on a JASCO J-810 CD spectropolarimeter equipped with a Peltier temperature control unit. Spectra were acquired using a 5 mm strain-free rectangular cell. CD melting profiles were acquired as full spectra from 5 to 95^{o}C in 1^{o}C steps, at <1^{o}C/minute. The buffer spectrum was collected under these conditions and used to correct the sample spectra. Melting curves were obtained by fitting each spectrum to a two-state model of a linear superposition of the 5^{o}C and 95^{o}C spectra. UV-Vis absorbance measurements were performed using a HP 8453 UV-Vis diode array spectrophotometer equipped with an Agilent 89090A Peltier temperature control unit. The melting profiles were acquired from 5 to 95^{o}C with a rate of <1^{o}C/min for all four traces (two heating and two cooling). UV melting profiles were obtained by monitoring the absorbance change at 295 nm, as well as by the aforementioned two-state model.

4.2.3. Fluorescence Titrations

All fluorescence measurements were performed on a Shimadzu RF-5301PC spectrofluorophotometer (Columbia, MD) at 25^{o}C. Titrations were carried out by making incremental additions of nucleic acid-1µM aza3 stock to a solution containing only 1 µM aza3. Aza3 was excited at 324 nm and emission spectra were collected from 325 to 700 nm. Excitation and emission slits were both set to 1.5 nm. Association constants were obtained by fitting integrated fluorescence intensity versus nucleic acid concentration to a single exponential as described previously.[17]

4.2.4. Surface Plasmon Resonance Measurements

Surface plasmon resonance experiments were performed with a BIACORE 2000 instrument. Biosensor experiments were conducted in degassed HEPES buffer (10 mM HEPES, 100 mM KCl, 3 mM EDTA, 0.00005 v/v of 10% P20 BIACORE surfactant, pH 7.4) at 25°C. The 5'-biotin-labeled tel24, tel26, d(CGAATTCGTTTTCGAATTCG) (dd3), d(CGCGCGCGTTTTCGCGCGCG) (dd4) and d(CCATATATATATATATAGCCCCCGCTATATATATATATATGG) (dd5) sequences were purchased from Midland Certified Reagent Company with HPLC purification. The DNA sequences were immobilized on a streptavidin-derivatized gold chip (SA chip from BIACORE) by manual injection of DNA stock solution with a flow rate of 1 µl/min until the SA chip surface was immobilized with the desired amount of DNA. Typically, a series of different ligand concentrations ranging from 10 nM to 10 µM were injected onto the chip with a flow rate of 50 µl/min for a period of 5 min followed by a dissociation period of 10 min. After each cycle, the chip surface was regenerated with a 20-sec injection of pH 2.0 glycine solution and multiple 1-min buffer injections. The number of binding sites and association constants were obtained from fitting plots of RU_{obs} versus C_{free}. To obtain the association constants the data generated were fitted to a three-site interaction model using Kaleidagraph for nonlinear least squares optimization of the binding parameters using the following equation:

$$RU_{obs} = (RU_{max} \times (K_1 \times C_{free} + 2K_1 \times K_2 C^2_{free} + 3K_1 \times K_2 \times K_3 C^3_{free})) / (1 + K_1 \times C_{free} + K_1 \times K_2 C^2_{free} + K_1 \times K_2 \times K_3 C^3_{free})$$

Where K1, K2 and K3 are macroscopic equilibrium constants for three types of binding sites, RU_{obs} is the SPR response at the steady state level, RU_{max} is the maximum

SPR response for binding one molecule per binding site, and C_{free} is the compound concentration in solution.

4.2.5. Isothermal Titration Calorimetry

Isothermal titration calorimetry experiments were performed using a Microcal VP-ITC. 5 µl of 200 µM aza3 was injected into 1.4 ml of 100 µM tel24 in the cell. Control titrations were conducted by injecting aza3 into the sample cell containing only buffer under the same conditions. The average heat of dilution was subtracted from the uncorrected heats, and the enthalpy of binding was calculated from the integrated heats and averaged. The experiments were done in duplicate, and the average of the first 50 injections for each run was used in order to obtain the ΔH distribution.

4.2.6. NMR Spectroscopy

NMR experiments were performed on a Bruker DRX-500 at 25^{o}C. ^{1}H spectra of the DNA samples collected in H_2O were used to verify proper folding of the G-quadruplex based upon imino proton chemical shifts. The 3-9-19 WATERGATE pulse sequence was used for water suppression. Aza3 titrations were performed by the dissolution of lyophilized aliquots of aza3 into a 450 µl of a 2 mM tel24 sample in D_2O (0.05 equivalents at each step). 128 transients were collected for each 1D ^{1}H. 2D NOESY spectra were collected at selected aza3 to tel24 ratios with 1024 points in F_2 and 512 blocks in F_1. The line broadening of the 1D ^{1}H spectrum was analyzed by fitting resolved resonances by Lorentzian functions with a linear offset. 2D NOESY cross peaks were assigned based on the assignments reported by Luu et al. The software packages SPARKY and SpinWorks were used to analyze 2D spectra. 1D rows were extracted from

2D NOESY data and analyzed by Lorentzian fits. The height changes of tel24 cross peaks in spectra acquired during aza3 titrations were normalized with respect to spectra acquired in the absence of aza3.

4.3. Aza3 is a High-Affinity, Selective G-Quadruplex Ligand

We investigated the binding of azacyanines to a G-quadruplex sequence, based on the human telomeric repeat $d(TTAGGG)_n$, for which a solution-state structure has been reported: $d(TTGGG(TTAGGG)_3A)$ (tel24).[18-20] To ascertain discrimination against duplex DNA, we also conducted binding studies with calf thymus DNA and two isomeric oligonucleotides d(GCGCATATATGCGC) (dd1) and d(GCGCAAATTTGCGC) (dd2). dd2 was specifically designed with an A-tract sequence element that is associated with a narrow minor groove and typically favors interaction with groove-binding ligands.[21]

The binding of aza3 to tel24 was first examined by UV-Vis and CD spectroscopy. In the presence of tel24, the longest wavelength absorption band of aza3 was red-shifted and hypochromic (Figure 4.2a), and a weak induced band was observed in the CD spectrum (Figure 4.2b). Monitoring the UV-Vis absorption of the G-quadruplex as a function of temperature also revealed that the presence of 1 equivalent of aza3 enhanced the thermal stability of tel24 relative to the unliganded species, raising the T_M from 65 to 67°C. The enhanced thermal stability of tel24 in the presence of aza3 and the spectral changes observed for aza3 in the presence of tel24 are all characteristic of ligand binding in the chiral environment of DNA.

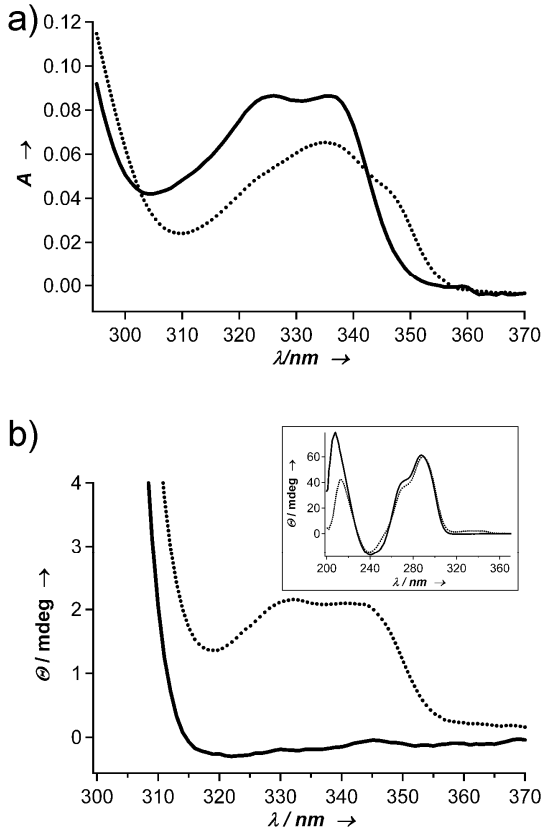
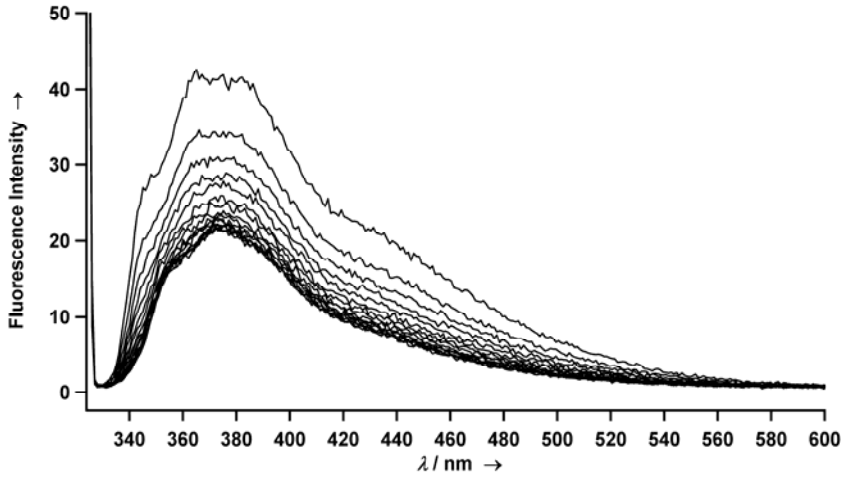


Figure 4.2. Long wavelength regions of the UV-vis and CD spectra of aza3 in the presence and absence of tel24. a) UV-vis spectra of 1:1 mixture of aza3 and tel24 at 5ºC (dashed line) and 95ºC (solid line). b) CD spectra of 1:1 mixture of aza3 and tel24 (dashed line) and of tel24 in the absence of aza3 (solid line). Insert shows full CD spectra.

The binding and structural selectivity of aza3 was characterized quantitatively by fluorescence titrations of aza3 with various nucleic acids (Figure 4.3). The ligand exhibits a profound structural preference, with ca. 100-fold selectivity for G-quadruplex vs. duplex DNA (Table 4.1). Aza3 is near what is considered the low end of binding constants (10^{3-4}) for Watson-Crick ligands and near the high end for quadruplex ligands (10^{6-7}); this result supports our size-selection hypothesis.[22] Additionally, aza3 binds more weakly to dd2 than dd1, providing some evidence against groove-binding.

a)



b)

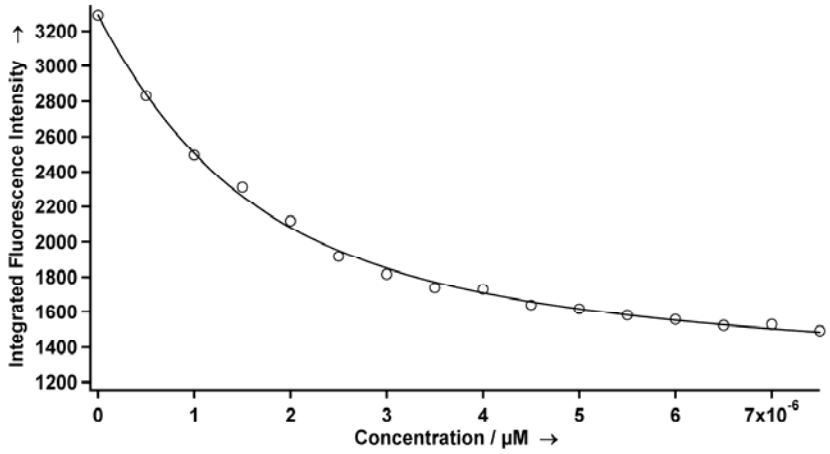


Figure 4.3. Representative fluorescence titration data. a) Fluorescence intensity measurements for 1µM solutions of aza3 in the presence of increasing concentrations of tel24. b) A plot of integrated fluorescence intensities of spectra shown in panel a as a function of DNA concentration (in units of oligonucleotide strand). Curve shown represents a single-exponential fit to the experimental data (circles).

Table 4.1. Association constants for aza3 with various DNAs, as determined by fluorescence spectroscopy.

Sequence	**Association Constant (M^{-1})**
tel24 (d(TTGGG(TTAGGG)$_3$A))	$1.28 \times 10^6 \pm (2.5 \times 10^5)$
Calf Thymus DNA	$9.2 \times 10^3 \pm (1.7 \times 10^3)$
dd1 (d(GCGCATATATGCGC))	$6.13 \times 10^3 \pm (1.2 \times 10^3)$
dd2 (d(GCGCAAATTTGCGC))	$1.14 \times 10^4 \pm (2.0 \times 10^3)$

4.4. Aza3 Binds the Human Telomere G-Quadruplex by a Mixed Intercalative-External Stacking Mode

Titration of aza3 into a 2 mM sample of tel24 revealed that the ligand was in intermediate exchange on the NMR timescale. A number of resonances broaden in a site-specific fashion, consistent with the local chemical environment of various residues being perturbed differentially by ligand binding. 1D ^{1}H NMR spectra of tel24 in the presence of increasing concentrations of aza3 are shown in Figure 4.4. Resonance line broadening was characterized by the fitting of Lorentzian functions to resolved resonances in 1D spectra, and by measuring the heights of aromatic-H1' crosspeaks in 2D NOESY spectra (Figures 4.5a and b).

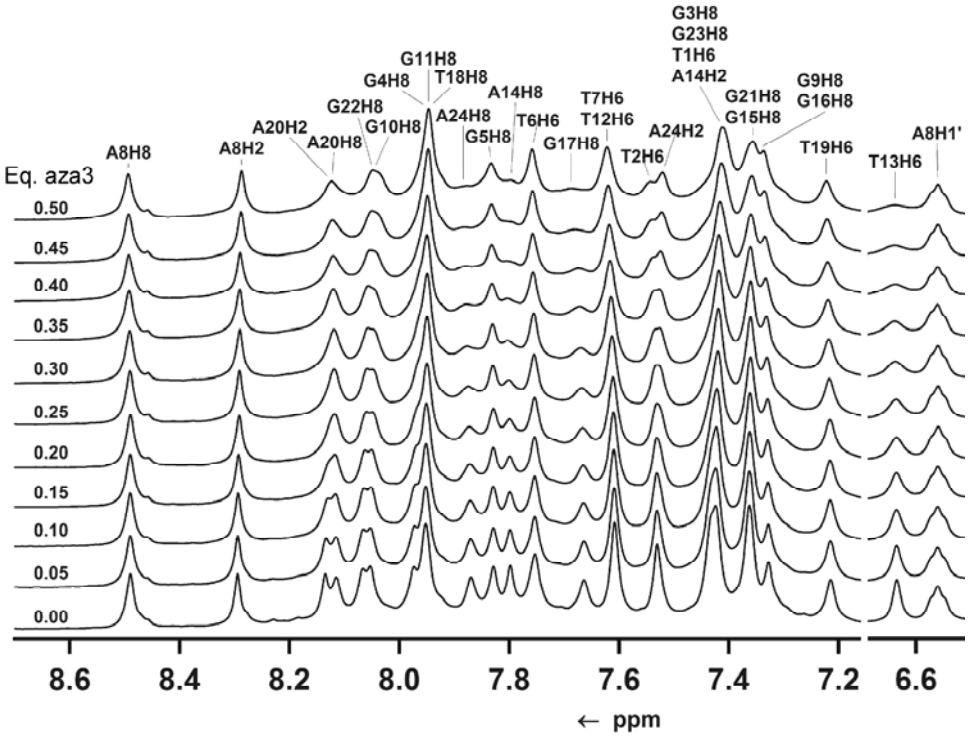


Figure 4.4. Aromatic region of ^{1}H-NMR spectra of tel24 in the presence of 0.00 to 0.50 molar equivalents of aza3.

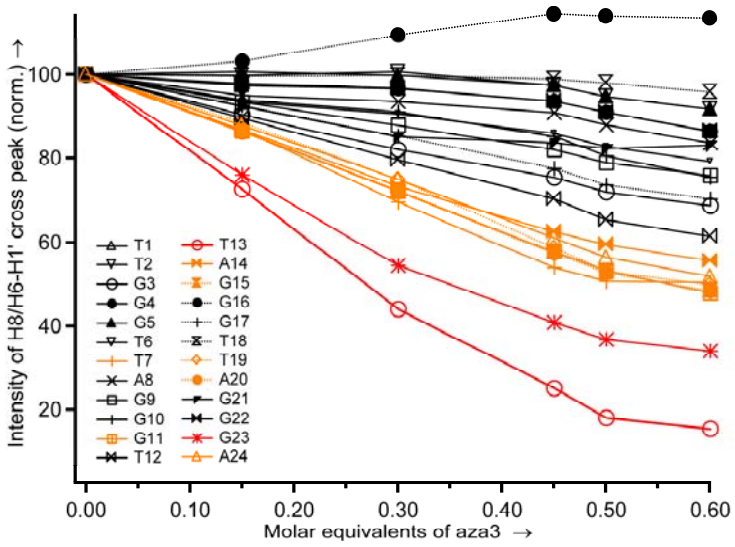


Figure 4.5. Attenuation of tel24 H6/H8-H1' NOE peak height by aza3. All cross peaks measured were in the direct dimension of a 2D NOESY for H1'.

As a qualitative method of determining the ligand binding site, the initial rates of decrease of NOESY cross peak intensities was plotted vs. the aza3 concentration (Figure 4.5) and their slopes were compared. The residues were separated into three groups based upon their apparent sensitivity to the presence of aza3. The solution structure of tel24 is shown in Figure 4.6 with residues colored according to these three groups.[20]

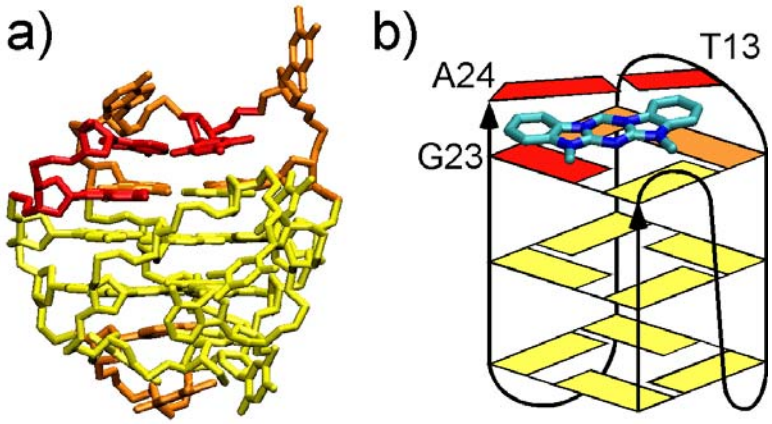


Figure 4.6. Structural information on tel24-aza3 binding determined from NMR experiments. a) The solution state structure of Tel24 (PDB ID 2GKU) with residues colored by rate of NOE attenuation per equivalent aza3 (red, high; orange, medium; yellow, low). For clarity, hydrogens and phosphate oxygens are not shown. b) A schematic representation of aza3 bound tel24, based upon NOE attenuation data, in which aza3 is positioned between the A·T base pair of T13 and A24 and G23 of the terminal G-quartet. For clarity, the loop bases are not indicated, except for those of T13 and A24.

The NMR data indicate the presence of two aza3 binding sites, one atop each exterior G-tetrad. Luu et al. reported that an A·T pair is formed by T1 and A20, and a reverse A·T pair is formed by T13 and A24, each being stacked on G-tetrads on opposite sides of the quadruplex. Interestingly, broadening occurs for both the aromatic residues of the external G-tetrads and these A·T base pairs (Figure 4.6), providing support for a

mixed intercalation-exterior stacking mode of binding (i.e., between the exterior tetrads and the capping A·T base pairs of the loops). The NMR data do not support a groove-binding mode of interaction, consistent with our results from fluorescence titration of aza3 with duplex DNA.

4.5. G-Quadruplex Binding and Duplex Selectivity is General to the Azacyanine Structural Class

"Model-free" ITC titrations (i.e., at very low ligand:binding site ratios) of aza3 into tel24 were performed to better characterize the thermodynamics of the interaction. A ΔH of -2.7±0.14 kcal/mol was obtained (Figure 4.7). With the 1.3×10^6 association constant determined by fluorescence titrations, this datum implies a ΔG of -8.3kcal/mol and a remarkable $-T\Delta S$ of -5.6 kcal/mol. There is a paucity of thermodynamic data on quadruplex-ligand interactions, but such entropy-driven binding has been reported previously for a remarkably different, tetracationic ligand, TmPyP4.[23]

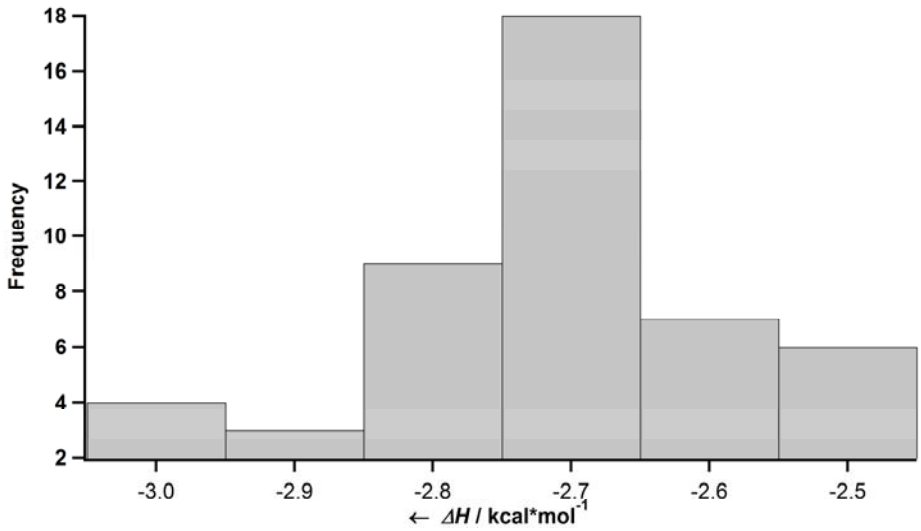


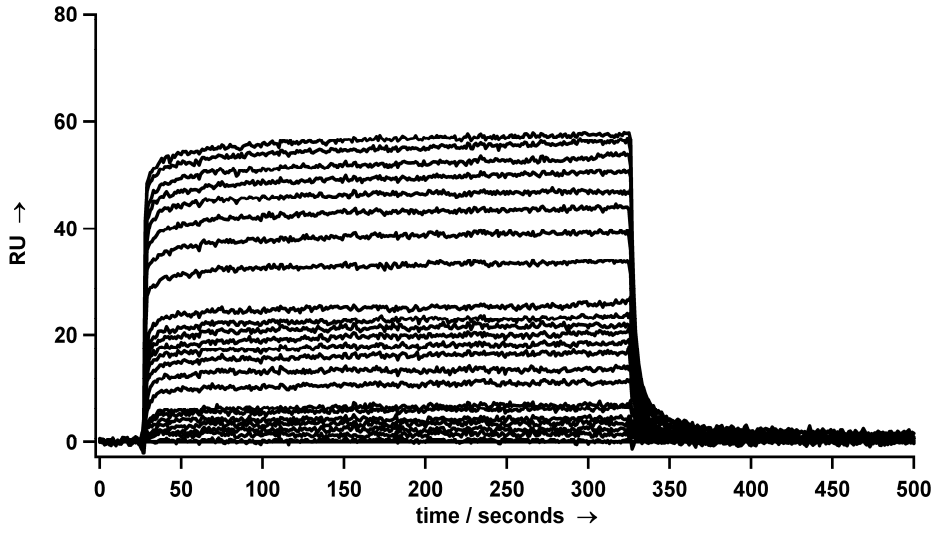
Figure 4.7. *ΔH* values for tel24-aza3 binding, measured by isothermal titration calorimetry. The data shown were obtained by performing two titrations of 50 injections each.

Finally, the binding of a broader spectrum of azacyanines (aza3-5) to tel24 was characterized by SPR. All ligands tested exhibited similar association constants (Table 4.2), suggesting that G-quadruplex binding is general to this class of compounds. Additionally, we have tested the binding affinity of all three azacyanine compounds to tel26 (Table 4.2), a twenty-six nucleotide DNA sequence that was also derived from the human telomere sequence and which has a fold similar to tel24.[19] All three azacyanine compounds showed similar binding affinities to tel26 and tel24. Interestingly, the SPR response curves were fit well for both DNA structures only by two- or three-site models, with one tight binding site and one or two approximately degenerate secondary sites for both sequences. (Table 4.2, Figures 4.8-4.9).

Table 4.2. Association constants determined by SPR for aza3-5 with d(TTGGG(TTAGGG)$_3$A) (tel24) and d(AAAGGG(TTAGGG)$_3$AA) (tel26). In addition to the K_1 value shown for the single strong binding site, all compounds have one or two much weaker binding sites with $K = 2 \times 10^5 \pm 1 \times 10^5\ M^{-1}$. Association constants for the duplex sequences were not determined because of their very low steady-state response even for high compound concentrations.

Compound	Sequence	Association Constant (M^{-1})
aza3	tel24	3.11×10^6
	tel26	4.09×10^6
aza4	tel24	2.97×10^6
	tel26	3.04×10^6
aza5	tel24	2.87×10^6
	tel26	4.74×10^6

a)



b)

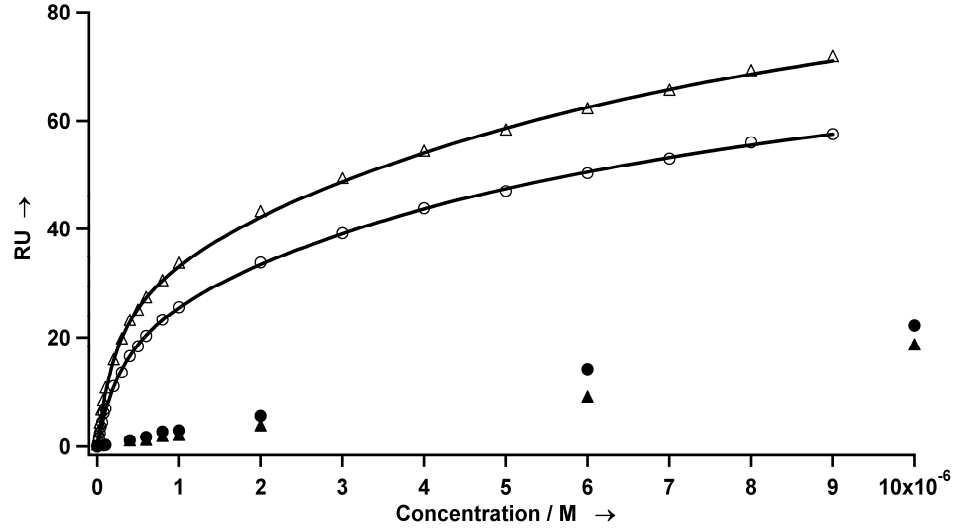


Figure 4.8. Representative SPR sensorgram data. a) SPR sensorgram for binding of aza3 to tel24. b) Response averages from the steady state regions of the sensorgrams of aza3 binding to different DNA sequences (open triangles: tel26, open circle: tel24, filled circle:dd4, filled triangle: dd5). The curves represent the best fit of the response averages from the steady state regions of the sensorgrams to three different binding site models.

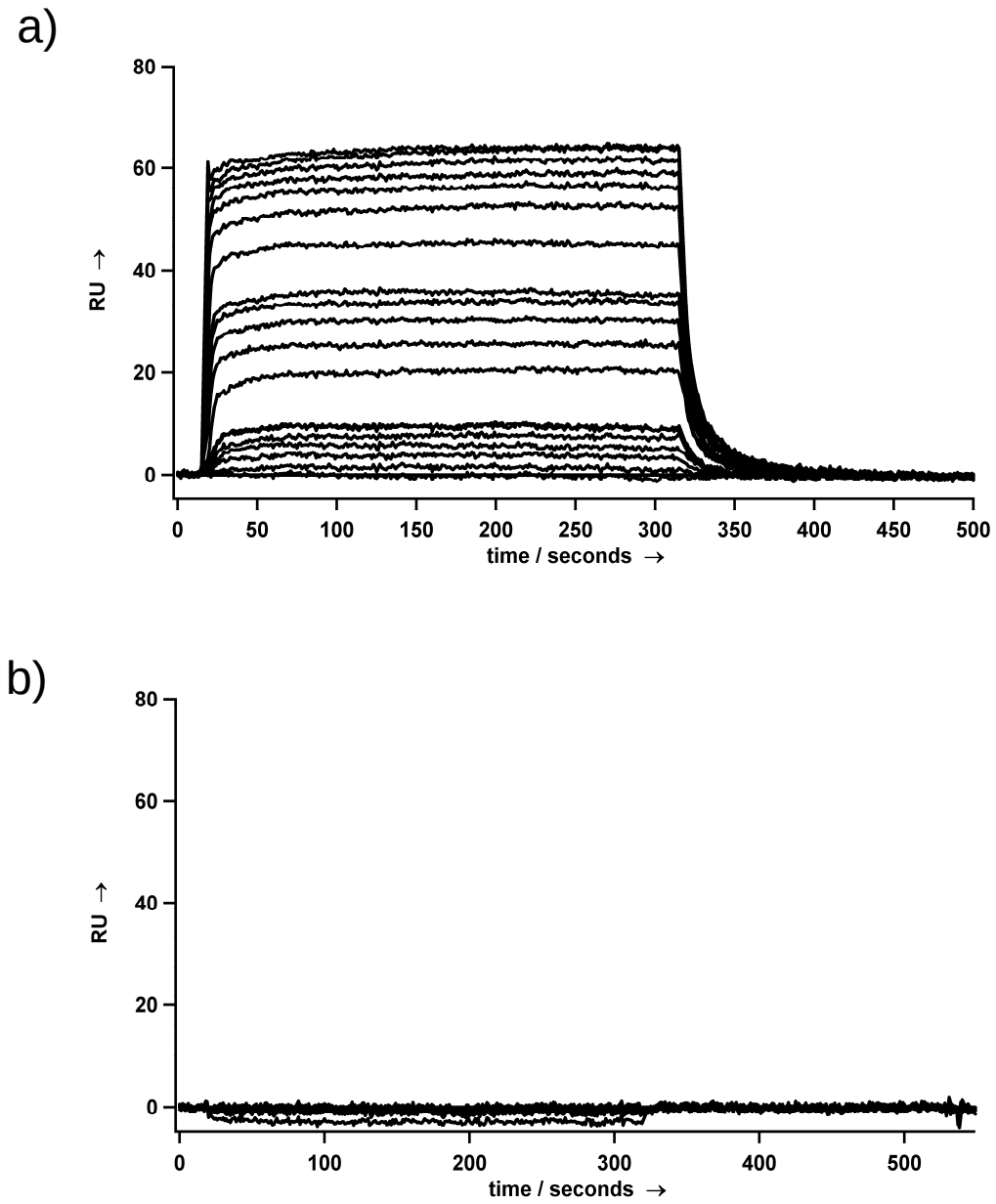


Figure 4.9. Within the SPR detection limit, no dd3-aza5 interaction is observed. SPR sensorgrams for binding of aza5 to a) tel24 and b) dd3.

4.6. Conclusion

G-quadruplex ligands hold great promise for use in the treatment of human disease. A common and valid critique of their medicinal utility relates to the gravity and wide range of off-target effects associated with binding to Watson-Crick DNA. Only by ameliorating such effects is the G-quadruplex a viable drug target. We have shown that a bispurine analogue, which is only slightly too large to intercalate a purine-pyrimidine base pair, exhibits strong and selective association with a G-quadruplex. Kurth's synthetic route to azacyanines provides a general and perhaps the most practical means thus far for the preparation of selective G-quadruplex ligands. A promising avenue of future research will be the decorating of azacyanines with the appropriate functional groups to bind specific promoter-related quadruplexes selectively.

4.7. References

1. Guittat, L., et al., *Targeting human telomerase for cancer therapeutics.* Cytotechnology, 2004. **45**(1-2): p. 75-90.

2. Keniry, M.A., *Quadruplex structures in nucleic acids.* Biopolymers, 2000. **56**(3): p. 123-146.

3. Rezler, E.M., D.J. Bearss, and L.H. Hurley, *Telomeres and telomerases as drug targets.* Current Opinion in Pharmacology, 2002. **2**(4): p. 415-423.

4. Rezler, E.M., D.J. Bearss, and L.H. Hurley, *Telomere inhibition and telomere disruption as processes for drug targeting.* Annual Review of Pharmacology and Toxicology, 2003. **43**: p. 359-379.

5. Satyanarayana, A., M.P. Manns, and K.L. Rudolph, *Telomeres, telomerase and cancer - An endless search to target the ends.* Cell Cycle, 2004. **3**(9): p. 1138-1150.

6. Siddiqui-Jain, A., et al., *Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.* PNAS, 2002. **99**(18): p. 11593-11598.

7. Han, F.X.G., R.T. Wheelhouse, and L.H. Hurley, *Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition.* Journal of the American Chemical Society, 1999. **121**(15): p. 3561-3570.

8. Izbicka, E., et al., *Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells.* Cancer Research, 1999. **59**(3): p. 639-644.

9. Harrison, R.J., et al., *Evaluation of by disubstituted acridone derivatives as telomerase inhibitors: the importance of G-quadruplex binding.* Bioorganic & Medicinal Chemistry Letters, 2004. **14**(23): p. 5845-5849.

10. Koeppel, F., et al., *Ethidium derivatives bind to G-quartets, inhibit telomerase and act as fluorescent probes for quadruplexes.* Nucleic Acids Research, 2001. **29**(5): p. 1087-1096.

11. Persil, Ö., et al., *Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding.* Journal of the American Chemical Society, 2004. **126**(28): p. 8644-8645.

12. Jain, S.S., M. Polak, and N.V. Hud, *Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne-driven duplex disproportionation.* Nucleic Acids Research, 2003. **31**(15): p. 4608-4615.

13. Polak, M. and N.V. Hud, *Complete disproportionation of duplex poly(dT).poly(dA) into triplex poly(dT).poly(dA).poly(dT) and poly(dA) by coralyne.* Nuc. Acid Res., 2002. **30**: p. 983-992.

14. Haddadin, M.J., J.M. Kurth, and M.M. Olmstead, *One-step synthesis of new heterocyclic azacyanines.* Tetrahedron Letters, 2000. **41**(30): p. 5613-5616.

15. Huang, K.S., et al., *Synthesis and reactions of some heterocyclic azacyanines.* Journal of Organic Chemistry, 2001. **66**(4): p. 1310-1315.

16. Chaires, J.B., *Structural selectivity of drug-nucleic acid interactions probed by competition dialysis*, in *DNA Binders and Related Subjects*, M.J. Waring and J.B. Chaires, Editors. 2005, Springer-Verlag Berlin: Berlin. p. 33-53.

17. Horowitz, E.D. and N.V. Hud, *Ethidium and proflavine binding to a 2',5'-linked RNA duplex.* Journal of the American Chemical Society, 2006. **128**(48): p. 15380-15381.

18. Dai, J.X., et al., *Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence.* Nucleic Acids Research, 2007. **35**(15): p. 4927-4940.

19. Dai, J.X., et al., *Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation.* Nucleic Acids Research, 2007. **35**(7): p. 2440-2450.

20. Luu, K.N., et al., *Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold.* Journal of the American Chemical Society, 2006. **128**(30): p. 9963-9970.

21. Tanious, F.A., et al., *DNA Sequence Dependent Monomer-Dimer Binding Modulation of Asymmetric Benzimidazole Derivatives.* J. Am. Chem. Soc., 2004. **126**: p. 143-153.

22. White, E.W., et al., *Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge.* Biophysical Chemistry, 2007. **126**(1-3): p. 140-153.

23. Freyer, M.W., et al., *Biophysical studies of the c-MYC NHE III1 promoter: Model quadruplex interactions with a cationic porphyrin.* Biophysical Journal, 2007. **92**(6): p. 2007-2015.

CHAPTER 5

STRONG HYDROGEN BONDING BY 8-AMINOGUANINE[d]

5.1. Introduction

Of the molecular recognition tools available to supramolecular chemists, hydrogen bond complementarity is arguably the most often-used and conceptually easiest to appreciate. Rationally-designed assemblies that use hydrogen bond complementarity are sometimes described as "DNA mimics" or "DNA-like," in reference to the selective hydrogen bonding of Watson-Crick base pairs. Here, we propose that a subclass of hydrogen bonds could further expand the supramolecular toolbox. Using a model system of oligonucleotides with modified guanine residues, we demonstrate the ability of strong hydrogen bonds to drive duplex formation with non-Watson-Crick base pairs.

Low-barrier hydrogen bonds (LBHBs),[1-9] short-strong hydrogen bonds, and Speakman-Hadži[10, 11] hydrogen bonds are names that have been applied to a type of strong hydrogen bond characterized by the following: a) proton-bound isoenergetic tautomers (sometimes described as pKa matching), b) an unusually strong interaction energy (≥10 kcal/mol), c) a far downfield ^{1}H-NMR chemical shift, and d) a barrier to proton transit that lies below or near the zero-point energy.[12] From a theoretical standpoint, two features contribute to the stability of these systems: electronic

---

[d] This chapter was adapted from previously published work. (Engelhart, AE; Morton, TH; Hud, NV. "Evidence of strong hydrogen bonding by 8-amino-guanine." *Chem. Comm*. (2009) 647-649.) It is reproduced by permission. © Royal Society of Chemistry (http://pubs.rsc.org/en/Content/ArticleLanding/2009/CC/b818409g).

stabilization due to proton binding; and a lowered zero-point energy, due to closely spaced vibrational levels for the asymmetric stretch.

Such phenomena could promote the self-assembly of supramolecular complexes with additional specificity and stability. In particular, the requirement for pKa matching could make such interactions more selective for driving the assembly of specific structures versus hydrogen bonding with water (pKa 15.7) and other high-dielectric solvents. While most discussions of this type of bonding have focused on anion-proton-anion interactions, we describe one of the less common cationic examples.

We present evidence that 8-amino-G (X) participates in selective, strong hydrogen bonds within a nucleic acid structure. This mode of molecular recognition has a quantum mechanical component and is complementary yet orthogonal to the electrostatic phenomenon reported by the Eschenmoser laboratory, in which the base pair with the largest ΔpKa (within a certain structural family) exhibited the strongest interaction.[13]

Over 30 years ago, Miles and colleagues demonstrated that poly(rX) forms a symmetric base-paired duplex with an N7 pKa of 9.4 (the N7 pKa of rXMP is 4.8).[14, 15] Using IR spectroscopy, the hemiprotonated poly(rX) self-structure (persisting up to ca. pH 10) was assigned as a duplex with proton-bound X^+X dimers (Figure 5.1a). At more basic pH, poly(rX) was shown to exist as a quadruplex, with the bases forming the familiar G-tetrad structure (Figure 5.1b). Thus, the work by Miles demonstrated that poly(rX) favors the X^+X base pair over G-tetrad formation at neutral pH. This interaction occurs with remarkable selectivity, as illustrated by the pKa shift of >4 units in the poly(rX) duplex as compared to the rXMP monomer.

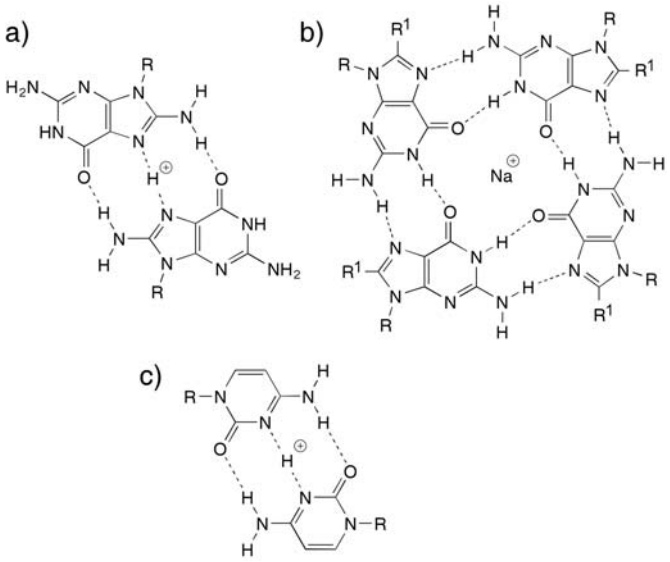


Figure 5.1. Strtuctures of protonated homo-base pairs and G-quartet. a) The X^+X base pair; b) the G-tetrad, in which R^1 is H for guanine and NH_2 for X; c) the C^+C base pair.

5.2. Experimental Procedures

5.2.1. Materials

DNA was prepared on an Expedite Synthesizer using standard protocols. Deprotection was performed over 20H in concentrated aqueous ammonia with 2-mercaptoethanol as recommended by the manufacturer for the 8-amino-dG phosphoramidite (Glen Research)

The DNA was purified on a 3mm 30% denaturing sequencing PAGE gel; the desired band was excised with a clean razor and extracted by the crush-and-soak method. Desalting was effected by SPE. The identity of the oligonucleotide was confirmed by low-resolution ESI-MS: found m/z=1922 (M-H), expected m/z=1922.

Solid phase extraction was performed on an ODS Sep-Pak Plus (Waters). The following solutions were pulled through the bed using a peristaltic pump at 1mL/minute. 10mL 100% MeCN; 10mL 50% MeCN, 50% 100mM triethylammonium acetate, pH 7

(TEAA); 100% TEAA; and the salt-containing DNA in ca. 50mL TEAA. At this point the bed was washed twice with 5mL portions of nanopure water. The DNA was then eluted with 5mL 40% aqueous acetonitrile, the volume was reduced by half on a vacuum centrifuge, and the remaining organic-depleted fraction was frozen and lyophilized.

The resulting solids were resuspended in water and briefly equilibrated over >100eq lithium sulfonate resin (Dowex 50X8). The resin was washed several times with water and the lithium-exchanged DNA was concentrated then lyophilized again.

The buffers used were 10mM acetate or cacodylate with 100mM metal chloride and were prepared at the desired pH at 5-10X. All samples were placed in boiling water for 5 minutes and annealed at 4.9-5X (50 μM total oligonucleotide concentration, 300 μM total base, 50mM buffer, 500mM salt) at 4ºC overnight.

5.2.2. Circular Dichroism

CD was performed on a JASCO J-810. These studies were performed at 10 μM total oligonucleotide concentration (60 μM total base), 10mM buffer, 100mM salt with 350-200nm scans and 1ºC steps. A ramp rate of 0.16ºC - 1ºC/minute was used. The full-wavelength spectrum at each temperature was fit to a two-state model, using the assumption that the 5ºC spectrum in the first heating corresponded to 100% duplex and the 60ºC spectrum corresponded to 0% duplex.

5.2.2. NMR

^{1}H-NMR spectra were collected on a Bruker DRX-500 at 275K (H_2O) or 280K (D_2O) and were the sum of 1024 transients. Spectra in 10% D_2O were observed using a 3-9-19 WATERGATE pulse sequence. Those in D_2O were 99.96% D and observed using a presaturation pulse.

5.3. The Proton-Bound 8-Aminoguanine Homo-Base Pair May Contain a LBHB

We hypothesize that the X^+X self-structure contains a LBHB. The combination of pairing-enforced geometry and pKa matching provide favorable conditions to observe such phenomena. Computational support for this possibility is provided by calculations at B3LYP/6-31G**, which demonstrate that the X^+X self-structure (Figure 5.1a) in the gas phase contains a hydrogen bond with a low barrier to N7-N7 H^+ transit. The gas-phase heat of dissociation for the proton-bound dimer (at 0 K, corrected for BSSE using counterpoise and for changes in zero point energy) is $\Delta H_{0K} = 44.3$ kcal mol^{-1}, and the barrier to proton transit (calculated as the transition state electronic energy minus that of the equilibrium geometry, with the transition state constrained to C_2 symmetry) is 2.5 kcal mol^{-1}. The N7-N7 distance is 2.764 Å at equilibrium, while it is 2.594 Å in the transition state (Table 5.1-5.2). Although these distances are longer than expected for LBHBs in sterically unhindered systems, the N7-N7 separation may be restricted by the flanking asymmetric NH_2-carbonyl hydrogen bonds.

Table 5.1. Coordinates for Hoogsteen proton-bound homodimer of 8-aminoguanine

Stoichiometry C10H13N12O2(1+)
Framework group C1[X(C10H13N12O2)]
Deg. of freedom 105
Full point group C1 NOp 1

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Ångstroms) X	Y	Z
1	6	0	-5.194368	-0.219603	-0.101662
2	7	0	-4.326756	-1.285710	-0.104282
3	6	0	-2.916710	-1.205681	-0.034956
4	6	0	-2.478046	0.141658	0.027503
5	6	0	-3.443728	1.143922	0.019074
6	7	0	-4.782420	1.030940	-0.033376
7	8	0	-2.270777	-2.268357	-0.041514
8	7	0	-1.197771	0.695061	0.093198
9	7	0	-2.739164	2.324242	0.087080
10	6	0	-1.392666	2.005719	0.134204
11	1	0	-3.149691	3.244692	0.146319
12	1	0	-4.677322	-2.232007	-0.197881
13	7	0	-0.437955	2.954305	0.266257
14	1	0	-0.658696	3.897801	-0.012889
15	1	0	0.535024	2.670762	0.135534
16	6	0	2.499706	-0.103587	0.004215
17	6	0	3.415411	-1.138504	0.026919
18	7	0	2.672879	-2.308178	0.095743
19	6	0	1.337616	-1.990705	0.110269
20	7	0	1.213907	-0.655077	0.056025
21	1	0	3.063163	-3.239255	0.122642
22	6	0	2.963316	1.238419	-0.075715
23	8	0	2.338247	2.298134	-0.114486
24	7	0	4.383450	1.262037	-0.117386
25	1	0	4.768894	2.196129	-0.200741
26	7	0	0.343150	-2.865882	0.177481
27	1	0	-0.653217	-2.559885	0.107822
28	1	0	0.542868	-3.853498	0.177579
29	6	0	5.211491	0.166002	-0.085208
30	7	0	4.753135	-1.073623	-0.007788
31	1	0	0.284016	-0.124487	0.069040
32	7	0	6.544305	0.373788	-0.164564
33	1	0	7.138589	-0.434062	-0.057448
34	1	0	6.955312	1.284458	-0.038544
35	7	0	-6.523267	-0.486188	-0.221713
36	1	0	-7.133085	0.308049	-0.093472
37	1	0	-6.892198	-1.382755	0.055459

Rotational constants (GHZ): 0.5638194 0.1152244 0.0958313
Standard basis: 6-31G(d,p) (6D, 7F)
There are 425 symmetry adapted basis functions of A symmetry.
SCF Done: E(RB+HF-LYP) = -1196.33343039 -V/T = 2.0092
Stoichiometry C10H13N12O2(1+)
Framework group C2[C2(H),X(C10H12N12O2)]
Deg. of freedom 53

Table 5.2. Proton shift TS for Hoogsteen proton-bound homodimer of 8-aminoguanine

Full point group C2 NOp 2
Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Ångstroms) X	Y	Z
1	6	0	-0.152426	5.146374	-0.172773
2	7	0	-1.241504	4.314492	-0.261215
3	6	0	-1.218490	2.900420	-0.173316
4	6	0	0.108142	2.423230	0.001839
5	6	0	1.135363	3.352851	0.079576
6	7	0	1.073999	4.691699	0.009664
7	8	0	-2.290137	2.287297	-0.257879
8	7	0	0.636963	1.129555	0.121132
9	7	0	2.290137	2.618098	0.252186
10	6	0	1.955763	1.282488	0.268782
11	1	0	3.216609	3.008319	0.347308
12	1	0	-2.167195	4.693400	-0.425970
13	7	0	2.855400	0.306478	0.451741
14	1	0	3.835044	0.536225	0.388679
15	1	0	2.595485	-0.664199	0.215887
16	6	0	-0.108142	-2.423230	0.001839
17	6	0	-1.135363	-3.352851	0.079576
18	7	0	-2.290137	-2.618098	0.252186
19	6	0	-1.955763	-1.282488	0.268782
20	7	0	-0.636963	-1.129555	0.121132
21	1	0	-3.216609	-3.008319	0.347308
22	6	0	1.218490	-2.900420	-0.173316
23	8	0	2.290137	-2.287297	-0.257879
24	7	0	1.241504	-4.314492	-0.261215
25	1	0	2.167195	-4.693400	-0.425970
26	7	0	-2.855400	-0.306478	0.451741
27	1	0	-2.595485	0.664199	0.215887
28	1	0	-3.835044	-0.536225	0.388679
29	6	0	0.152426	-5.146374	-0.172773
30	7	0	-1.073999	-4.691699	0.009664
31	1	0	0.000000	0.000000	0.112242
32	7	0	0.356795	-6.478780	-0.317385
33	1	0	-0.440657	-7.071328	-0.141219
34	1	0	1.270303	-6.886189	-0.195273
35	7	0	-0.356795	6.478780	-0.317385
36	1	0	0.440657	7.071328	-0.141219
37	1	0	-1.270303	6.886189	-0.195273

Rotational constants (GHZ): 0.5684964 0.1181191
0.0985323
Standard basis: 6-31G(d,p) (6D, 7F)
There are 213 symmetry adapted basis functions of A symmetry.
There are 212 symmetry adapted basis functions of B symmetry.
SCF Done: E(RB+HF-LYP) = -1196.32944490 -V/T = 2.0092

5.4. 8-Aminoguanine-Rich Oligonucleotides Form pH-Dependent Self-Assemblies, Similar to Cytosine-Rich Oligonucleotides

As an experimental means to characterize the ability of X^+X pairing to drive the self-assembly of oligonucleotides, we prepared and analyzed d(TXXXXT) (TX_4T). Our results support the formation of a structure by oligo(dX) analogous to that reported by Miles for poly(rX). TX_4T in lithium or sodium buffer at pH 4.5 exhibits a cooperative, reversible melt (T_M = 47ºC, Figure 5.2), which is destabilized at pH 7.0 in lithium or sodium buffer. Both of these observations indicate that the TX_4T self-structure at low pH is not a G-quadruplex, because either reducing the population of protonated dX bases or changing associated cations from lithium to sodium would stabilize a G-quadruplex. We note as well that the low-pH TX_4T melting transition is less cooperative than that observed in the tetramolecular sodium G-quadruplex formed by d(TGGGGT) (Figure 5.3), further confirming Miles' proposed bimolecularity. The TX_4T self-structure is, however, of similar, unusually high thermal stability. In comparison to Watson-Crick base pairing, at least six G·C base pairs are required for a comparable T_M at neutral pH with the given salt and oligonucleotide concentration.[16]

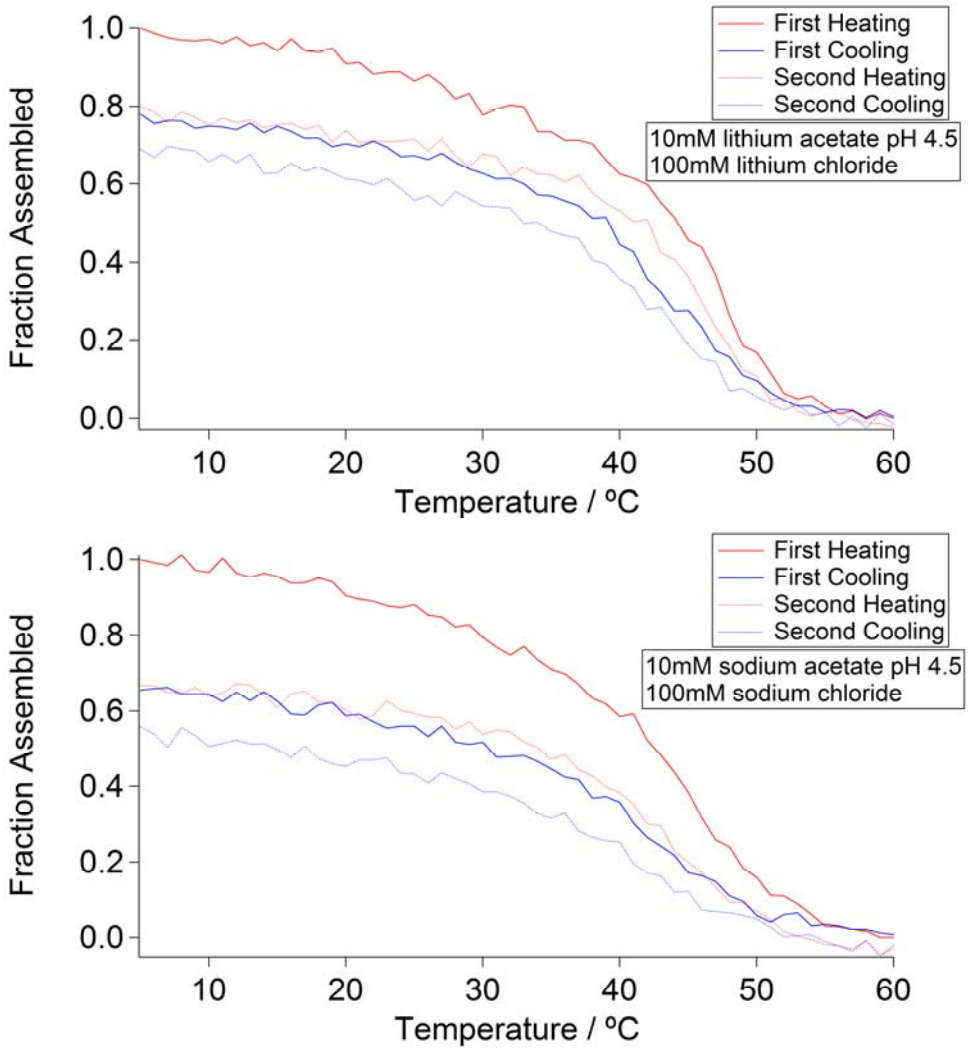


Figure 5.2. Thermal denaturation of TX_4T in lithium and sodium-containing buffers.

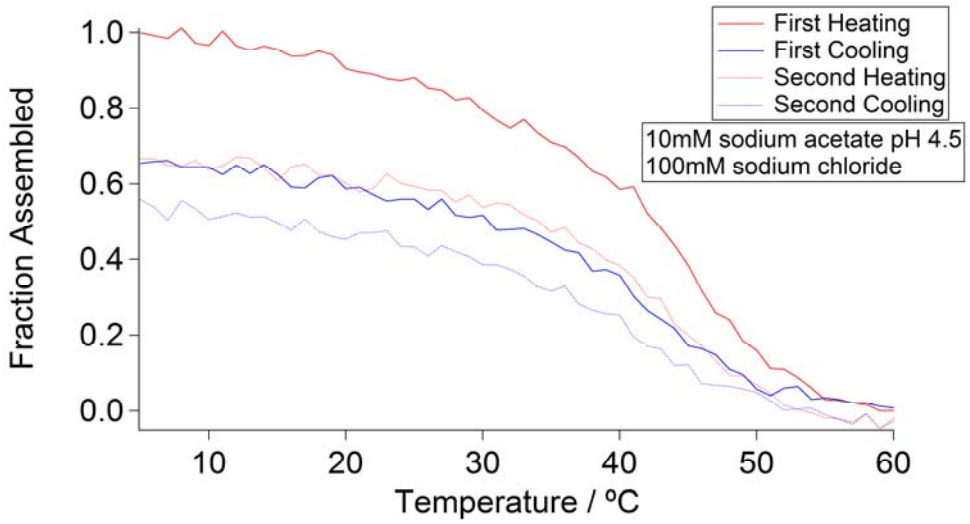


Figure 5.3. Thermal denaturation of TG_4T in sodium-containing buffer. Irreversible changes in the CD spectrum persisted even after overnight incubation at 5ºC, due to the slow, tetramolecular association of the G-quadruplex.

As mentioned above, a downfield ^1H chemical shift has been invoked as evidence of strong hydrogen bonding, with ^1H chemical shifts of $\delta \geq 15$ ppm typically considered indicative of such interactions. While hydrogen-bound imino protons in duplex and quadruplex DNA structures are known to exhibit downfield shifts (δ = 10-14 ppm), TX_4T exhibits somewhat lower-field resonances, with four singlets observed between 14.8 and 15.3 ppm (Figure 5.4). This observation also suggests that the TX_4T self-structure is parallel-stranded (a symmetric, antiparallel structure would be expected to exhibit only two imino resonances). Mergny and coworkers recently reported NMR spectra of quadruplexes formed by all four TG_4T analogues with single dG-to-dX mutations and reported no imino resonances downfield from 12.2 ppm.[17] Further, the introduction of a single dX substitution resulted in significant polymorphism and destabilization of the quadruplex in all but one position. The imino resonances of TX_4T, by contrast, are well-

dispersed. Additionally, only one set of non-exchangeable proton resonances are present (Figure 5.5), again supporting one parallel-stranded structure.

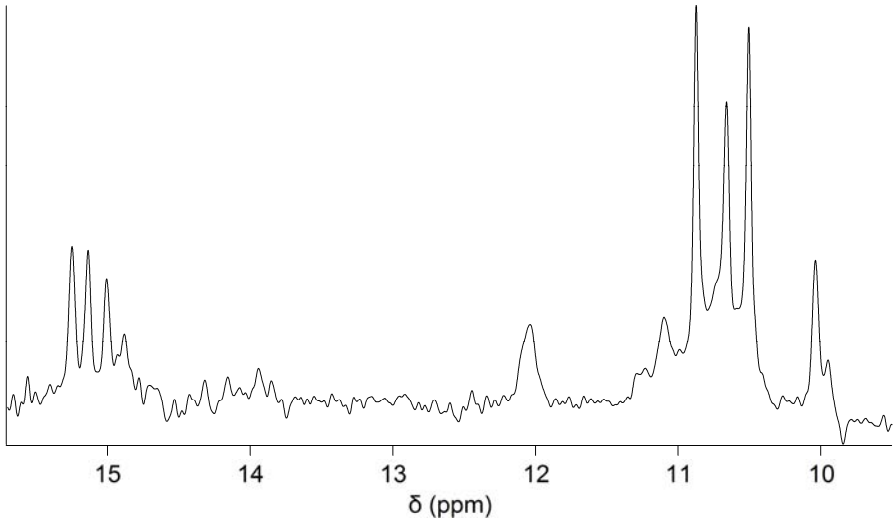


Figure 5.4. 500 MHz ^{1}H NMR spectrum of imino and amino protons of 1 mM TX_4T at 275 K. Sample was lithium-exchanged with no added salt, 10% D_2O, pH 5.

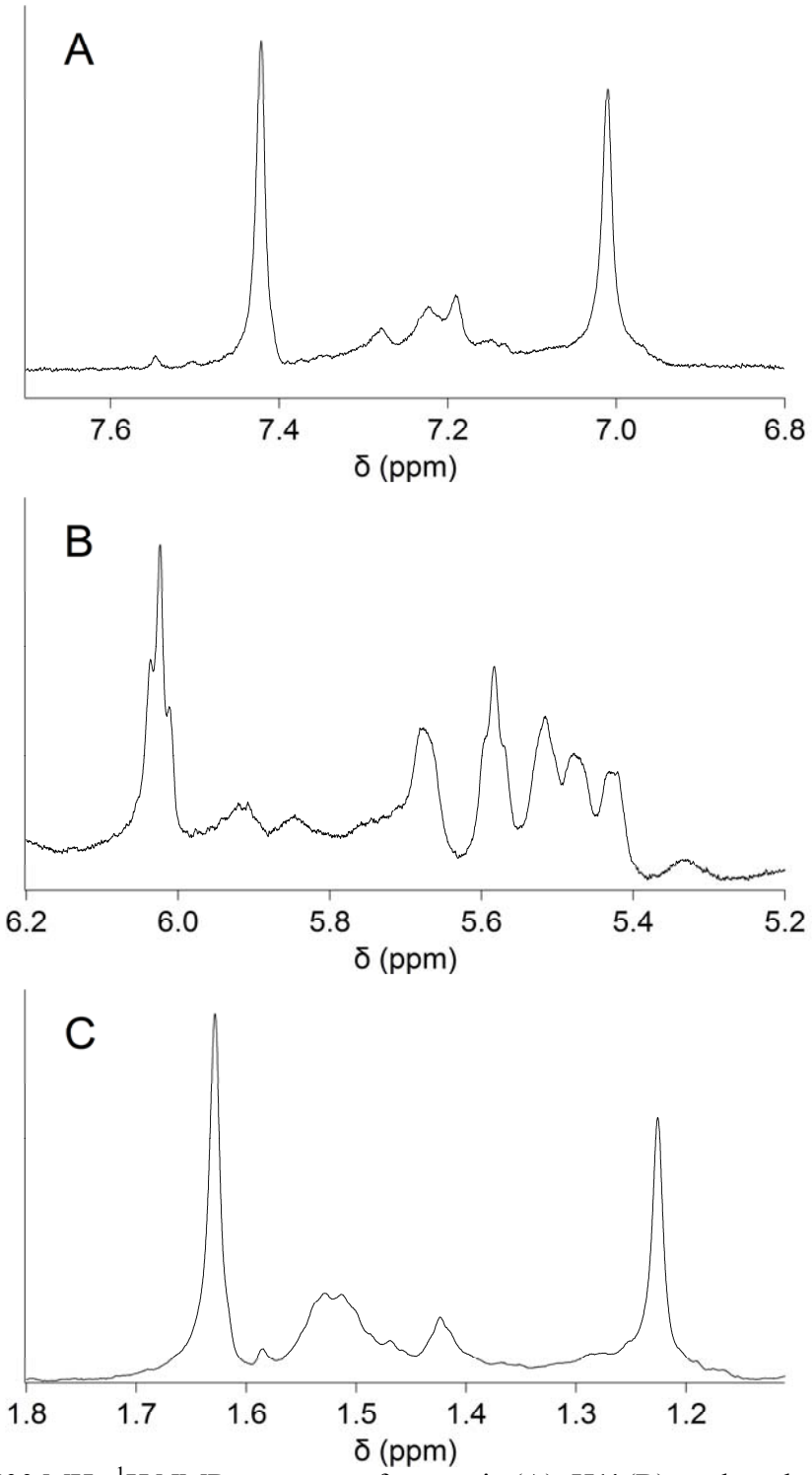


Figure 5.5. 500 MHz ^{1}H NMR spectrum of aromatic (A), H1′ (B), and methyl (C) protons of 1 mM TX_4T at 275 K. Sample was lithium-exchanged with no added salt, 10% D_2O, pH 5.

We note that a previously reported DNA structure formed by a natural nucleotide may also contain LBHBs. Oligonucleotides containing runs of cytosine residues are known to form so-called i-motif structures with C^+C base pairs (Figure 1c). The two residues involved in an i-motif base pair are the same,[18, 19] so the imine nitrogens are pKa matched. NMR spectroscopy studies of i-motif structures also reveal low-field (≥15 ppm) resonances for the imino protons of the hemiprotonated C^+C base pairs.[20] Furthermore, Han and coworkers recently calculated comparable interaction energies, barrier heights, and inter-heavy atom separation for the symmetric C^+C base pair as presented above for the X^+X base pair.[21] They predicted a 3.6-4.6 kcal/mol barrier to proton transit across the analogous N3···H^+···N3 hydrogen bond.

As a means to compare the relative stability of the TX_4T self-structure with the i-motif, we measured the melting temperature of the i-motif formed by TC_4T under the same conditions used in the TX_4T studies discussed above. At pH 4.5, TC_4T exhibited a cation-insensitive, cooperative melting transition at 44ºC. Given the similar stability of the TX_4T and the TC_4T self-structures, it appears the principle of strong hydrogen bonding between hemiprotonated pKa-matched nucleobases may be general.

5.5. Conclusion

Regardless of the exact type of hydrogen bonds formed by duplex TX_4T (with or without LBHBs), it is clear that 8-amino-G, in RNA and DNA, is able to form unique, pH-specific interactions with itself that appear to be associated with pKa matching. Lippert and Schmidt have described other ionic guanine derivatives that exhibit proton-bound dimers even at the single-base level, providing further evidence for the general principle that pKa-matched proton-bound dimers are tightly bound.[22, 23] These

interactions are of unusual strength and present a means of selecting against G-quadruplex formation: TX_4T, when operated at low to neutral pH, forsakes the very stable quadruplex formed by TG_4T. Combined with observations reported previously for i-motif DNA, we have illustrated that pKa-matched proton-bound molecular recognition motifs are a potentially powerful tool by which to drive selective, strong self-assembly. Computational studies also suggest that additional LBHB-stabilized base pairs can be formed by adenine nucleotides that would be compatible with flanking Watson-Crick duplexes.[24]

5.6. References

1. Kreevoy, M.M. and T.M. Liang, *Structures and isotopic fractionation factors of complexes, A1HA2-.* Journal of the American Chemical Society, 1980. **102**(10): p. 3315-3322.

2. Kato, Y., L.M. Toledo, and J. Rebek, *Energetics of a low barrier hydrogen bond in nonpolar solvents.* Journal of the American Chemical Society, 1996. **118**(36): p. 8575-8579.

3. GarciaViloca, M., A. GonzalezLafont, and J.M. Lluch, *Theoretical study of the low-barrier hydrogen bond in the hydrogen maleate anion in the gas phase. Comparison with normal hydrogen bonds.* Journal of the American Chemical Society, 1997. **119**(5): p. 1081-1086.

4. Kumar, G.A. and M.A. McAllister, *Characterization of low-barrier hydrogen bonds. 8. Substituent effects on the strength and geometry of the formic acid formate anion model system. An ab initio and DFT investigation.* Journal of the American Chemical Society, 1998. **120**(13): p. 3159-3165.

5. Alkorta, I. and J. Elguero, *Basicity and proton transfer in proton sponges and related compounds: An ab initio study.* Struct. Chem., 2000. **11**(6): p. 335-340.

6. Zhu, H.F., et al., *Supramolecular architectures constructed by strong hydrogen bonds. Crystal structures of novel one-dimensional polycatenane and three-dimensional interpenetrated network.* Chemistry Letters, 2002(9): p. 898-899.

7. Hudson, B.S. and N. Verdal, *Vibrational dynamics in short, strong symmetric hydrogen bonds: General considerations and two examples.* Physica B-Condensed Matter, 2006. **385**: p. 212-215.

8. Frey, P.A., *Isotope Effects in the Characterization of Low Barrier Hydrogen Bonds*, in *Isotope Effects in Chemistry and Biology*, A. Kohen and H.-H. Limbach, Editors. 2006, CRC Press: Boca Raton. p. 975-993.

9. Lau, J.S. and C.L. Perrin, *Isotope effects and symmetry of hydrogen bonds in solution: single- and double well potential*, in *Isotope Effects in Chemistry and Biology*, A. Kohen and H.-H. Limbach, Editors. 2006, CRC Press: Boca Raton. p. 231-252.

10. Bratoz, S. and D. Hadzi, *Infrared spectra of molecules with hydrogen bonds.* Journal of Chemical Physics, 1957. **27**(5): p. 991-997.

11. Speakman, J.C., *The crystal structure of, and the hydrogen bond in, sodium hydrogen diacetate.* Proc. Chem. Soc. London, 1959(10): p. 316-316.

12. Yaghmaei, S., et al., *Chelation of a proton by an aliphatic tertiary diamine.* Journal of the American Chemical Society, 2008. **130**(25): p. 7836-7838.

13. Mittapalli, G.K., et al., *Mapping the landscape of potentially primordial informational oligomers: oligodipeptides tagged with 2,4-disubstituted 5-aminopyrimidines as recognition elements.* Angewandte Chemie-International Edition in English, 2007. **46**(14): p. 2478-2484.

14. Hattori, M., J. Frazier, and H.T. Miles, *Ordered forms of 5'-8-aminoguanylic acid.* Biopolymers, 1975. **14**(10): p. 2095-2106.

15. Hattori, M., J. Frazier, and H.T. Miles, *Poly(8-aminoguanylic acid) - formation of ordered self-structures and interaction with poly(cytidylic acid).* Biochemistry, 1975. **14**(23): p. 5033-5045.

16. SantaLucia, J. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**: p. 415-440.

17. Gros, J., et al., *8-Amino guanine accelerates tetramolecular G-quadruplex formation.* Chem Commun, 2008(25): p. 2926-2928.

18. Gehring, K., J.L. Leroy, and M. Gueron, *A tetrameric DNA-structure with protonated cytosine.cytosine base-pairs.* Nature, 1993. **363**(6429): p. 561-565.

19. Sponer, J., et al., *Base stacking and hydrogen bonding in protonated cytosine dimer: The role of molecular ion-dipole and induction interactions.* Journal of Biomolecular Structure & Dynamics, 1996. **13**(4): p. 695-706.

20. Esmaili, N. and J.L. Leroy, *i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats.* Nucleic Acids Research, 2005. **33**(1): p. 213-224.

21. Han, S.Y. and H.B. Oh, *Theoretical study of the ionic hydrogen bond in the isolated proton-bound dimer of cytosine.* Chemical Physics Letters, 2006. **432**(1-3): p. 269-274.

22. Schmidt, A., N. Kobakhidze, and M.K. Kindermann, *Molecular recognition of modified nucleobases. Self-complementarity and base-pairing of betainic guanine model compounds.* J. Chem. Soc., Perkin Trans. 1, 2002(7): p. 982-990.

23. Roitzsch, M. and B. Lippert, *Structural precursor of the hemideprotonated guanine pair.* Chemical Communications, 2005(48): p. 5991-5993.

24. Hud, N.V. and T.H. Morton, *DFT energy surfaces for aminopurine homodimers and their conjugate acid ions.* Journal of Physical Chemistry A, 2007. **111**(17): p. 3369-3377.

CHAPTER 6

DNA AND RNA IN ANHYDROUS MEDIA: EVIDENCE FOR DUPLEX, TRIPLEX, AND G-QUADRUPLEX SECONDARY STRUCTURES IN A DEEP EUTECTIC SOLVENT[e]

6.1. Introduction

Room-temperature ionic liquids (RTIL) have generated tremendous interest as nonvolatile media that provide favorable solution-state environments for a wide range of chemical reactions.[1-6] Recently, Abbott *et al.* developed room temperature deep eutectic solvents (DES), a closely related class of solvents with physical properties and phase behaviors very similar to RTIL.[7, 8] These eutectic mixtures are attractive alternatives to RTIL, as DES can be less expensive, more synthetically accessible, nontoxic and biodegradable. Here, we report that nucleic acids can form several secondary structures that reversibly denature with heating in a water-free DES. In some cases, the nucleic acid sequences studied exhibit different relative stabilities and different secondary structures in the DES, compared to aqueous media.

The melting point of choline chloride (ChCl) is 302 ˚C and that of urea is 133 ˚C, while a 1:2 ChCl:urea mixture melts at 12 ˚C.[7] This DES dissolves organic ions, inorganic salts that are only sparingly soluble in water, as well as metal oxides (e.g.,

---

[e] This chapter was adapted from the pre-peer reviewed version of previously published work. (Mamajanov, I; Engelhart, AE; Bean, HD; Hud, NV. "DNA and RNA in anhydrous media: duplex, triplex, and G-quadruplex secondary structures in a deep eutectic solvent." Angew. Chem. Intl. Ed. Engl. (2010) DOI: 10.1002/anie.201001561.). It is reproduced with permission.

CuO) – properties that are more similar to those of ionic liquids than to molecular solvents.[7, 8] These phase and solvent properties make this DES an attractive solvent of this class in which to explore nucleic acid secondary structure.

6.2. Experimental Procedures

6.2.1. Materials

The DNA strands were purchased from Integrated DNA Technologies (Coralville, IA) and Dharmacon (Lafayette, CO), respectively. The sequences used in the study are listed in Table 6.1.

Table 6.1. Nucleic acids studied in choline chloride:urea DES.

Name	Sequence
32-bp DNA Duplex	d(GGT GTC AGT AAG CCA TTC GAG ATC CTC ATA GT) and d(ACT ATG AGG ATC TCG AAT GGC TTA CTG ACA CC)
$d(AT)_{16}$	d(ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT AT)
$d(A_4T_4)_4$	d(AAA ATT TTA AAA TTT TAA AAT TTT AAA ATT TT)
$d(CG)_8$	d(CGC GCG CGC GCG CGC G)
dA_{16}	d(AAA AAA AAA AAA AAA A)
dT_{16}	d(TTT TTT TTT TTT TTT T)
TBA	d(GGT TGG TGT GGT TGG)
12-bp RNA Duplex	r(CGG CGC GGC GGG) and r(CCC GCC GCG CCG)

The choline chloride/urea DES was prepared by heating a 1:2 molar mixture of choline chloride (Acros Organics) and urea (Fisher) anhydrous solids at 100 °C until a liquid was formed (ca. 2 hours). Water content, measured employing the Hydranal® moisture test kit (Sigma-Aldrich) based on Karl-Fischer analysis, was established to be 0.23%. The RTIL 1-methylimidazolium tetrafluoroborate (HMIm BF_4) was synthesized

as described previously.[9] Solutions of nucleic acids in the DES and RTIL were prepared by mixing an aqueous stock solution of DNA or RNA with the DES and then subjecting the mixture to vacuum centrifugation until a constant mass was reached (at least 12 h), a process that was determined (based on gravimetric analysis) to quantitatively remove added water from the anhydrous DES and RTIL.

6.2.2. CD Spectroscopy

CD spectra were acquired on a Jasco J-810 spectropolarimeter equipped with Peltier temperature controller. 1 mm pathlength cells were used for DES and aqueous samples (1.6 m*m* nucleotide concentration). The range of 235-350 nm was scanned at a rate of 500 nm/min, averaged over 5 measurements in case of the DES and water. For thermal denaturation studies, measurements were taken over two heating/cooling cycles from 5 – 95 °C at increments of 5 °C. For of spectra acquired in the RTIL, 16 mm nucleotide concentrations and 0.1 mm pathlength detachable cells were utilized to minimize interference due to HMIm absorption. The use of detachable cells precluded long experiments; therefore, thermal denaturation studies were not conducted.

Melting curves were generated by plotting the elipticity at a single wavelength versus temperature. The T_m was derived from a sigmoidal fit about the transition.

6.3. Deep Eutectic Solvents Solvate a Variety of Nucleic Acid Duplex Structures

Circular dichroism (CD) was used to monitor nucleic acid structure.[10] A 32-bp DNA duplex of mixed GC/AT sequence composition in a low-salt buffer (100 m*m* NaCl, 10 m*m* sodium phosphate, pH 7) exhibits a CD spectrum consistent with a B-form duplex, as expected (Figure 6.1a). In the DES, the positive band is both significantly more intense and blue-shifted (Figure 6.1a), indicating a change in secondary structure (see below). In this buffer, the duplex exhibits a cooperative, reversible melting transition with a midpoint (T_m) at 73 °C. In the DES, the cooperative, reversible transition is retained (Figure 6.2), but the T_m is depressed to 37 °C.

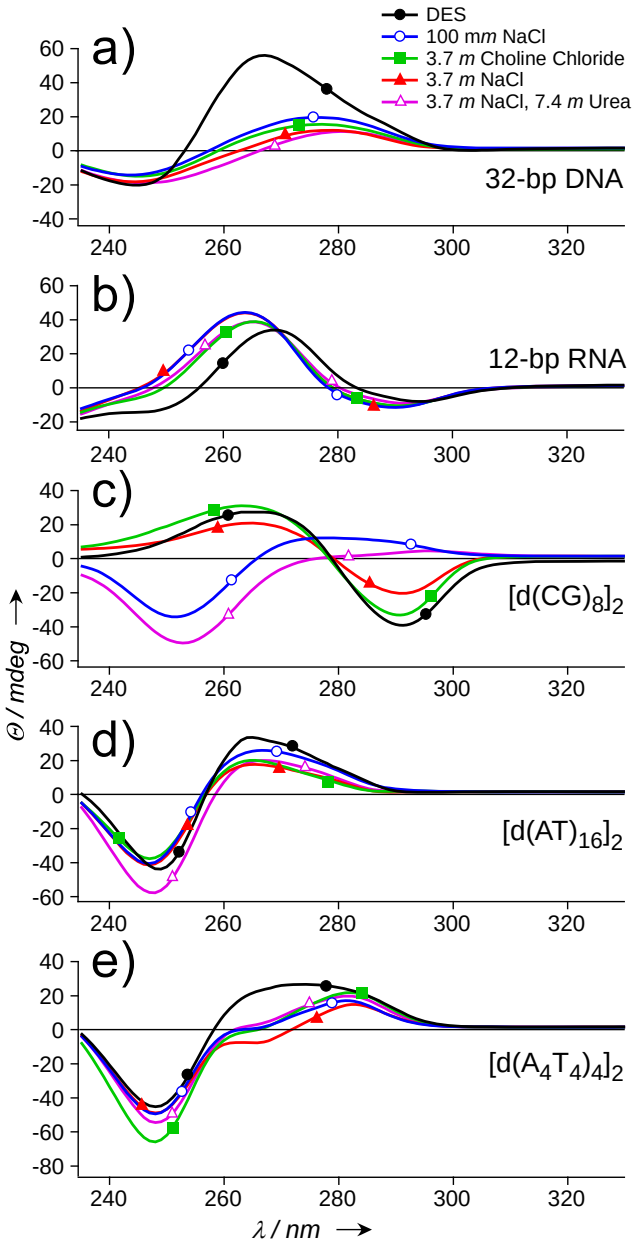


Figure 6.1. CD spectra of DNA and RNA duplexes in the DES. The concentration of all samples was 1.6 m*m* in nucleotide. Spectra were collected in 1 mm path length cells. All aqueous solutions contained 10 m*m* sodium phosphate, pH 7. Spectra were acquired at 5°C.

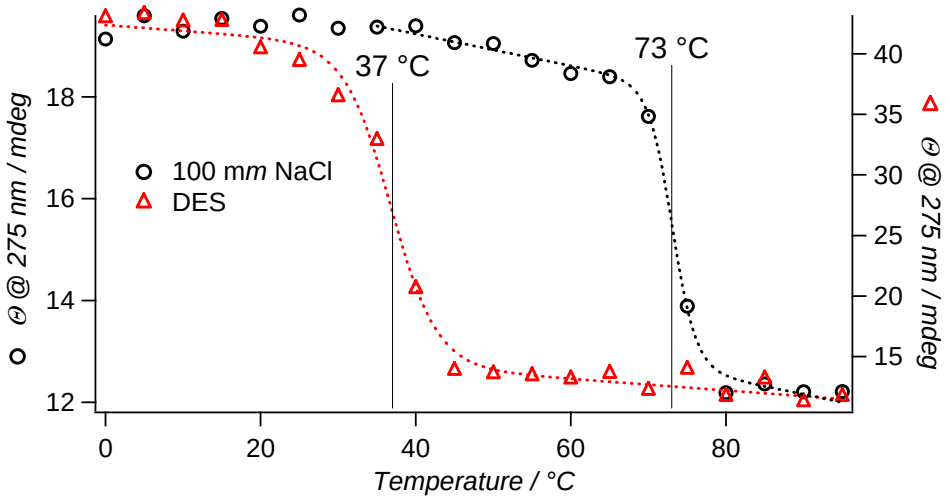


Figure 6.2. Melting curves of 32-bp DNA duplex in the DES and in aqueous solution. The aqueous solution contained 100 mM NaCl, 10 mM Na phosphate buffer, pH=7.

The CD spectrum of the 32-bp DNA duplex in the DES is indicative of an A-form helix.[10] Most mixed-sequence DNA duplexes will undergo a B- to A-form helix transition when subjected to dehydrating conditions and high ionic strength,[11-14] which is certainly descriptive of the DES. Consistent with these results, an RNA duplex (which adopts the A-form helix, even in low-salt aqueous solution) exhibits similar CD spectra in the DES and in varied aqueous buffers, although the spectrum in the DES is red-shifted by 5 nm (Figure 6.1b).

It is perhaps not surprising that duplex stability is depressed in the DES relative to aqueous solution (Table 6.2), as urea is commonly used as a denaturant, and the DES contains 7.4 *m* urea and 3.7 *m* ChCl. In an aqueous solution containing 7.4 *m* urea and 3.7 *m* NaCl, the T_m of the 32-bp DNA duplex is 65 °C, both T_ms being significantly higher than the T_m observed in the DES. In 3.7 *m* aqueous ChCl, the T_m is 83 ˚C (Table 6.2). Thus, the depressed T_m in the DES cannot be entirely attributed to either solvent

component. It is the result of distinct solvent properties of the DES, the elimination of bulk water, or both. The T_m values observed for the RNA duplex support this conclusion (Table 6.2).

Table 6.2. T_m (°C) of DNA and RNA duplexes in various solvents. Aqueous solutions contained 10 m*m* sodium phosphate, pH 7. $d(CG)_8$ T_M was not determined, due to a broad melting transition. 32-bp DNA and 12-bp RNA sequences are described in Table 6.1.

Nucleic Acid	100 m*m* NaCl	DES	3.7 *m* NaCl	3.7 *m* ChCl	3.7 *m* NaCl 7.4 *m* Urea
32-bp DNA	73	37	85	83	65
12-bp RNA	≥95	44	>95	69	83
$d(CG)_8$	≥85	44	nd	72	76
$d(AT)_{16}$	57	29	68	77	38
$d(A_4T_4)_4$	57	26	73	79	52

Given the apparent preference of the mixed GC/AT DNA duplex for the A-form helix in the DES, we considered whether duplexes with CG repeats, which can exhibit a B- to Z-form transition in high salt or dehydrating conditions, might also exhibit an alternative structure in the DES. For this investigation, the oligonucleotide $d(CG)_8$ was selected. The CD spectrum of $d(CG)_8$ in the DES is dramatically different from that obtained in low-salt buffer (Figure 6.1c). The inverted CD spectrum obtained in DES, with a negative band at 290 nm and a positive band at 260 nm, is indicative of a left-handed Z-form helix.[15, 16] Similar spectra are observed in aqueous solutions containing 3.7 *m* ChCl or 3.7 *m* NaCl, but not 3.7 *m* NaCl and 7.4 *m* urea (Figure 6.1c). The spectra in 3.7 *m* ChCl and the DES are particularly similar, strongly suggesting that

the helical structures of $d(CG)_8$ in these two environments, one aqueous and one anhydrous, are the same.

DNA polymers with AT repeats can also adopt helical structures that deviate from the canonical B-form, including cation-dependent variants of the B- and A-form helices in solution and in fibers,[17-19] and the Z-form at high ionic strength and elevated temperature.[20] We therefore examined the oligonucleotide $d(AT)_{16}$. Unlike the mixed-sequence 32-bp DNA or $d(CG)_8$, $d(AT)_{16}$ does not exhibit an appreciably different CD spectrum in the DES (Figure 6.1d), suggesting the helical structure is similar to that in the aqueous solutions.

The duplex formed by the sequence $d(A_4T_4)_4$ was also of interest, because it contains four A-tract sequence elements, which are defined as four or more A·T base pairs without a TpA step. These sequence elements are known to adopt an altered B-form helical structure, designated B*, with an unusually narrow minor groove and high base pair propeller twist,[21-24] and a propensity for cation localization in the minor groove.[25-28] For $d(A_4T_4)_4$, any B*-form helical structure is predicted to be interspersed with B-form helical structure, as the B*-form is disrupted by TpA steps.[23] The CD spectrum of $d(A_4T_4)_4$ in low-salt buffer at 5 °C is consistent with a mixed B-/B*-form structure (Figure 1e), and spectra of the same general shape are observed for other aqueous buffer conditions. In the DES, however, $d(A_4T_4)_4$ exhibits a significantly different CD spectrum (Figure 6.1e). Although the origin of this spectral difference is not obvious, it may represent a significant change in secondary structure. Nevertheless, a cooperative transition at 26 °C indicates that $d(A_4T_4)_4$ does form a secondary structure in the DES (Figure 6.3).

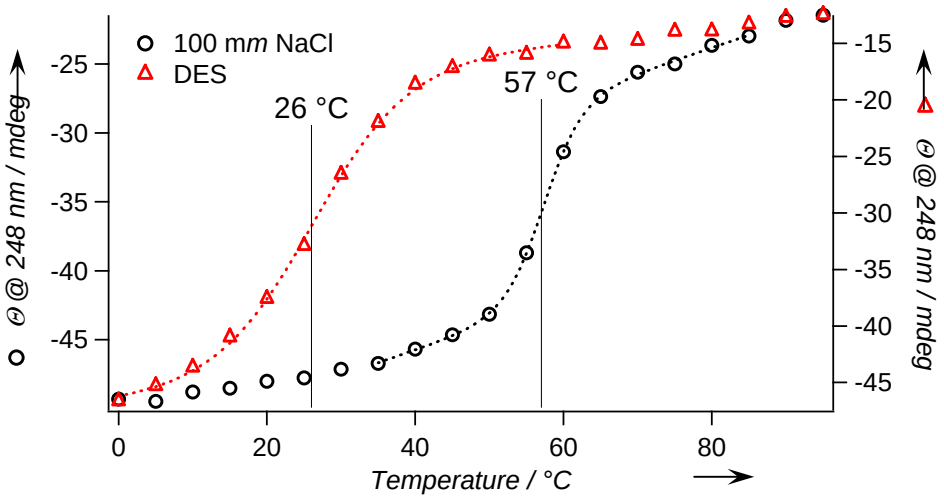


Figure 6.3. Melting curves of $d(A_4T_4)_4$ strand in the DES and in aqueous solution. The aqueous solution contained 100 mM NaCl, 10 mM Na phosphate buffer, pH=7.

6.4. Deep Eutectic Solvents Solvate Non-Duplex Nucleic Acid Structures

Nucleic acids can also form triplex structures, in which a homo-pyrimidine strand forms Hoogsteen base pairs with the purines of a homo-purine·homo-pyrimidine Watson-Crick duplex.[29, 30] To investigate potential triplex formation within the DES, we prepared samples of $d(A)_{16}$ and $d(T)_{16}$ in a 1:2 molar ratio. Triplexes typically require divalent cations or high monovalent cation concentrations to be stable at room temperature.[31] A 1:2 mixture of $d(A)_{16}$ and $d(T)_{16}$ in buffer with 50 m*m* $MgCl_2$ exhibits a CD spectrum consistent with triplex formation (Figure 2a).[32] This system exhibits a third strand T_m of 20 °C, and a $d(A)_{16}$·$d(T)_{16}$ duplex T_m of 44 °C (Figure 6.4). In the DES, the CD spectrum of this system is significantly different (Figure 6.1). Furthermore, only one melting transition is observed at 66 °C (Figure 6.4). At first, we thought these observations were an indication that the triplex was not stable in the DES. However, the

addition of the spectra of samples containing 1:1 $d(A)_{16}{\cdot}d(T)_{16}$ and only $d(T)_{16}$ in the DES, at concentrations that would occur if the triplex was completely unstable in the DES at 15 °C (i.e., no interaction between the duplex and third homopyrimidine strand), yielded a spectrum that differed from that of the putative triplex in the DES (Figure 6.5). Thus, the additional molar equivalent of the $d(T)_{16}$ strand appears to interact with the $d(A)_{16}{\cdot}d(T)_{16}$ duplex.

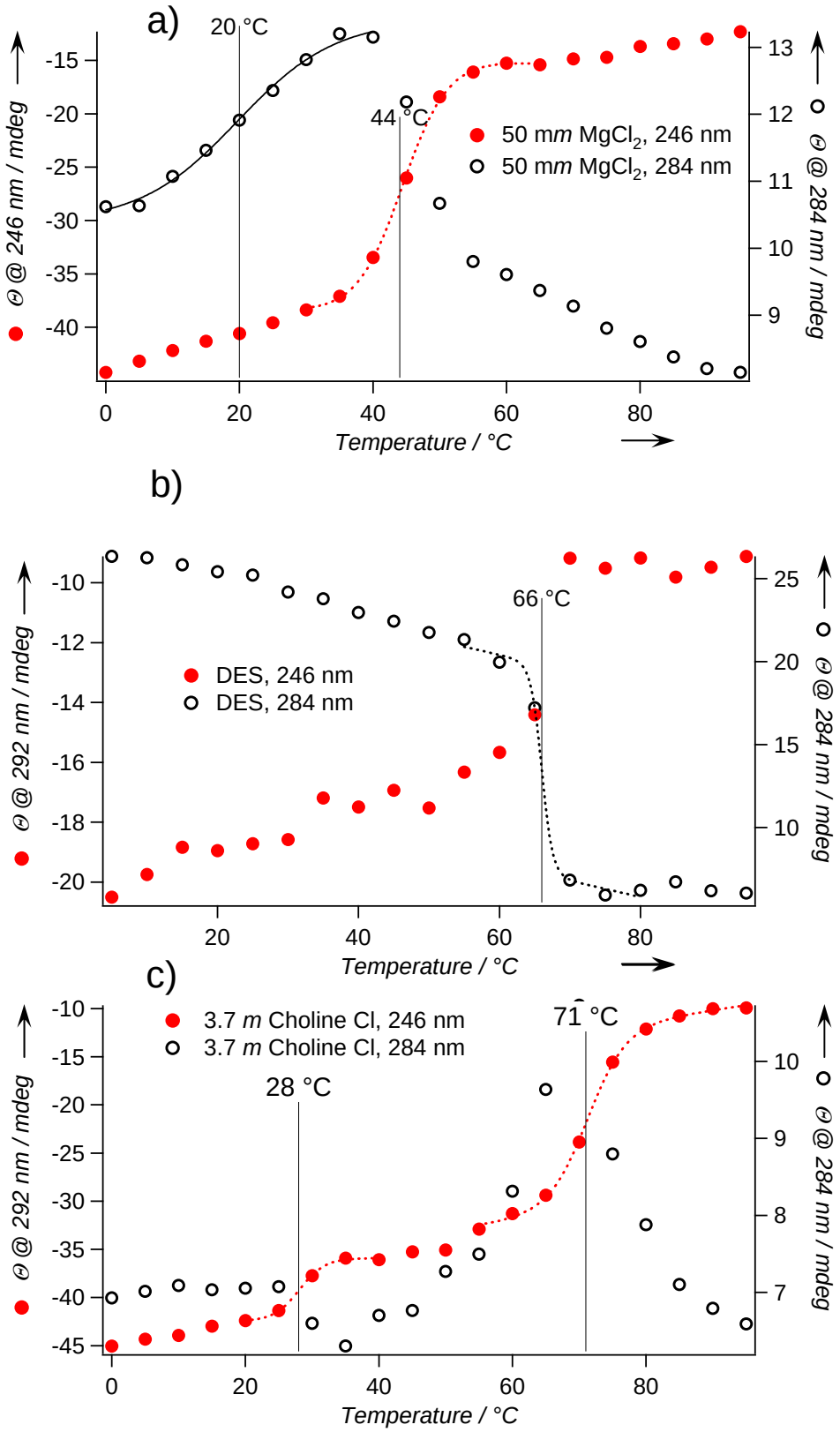


Figure 6.4. Melting curves of $d(A)_{16}$ and $d(T)_{16}$ in a 1:2 molar ratio in various solvents. a) aqueous solution (50 mM $MgCl_2$, 10 mM Na phosphate buffer, pH=7) b) DES , and c) aqueous choline chloride (3.7M) solution.

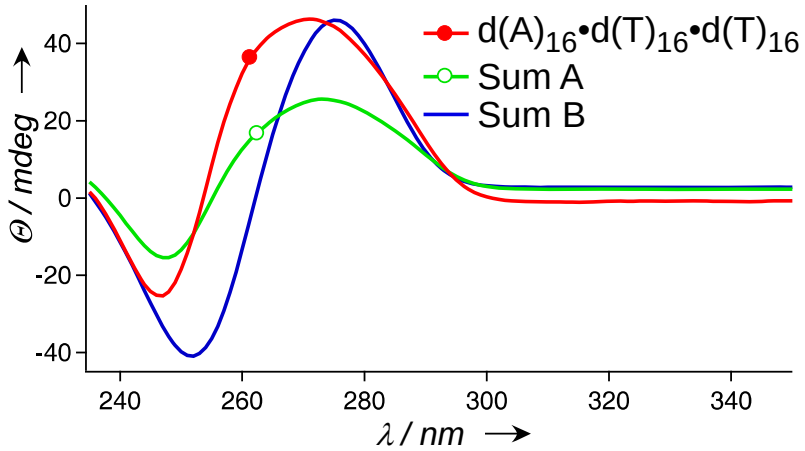


Figure 6.5. The CD spectrum of $d(A)_{16}$-2 $d(T)_{16}$ in the DES is not recreated by addition of the $d(A)_{16}$-$d(T)_{16}$ and $d(T)_{16}$ spectra. CD spectrum of $d(A)_{16}$ and $d(T)_{16}$ (molar ratio 1:2, 1.6m*m* total nucleoside concentration, 15˚C) in DES (red); Sum A is a numerical sum of spectra of [0.533m*m* $d(A)_{16}$] and [1.06m*m* of $d(T)_{16}$] in DES at 15˚C (blue). Sum B is a numerical sum of spectra of [0.533m*m* $d(A)_{16}$ and $d(T)_{16}$] and [0.533m*m* $d(T)_{16}$] at 15˚C (green).

As an additional means to verify triplex formation in the DES, a Job plot was used to investigate the stoichiometry of $d(T)_{16}$-$d(A)_{16}$ interaction (Figure 6.6). The inflection observed in this plot at x=0.67 supports triplex formation (i.e., a 1 $d(A)_{16}$:2 $d(T)_{16}$ complex). Surprisingly, but consistent with the observed single melting transition, the Job plot analysis did not indicate the formation of a 1:1 complex between $d(A)_{16}$ and $d(T)_{16}$ in the DES. These results suggest that the $d(A)_{16}$·$[d(T)_{16}]_2$ triplex is of equal or greater stability than the $d(A)_{16}$·$d(T)_{16}$ duplex, a situation that has been observed previously under high salt conditions.[31]

In a buffer containing 3.7 *m* ChCl, the 1 $d(A)_{16}$:2$d(T)_{16}$ mixture exhibits a spectrum very similar to a buffer containing 50 m*m* $MgCl_2$ (Figure 6.6). The T_m for the triplex in 3.7 *m* ChCl is 28 °C, and the duplex is 71 °C (Figure 6.4). These results illustrate that aqueous choline, at high concentration, stabilizes a DNA triplex and a homopurine-homopyrimidine duplex.

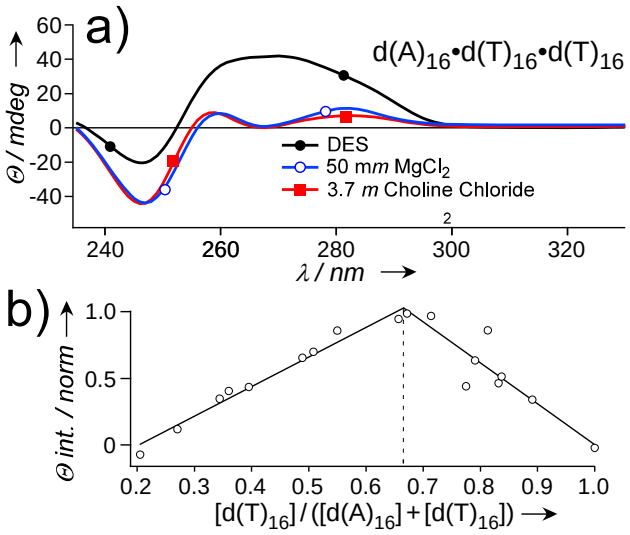


Figure 6.6. Analysis of the stoichiometry of the complex formed between $d(A)_{16}$ and $d(T)_{16}$. a) CD spectra of $d(A)_{16}$ and $d(T)_{16}$ in a 1:2 molar ratio, with a total nucleotide concentration of 1.6 m*m*. Aqueous buffers \contained 10 m*m* sodium phosphate, pH 7. Spectra collected at 5°C. b) Job plot analysis at 5 °C of $d(T)_{16}$ and $d(A)_{16}$ mixtures with a constant nucleotide concentration of 1.6 m*m*. Y-axis is integrated intensity of CD signal from 235-300 nm. Left line is best fit for x=0.2 to 0.67; right line is best fit for x=0.67 to 1.0.

G-quadruplex structures are currently of substantial interest in medicine, self-assembly and nanotechnology.[33, 34] In the present study, we used the thrombin-binding aptamer (TBA)[35] to test the stability of the G-quadruplex secondary structure within the DES. TBA, like other G-quadruplexes, is stabilized by the coordination of cations (e.g., Na^+ or K^+) in the center of G-tetrads. As expected, TBA does not form a stable G-quadruplex in the DES, as the only cation present in significant concentration, choline, is too large for coordination by G-tetrads. However, the addition of 100 m*m* KCl to the DES gave a spectrum similar to that observed for TBA in aqueous buffer containing 100 m*m* KCl (Figure 6.7). As in aqueous buffer, TBA exhibits a cooperative, reversible thermal transition in the DES with 100 m*m* KCl (Figure 6.8).

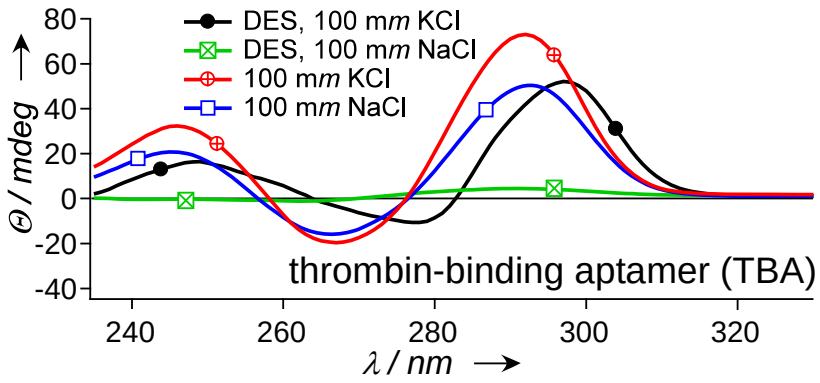


Figure 6.7. CD spectra of the thrombin binding aptamer (TBA) in various solvents. Samples were 1.6 m*m* in nucleotide. Aqueous solutions contained 10 m*m* sodium phosphate, pH 7. Spectra were recorded at 5°C.

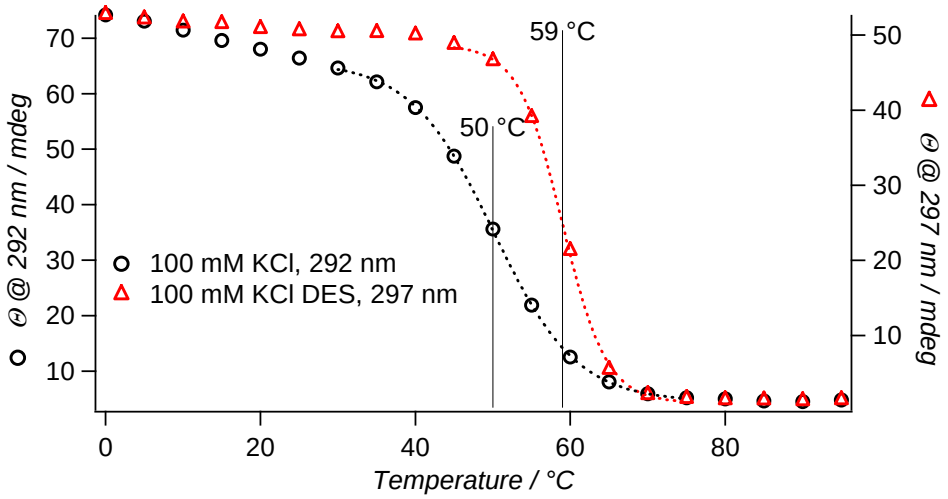


Figure 6.8. Melting curves of TBA in DES containing 100 m*m* KCl and in water containing 100 mM KCl and 10 mM Na phosphate buffer (pH=7)

The stability of the TBA G-quadruplex in the DES with added cations is quite different from that in aqueous buffer. With 100m*m* KCl present, the T_m is 59 °C in DES, compared to 50 °C in aqueous buffer. In contrast, the T_m of TBA in aqueous buffer is 20 °C if 100m*m* NaCl is present, while TBA does not form a stable G-quadruplex in the DES with 100m*m* NaCl (Figure 6.7). Thus, the T_m of the K^+ form of TBA *increases* by 9 °C in the DES compared to buffer, and the T_m of the Na^+ form of TBA *decreases* by at least 20 °C. Given that G-quadruplex stability is strongly coupled to cation solvation free energy,[36] the greater stability difference between the Na^+ and K^+ forms of TBA in the DES could result from differing solvation free energies of Na^+ and K^+ in these two solvents.

6.5. Room-Temperature Ionic Liquids Also Solvate Nucleic Acid Duplex Structures

Finally, as an initial investigation of nucleic acid duplex stability in a RTIL, we examined the 32-bp DNA duplex and $d(CG)_8$ in a popular ionic liquid, N-methylimidazolium tetrafluoroborate (HMIm BF_4). In HMIm BF_4, $d(CG)_8$ exhibits secondary structure formation, while the 32-bp DNA duplex is at least partially denatured – despite these duplexes' comparable stability in the DES (Figure 6.9).

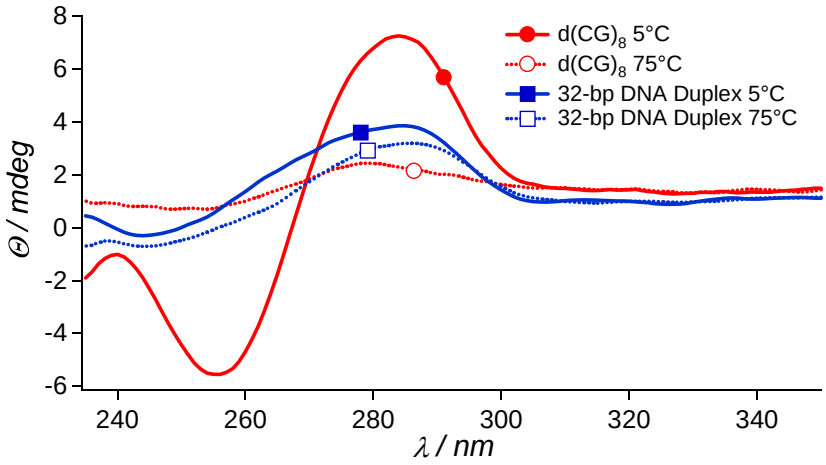


Figure 6.9. CD spectra of 32-bp DNA and $d(CG)_8$ in HMIm BF_4 at 5 and 75°C.

Neither duplex exhibits the B- to A- or Z- transition apparent in the DES. Like molecular solvents, specific DES and RTIL appear to vary in their relationship to the secondary structure of the (bio)polymers they dissolve. The structural similarity of choline to betaine and tetraalkyl ammonium ions may be responsible for some of the DES's characteristics in this regard, as these ions have been reported to exhibit sequence-dependent modulation of DNA duplex stability.[37-39]

6.6. Conclusion

There have been earlier reports of DNA maintaining a duplex structure in nonaqueous solvent, but these reports are apparently limited to glycerol and ethylene glycol.[40-43] Previous studies of DNA structural integrity in RTIL were limited to hydrated ionic liquids.[6] To the best of our knowledge, this is the first demonstration that no fewer than four distinct nucleic acid structures can exist in DES or RTIL. Given previous reports that these solvents can support protein enzyme catalysis, the possibility

that catalytic nucleic acids and protein enzyme-nucleic acid complexes could be used in these solvents is feasible and enticing. Furthermore, in addition to their ability to support biomolecular structures, these solvents' nonvolatility, low water activity, and simple molecular structures makes them attractive and potentially prebiotically plausible milieus for the chemical polymerization of nucleic acids and peptides.

6.7. References

1. Welton, T., *Room-temperature ionic liquids. Solvents for synthesis and catalysis.* Chemical Reviews, 1999. **99**(8): p. 2071-2083.

2. Huddleston, J.G., et al., *Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazolium cation.* Green Chemistry, 2001. **3**(4): p. 156-164.

3. Sheldon, R., *Catalytic reactions in ionic liquids.* Chemical Communications, 2001(23): p. 2399-2407.

4. Rogers, R.D. and K.R. Seddon, *Ionic liquids - Solvents of the future?* Science, 2003. **302**(5646): p. 792-793.

5. Welton, T., *Ionic liquids in catalysis.* Coordination Chemistry Reviews, 2004. **248**(21-24): p. 2459-2477.

6. Vijayaraghavan, R., et al., *Long-Term Structural and Chemical Stability of DNA in Hydrated Ionic Liquids.* Angewandte Chemie International Edition, 2010. **49**(9): p. 1631-1633.

7. Abbott, A.P., et al., *Novel solvent properties of choline chloride/urea mixtures.* Chemical Communications, 2003(1): p. 70-71.

8. Abbott, A.P., et al., *Deep eutectic solvents formed between choline chloride and carboxylic acids: Versatile alternatives to ionic liquids.* Journal of the American Chemical Society, 2004. **126**(29): p. 9142-9147.

9. Zhu, H.P., et al., *Bronsted acidic ionic liquid 1-methylimidazolium tetrafluoroborate: a green catalyst and recyclable medium for esterification.* Green Chemistry, 2003. **5**(1): p. 38-39.

10. Ardhammar, M.N., B., Kurucsev, T., *Circular Dichroism: Principles and Applications*. Second ed, ed. N. Berova, Nakanishi, K., Woody, R.W. 2000: John Wiley & Sons, Inc. 741-768.

11. Franklin, R.E. and R.G. Gosling, *Molecular Configuration in Sodium Thymonucleate.* Nature, 1953. **171**(4356): p. 740-741.

12. Basham, B., G.P. Schroth, and P.S. Ho, *An A-DNA triplet code: Thermodynamic rules for predicting A- and B-DNA.* Proc. Natl. Acad. Sci. USA, 1995. **92**: p. 6464-6468.

13. Tolstorukov, M.Y., et al., *Sequence-dependent B <-> A transition in DNA evaluated with dimeric and trimeric scales.* Biophysical Journal, 2001. **81**(6): p. 3409-3421.

14. Hud, N.V. and J. Plavec, *A unified model for the origin of DNA sequence-directed curvature.* Biopolymers, 2003. **69**(1): p. 144-159.

15. Pohl, F.M. and T.M. Jovin, *Salt-induced cooperative conformational change of a synthetic DNA - Equilibrium and kinetic studies with poly(dG-dC).* Journal of Molecular Biology, 1972. **67**(3): p. 375-396.

16. Wang, A.H.J., et al., *Molecular-structure of a left-handed double helical DNA fragment at atomic resolution.* Nature, 1979. **282**(5740): p. 680-686.

17. Brahms, S., J. Brahms, and K.E. Vanholde, *Nature of conformational-changes in poly[d(A-T)-d(A-T)] in premelting region.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(10): p. 3453-3457.

18. Millane, R.P., et al., *Structure of a pleiomeric form of poly d(AT)-poly d(AT).* Nucleic Acids Research, 1984. **12**(13): p. 5475-5493.

19. Thomas, G.J. and J.M. Benevides, *An A-helix structure for poly(dA-dT) poly(dA-dT).* Biopolymers, 1985. **24**(6): p. 1101-1105.

20. Klump, H.H., E. Schmid, and M. Wosgien, *Energetics of Z-dna formation in poly-d(A-T), poly-d(G-C), and poly-d(A-C) poly-d(G-T).* Nucleic Acids Research, 1993. **21**(10): p. 2343-2348.

21. Dickerson, R.E. and H.R. Drew, *Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure.* J. Mol. Biol., 1981. **149**: p. 761-786.

22. Nelson, H.C.M., et al., *The structure of an oligo(dA)-oligo(dT) tract and its biological implications.* Nature, 1987. **330**: p. 221-226.

23. Chuprina, V.P., et al., *Sequence effects on local DNA topology.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 9087-9091.

24. Shatzky-Schwartz, M., et al., *X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature.* J. Mol. Biol., 1997. **267**: p. 595-623.

25. Young, M.A., B. Jayaram, and D.L. Beveridge, *Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets.* Journal of the American Chemical Society, 1997. **119**(1): p. 59-69.

26. Shui, X.Q., et al., *Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations.* Biochemistry, 1998. **37**(48): p. 16877-16887.

27. Hud, N.V. and M. Polak, *DNA-cation interactions: the major and minor grooves are flexible ionophores.* Current Opinion in Structural Biology, 2001. **11**(3): p. 293-301.

28. Hud, N. and F. Feigon, *Characterization of Divalent Cation Localization in the Minor Groove of the A_nT_n and T_nA_n DNA Sequence Elements by ^{1}H NMR Spectroscopy and Manganese(II).* Biochemistry, 2002. **41**: p. 9900-9910.

29. Felsenfeld, G., D.R. Davies, and A. Rich, *Formation of a 3-stranded polynucleotide molecule* Journal of the American Chemical Society, 1957. **79**(8): p. 2023-2024.

30. Sklenar, V. and J. Feigon, *Formation of a Stable Triplex from a Single DNA Strand.* Nature, 1990. **345**(6278): p. 836-838.

31. Plum, G.E., *Thermodynamics of oligonucleotide triple helices.* Biopolymers, 1997. **44**(3): p. 241-256.

32. Pilch, D.S., C. Levenson, and R.H. Shafer, *Structural-analysis of the (dA)10-2(dT)10 triple helix.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(5): p. 1942-1946.

33. Davis, J.T., *G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry.* Angew. Chem. Int. Ed. Engl., 2004. **43**(6): p. 668-698.

34. Davis, J.T. and G.P. Spada, *Supramolecular architectures generated by self-assembly of guanosine derivatives.* Chemical Society Reviews, 2007. **36**(2): p. 296-313.

35. Schultze, P., R.F. Macaya, and J. Feigon, *3-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG).* Journal of Molecular Biology, 1994. **235**(5): p. 1532-1547.

36. Hud, N.V., et al., *The selectivity for K^+ versus Na^+ in DNA quadruplexes is dominated by relative free energies of hydration: A thermodynamic analysis by ^{1}H NMR.* Biochemistry, 1996. **35**(48): p. 15383-15390.

37. Melchior, J., W.B. and P.H. von Hippel, *Alteration of the Relative Stability of dA - dT and dG* dC Base Pairs in DNA.* Proc Natl. Acad. Sci. USA, 1973. **70**: p. 298-302.

38. Marky, L.A., D. Patel, and K.J. Breslauer, *Effect of tetramethylammonium ion on the helix-to-coil transition of poly(deoxyadenylylthymidine) - a nuclear magnetic-resonance and calorimetric investigation.* Biochemistry, 1981. **20**(6): p. 1427-1431.

39. Rees, W.A., et al., *Betaine can eliminate the base pair composition dependence of DNA melting.* Biochemistry, 1993. **32**(1): p. 137-144.

40. Eliasson, R., et al., *Stability of deoxyribonucleic acid in glycol solution.* Biochimica et Biophysica Acta, 1963. **68**(2): p. 234-239.

41. Bonner, G. and A.M. Klibanov, *Structural stability of DNA in nonaqueous solvents.* Biotechnology and Bioengineering, 2000. **68**(3): p. 339-344.

42. Luzzati, V., et al., *Structure transitions observed in DNA + poly A in solution as function of temperature + pH.* Journal of Molecular Biology, 1964. **10**(1): p. 28-41.

43. Green, G. and H.R. Mahler, *Optical properties of DNA in ethylene glycol.* Biopolymers, 1968. **6**(10): p. 1509-1514.

CHAPTER 7

BISULFITE ACCESSIBILITY IN G-TRACTS[f]

7.1. Introduction

Several major chromosomal fragile sites in human lymphomas, including the bcl-2 major breakpoint region (MBR), bcl-1 major translocation cluster (MTC), and c-myc exon 1-intron 1 boundary, contain distinctive sequences of consecutive cytosines that exhibit a high degree of reactivity with the structure-specific chemical probe bisulfite. The sequence-specific structural variations of the dsDNA helix are well-documented.[1-4] For example, DNA sequences containing four or more A or T residues without a d(TpA) step can adopt a variation of the B-form helix with a narrower minor groove, whereas sequences with consecutive G (or C) bases are more prone to adopt the A-form helix under dehydrating conditions.[5] The A-form helix observed in RNA and DNA-RNA hybrid duplexes is present during transcription and replication, and sometimes when DNA is bound to proteins.[6-10] Most reported DNA crystal structures exhibit either the B- or A-form helical structure, with a small fraction exhibiting an intermediate structure. A strong correlation exists between the propensity of a given DNA sequence to adopt a non-B-form structure and the interaction of solvent (water and ions) with the major and minor grooves [5], but neither the natural range of DNA conformations under

---

[f] This chapter was adapted from previously published work. (Tsai, AG; Engelhart, AE; Hatmal, MM; Houston, SI; Hud, NV; Haworth, IS; Lieber, MR. "Conformational variants of duplex DNA correlated with cytosine-rich chromosomal fragile sites." J. Biol. Chem. (2009) 284:7157-7164.). It is reproduced with permission.

physiologic solution-phase conditions, nor how subtle variations in DNA structure may influence the activity of DNA binding proteins, is well understood.

The structure of a particular sequence of double-stranded DNA (dsDNA) is often inferred from its pattern of reactivity with structure-specific chemical or enzymatic probes. After treatment with such a probe, the DNA is typically cleaved at reactive sites, and the positions of these cleaved sites are calculated from the fragment lengths observed using gel-based detection methods such as direct labeling, primer extension, and ligation-mediated PCR.

Unlike cleavage-based methods, bisulfite converts single-stranded, unmethylated cytosines into uracils, which can be cloned and sequenced.[11-15] This is technically much easier to perform and reproduce, ensures specificity to the sequence of interest, and allows one to examine specific clones and, thereby, resolve multiple structures. Previous work has demonstrated the utility of bisulfite in analyzing conceptually simple, stable structures such as R-loops and cruciforms.[11, 12] For these structures, the reasons for the observed pattern of bisulfite reactivity are very clear: they contain completely unpaired cytosines, and those cytosines are the primary ones to react.

However, in genomic regions without demonstrated R-looping or cruciforming ability, the reasons for bisulfite reactivity are less clear, as at the bcl-1 major translocation cluster (MTC) and possibly the bcl-2 major breakpoint region (MBR).[13] Reactivity is largely symmetrical between the two strands, increases with supercoiling, and does not depend on the source of the DNA. Because the structural basis for reactivity in these presumably double-stranded regions is unclear, we performed bisulfite probing on oligonucleotides containing sequences previously characterized by other groups using

other methods.[16, 17] Integration of the genomic in situ and in vitro studies with additional methods done here (circular dichroism) and elsewhere (X-ray crystallography and NMR) have permitted us to formulate a coherent view of how regions with consecutive cytosine induce a conformation that is intermediate between B-form and A-form DNA. The frequency of this conformation in zones of chromosomal fragility is likely to be relevant to understanding why they are fragile.

7.2. Experimental Procedures

7.2.1 Materials

DNA oligos for bisulfite analyses were ordered from Operon Biotechnologies and oligos containing 2'-O-methyl RNA were ordered from IDT. All oligos were purified by denaturing PAGE. 650 pmol of gel-purified long strand oligo was mixed with 1,300 pmol of gel-purified short strand oligo in 65 μL TE with 100 mM NaCl in a screw-cap tube, and annealed by boiling the tube in 1 L water for 5 minutes, then allowing it to cool to room temperature overnight. For CD analyses, DNA and RNA oligonucleotides from IDT (Coralville, IA) were used as received.

Substrates are listed long strand-short strand as follows: GGGCCC, AT160-AT161; GGCGCC, AT162-AT163; GGCC, AT164-AT165; GCGC, AT166-AT167; CGCG, AT168-AT169; GCGCGC, AT183-AT184; 2'-O-methyl-GGGCCC, AT183-AT259; 2'-O-methyl-GCGCGC, AT160-AT264. Oligonucleotide sequences are as follows: AT160, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TGG GCC CAT GCG ATC CTT GAA GGA ATT TGG AGA GAG GGG T); AT161, d(TCA AGG ATC GCA TGG GCC CAT GAA CGC AGG T); AT162, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TGG CGC CAT GCG ATC CTT GAA GGA ATT TGG

AGA GAG GGG T); AT163, d(TCA AGG ATC GCA TGG CGC CAT GAA CGC AGG T); AT164, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TGG CCA TGC GAT CCT TGA AGG AAT TTG GAG AGA GGG GT); AT165, d(TCA AGG ATC GCA TGG CCA TGA ACG CAG GT); AT166, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TGC GCA TGC GAT CCT TGA AGG AAT TTG GAG AGA GGG GT); AT167, d(TCA AGG ATC GCA TGC GCA TGA ACG CAG GT); AT168, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TCG CGA TGC GAT CCT TGA AGG AAT TTG GAG AGA GGG GT); AT169, d(TCA AGG ATC GCA TCG CGA TGA ACG CAG GT); AT183, d(GTG GGG TTA TTG TGG GTG TAC CTG CGT TCA TGC GCG CAT GCG ATC CTT GAA GGA ATT TGG AGA GAG GGG T); AT184, d(TCA AGG ATC GCA TGC GCG CAT GAA CGC AGG T); AT259, TCA AGG ATC G (CAU GCG CGC AUG)$_{2'\text{-O-methyl RNA}}$AAC GCA GGT; AT264, d(TCA AGG ATC G(CAU GGG CCC AUG)$_{2'\text{-O-methyl RNA}}$AAC GCA GGT); AT27, d(ACC CCT CTC TCC AAA TTC CT); AT131, d(GTG GGG TTA TTG TGG GTG T).

7.2.2 Bisulfite Analysis of Duplex Oligonucleotides

As required for bisulfite probing, we lengthened the ends to improve stability at 37° C and lengthened one strand to provide space for PCR primers. As a result, each substrate is composed of a long strand (68 to 70 nt) and a short strand (29 to 31 nt), but only the long strand can be PCR-amplified and sequenced, and thus only long strand C to U conversions can be detected. The flanking single-stranded regions contain no cytosines in order to prevent any PCR bias towards less-reacted molecules, and this arrangement also allows us to see reactivity at the ends of the double-stranded portion.

The bisulfite mixture was prepared by mixing 0.5 g $NaHSO_3$ (Sigma) with 0.525 mL of ddH_2O and 0.2626 mL of 2 M NaOH, then mixing 457.5 μL of the resulting solution with 12.5 μL of 20 mM hydroquinone. 15 μL of annealed oligo was mixed with 235 μL of bisulfite mixture and put into an air incubator at 37° C for 16 hours. After treatment, oligos were repeatedly ethanol precipitated until the pellet could be dissolved in 20 μL TE, and 5 μL was run on 5% native mini-PAGE along with small amounts of component single-stranded oligos and untreated substrate more than one lane away. The gel was stained with 0.25 μg/mL ethidium bromide in deionized water for 20 minutes with light shaking, transferred to a UV transilluminator overlain with clean plastic wrap, and the double-stranded form cut out with a clean razor blade. The gel slice was treated 4 times with 0.3 M NaOH for 5 minutes at room temperature, washed 4 times with TE for 5 minutes at room temperature, then crushed and soaked in 200 μL PAGE diffusion buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1% SDS, 1 mM EDTA, pH 8.0) at 37° C, 250 rpm for >8 hours. The supernatant was precipitated with ethanol and resuspended in 20 μL TE. 0.5 μL was used for a 10 μL PCR with primers AT131 and AT27 using NEB Taq DNA polymerase using 30 cycles of 94° C, 0:30; 55° C, 0:30; 72° C, 0:30; with an initial denaturation 94° C, 2:00 and final extension 72°C, 2:00. Primers are designed to amplify only the long strand and do not anneal to any cytosines on the long strand, preventing bisulfite conversion PCR bias. PCR products were checked by running on 5% native PAGE, Invitrogen topo TA cloned, and individual molecules sequenced on a Li-Cor IR^2 DNA analyzer using M13 forward primer according to manufacturers' directions. Molecules with bisulfite conversions at all or all but one cytosine were interpreted as having become single-stranded either before or during

bisulfite treatment and not used for analysis. 16 molecules of data were obtained for each PCR; two PCRs were done for each bisulfite treatment; and two bisulfite treatments were done for each substrate, giving 64 molecules of data per substrate.

7.2.3. Calculation of Electrostatic Potentials

Coordinates for $[d(CATGGGCCCATG)]_2$ and $[d(CATGCGCGCATG)]_2$ in the canonical B-form helix were generated using NUCGEN in the AMBER8 suite of programs.[18] These coordinates were converted to PQR format with AMBER charges using PDB2PQR.[19] The electrostatic potential of these structures was calculated using the nonlinear Poisson-Boltzmann equation as implemented in APBS.[20] The bisulfite concentration and temperature were those used in the bisulfite probing experiments. The sodium ion was assigned a radius of 2 Å, while the bisulfite and sulfite ion radii were assigned 4 Å. Electrostatic potentials were converted into bisulfite concentration maps using the relationship $C(x,y,z) = C_b e^{(-(Z*\Phi(x,y,z)))}$, where C_b is the bulk ion concentration, Z is the ion valence, and $\Phi(x,y,z)$ is the electrostatic potential.[21] These maps were visualized using the VMD software package using a spherical probe of radius 4 Å.[22]

7.2.4. Circular Dichroism

Circular dichroism was performed on solutions containing 10 μM total oligonucleotide in a 1:1 mixture of disodium phosphate and monosodium phosphate in 5 mm rectangular cells on a JASCO J-810 at 5°C.

7.3. Bisulfite reactivity at the bcl-1 MTC and bcl-2 MBR

Bisulfite catalyzes the deamination of single-stranded, unmethylated cytosine to uracil.[11] First, the bisulfite anion adds to the 5-6 double-bond of cytosine, breaking aromaticity and resulting in cytosine sulfonate.[23, 24] When protonated, this modified cytosine can undergo a rate-limiting hydrolytic deamination to uracil sulfonate. Later, mild base and dilution can eliminate the sulfonate group, yielding uracil. Stacking within duplex DNA is thought to protect cytosines from bisulfite attack, providing the basis for structural specificity (i.e., sensitivity of single-stranded regions).[11-13] Additionally, the presence of two polyanionic backbones makes the close approach of the bisulfite anion electrostatically unfavorable.

Several chromosomal fragile sites are in close proximity to bisulfite-reactive regions rich in stretches of consecutive cytosines (Figures 7.1, 7.2, and [13]). An abnormally high proportion of breakpoints from t(14;18)(q32;q21) translocations occur in a region termed the bcl-2 major breakpoint region (MBR), and bisulfite reactivity at the MBR was previously reported by Lieber et al.[13] The sum of reactivity on the top and bottom strands is plotted in the top graph of Figure 7.1A. One can see the correlation with the number of consecutive cytosines in the second plot, and this correlation is better than with any other sequence composition (lower two plots).

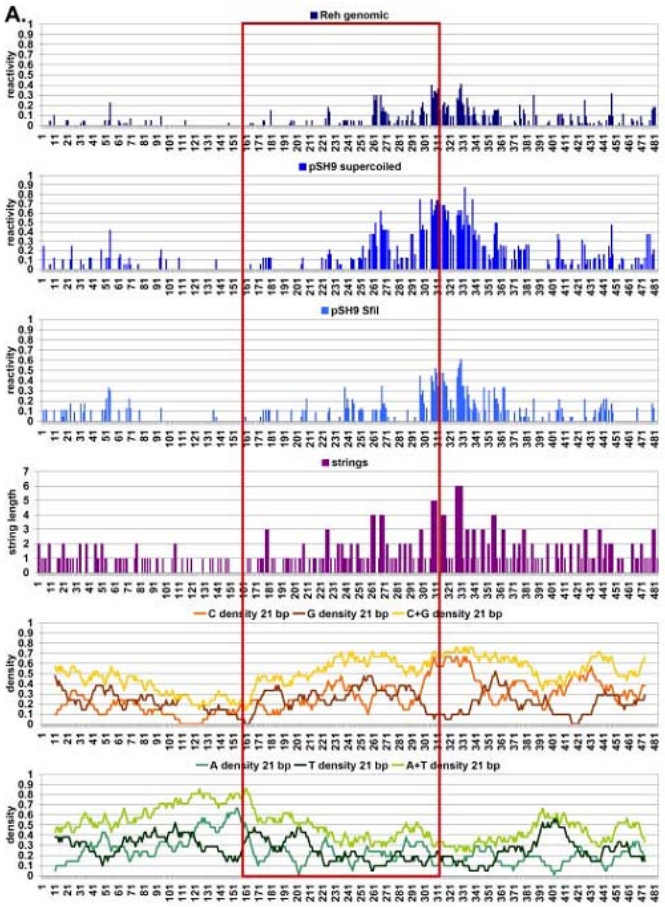


Figure 7.1. Correlation of bcl-2 MBR and bcl-1 MTC bisulfite reactivity with sequence features. (A) Bisulfite reactivity of the bcl-2 MBR has been published and is plotted in the top panel.[13] The x-axis denotes every base position along the 528 bp fragment (which includes A and T, in addition to C and G). C $\rightarrow$ T and G $\rightarrow$ A conversion percentages are plotted, calculated as the number of molecules converted at that particular position, divided by the total number of molecules sequenced for that strand. Bases in the second panel are aligned with those in the top panel, and plot the C or G "string" length, or the length of consecutive Cs or Gs within which a particular C or G position is located. For instance, ACGGGGCCGGT would have values of 0, 1, 4, 4, 4, 4, 2, 2, 2, 2, 0 for each base position, from left to right. The bottom two graphs show the C, G, C+G, A, T, and A+T densities calculated using a moving window of 21 bp. While there is some correlation between C and C+G density and bisulfite reactivity, the fine peak structure of reactivity matches best with the string length.
(B) Bisulfite reactivity at the bcl-1 MTC. Reactivity and sequence features for the bcl-1 MTC are plotted similar to Figure 7.1A, but using the data from Fig 7.2 for the three bisulfite plots.

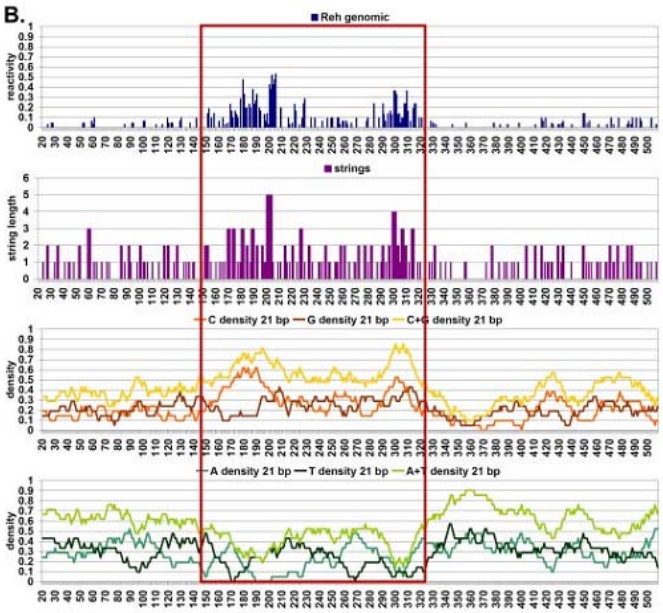


Figure 7.1 continued.

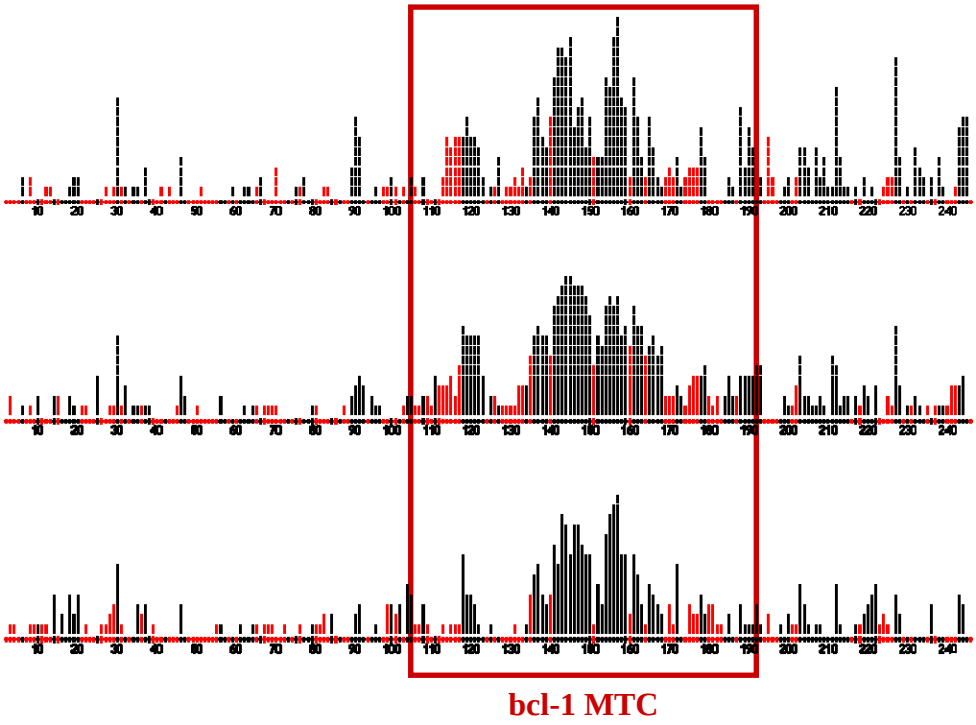


Figure 7.2. Bisulfite reactivity of genomic DNA, supercoiled DNA and linear plasmid DNA containing the the bcl-1 major translocation cluster (MTC). Extracted Reh Genomic DNA: 7.8% total conversion frequency, 44 top strands, 20 bottom strands; pSH 9 Supercoiled: 19.4% total conversion frequency. 19 top strands, 8 bottom strands; pSH 9 Sfil-Cut: 11.5% total conversion frequency, 23 top strands, 9 bottom strands.

Because breakpoints also occur frequently in the bcl-1 major translocation cluster (MTC) from t(11;14)(q13;q32) translocations, we decided to probe this region using bisulfite as well.[25] Bisulfite reactivity occurs primarily at the downstream edge of the bcl-1 MTC (top plot of Figure 7.1B and all of Figure 7.2). Two peaks of reactivity are observed. The first and smaller peak occurs at the semi-palindromic sequence cgaggggaagcccctcc. The second, broader, and larger peak occurs at the sequence ccctctaagccccctctccccgtcacatccccccgaccctgcc. Note the stretches of consecutive Cs within both sequences.

The pattern of reactivity is very similar between the different substrates – extracted genomic DNA, supercoiled plasmid, or linearized plasmid. However, the overall level of bisulfite reactivity is higher on supercoiled plasmid than on linearized plasmid or genomic DNA (Fig 7.2). Supercoiling destabilizes the duplex, accelerating DNA breathing and allowing it to more easily form non-B-DNA structures.

Because the top and bottom strands of the original bisulfite-probed DNA molecule are separated during PCR and separately cloned, each sequence obtained gives information for deamination events on only one of the two original strands. When the reactivity is rescaled to balance the top and bottom strand representation, it is roughly symmetrical between the two strands (top plot of Figure 7.1B; compare with Figure 7.2, which displays top strand in black and bottom strand in red).

To rule out our primers biasing against binding to converted regions, we used a second set of outer primers to confirm that few conversions occur at the binding sites of the first set. Another strategy – designing two sets of primers, one set insensitive to top

strand conversions and the other insensitive to those on the bottom – was also performed. These show no difference in pattern.

The pattern does not change if bisulfite treatment is done on intact Reh cells embedded in 1% low-melt agarose, or intact *E. coli* embedded in agarose after transformation with plasmid containing the bcl-1 sequence. Any proteins that might bind to the DNA are likely denatured by the high salt concentration of the bisulfite solution. The pattern also does not change if the bisulfite solution is adjusted to pH 6, though overall reactivity is decreased about three-fold – consistent with the three-fold lower efficiency of the bisulfite-catalyzed deamination reaction itself at pH 6 compared to pH 5.2.[24] The reactivity increases proportionately if DNA is treated for 24 hours rather than the usual 16 hours; the relative reactivity remains the same. Nor does it change if the two peaks are separated and cloned into different plasmids, or if the plasmid or genomic DNA are pretreated with RNase H, or if the plasmid is extracted from *E. coli* by a non-denaturing method, rather than alkaline lysis. Therefore, the specific pattern of bisulfite reactivity appears to depend primarily upon the DNA sequence.

7.4. Analysis of DNA Duplexes Using Bisulfite

In order to determine the basis for the full range of reactivity of bisulifte with DNA, we probed several well-characterized and structurally divergent duplexes with bisulfite: d(…catGGGCCCatg…)$_2$, d(…catGGCGCCatg…)$_2$, and d(…catGCGCGCatg…)$_2$ (Figure 7.3 and tabulated in the top 3 rows of Table 7.1). Consistent with its single-strand specificity, bisulfite virtually always detected the frayed edges of the duplex. Most often, only one of the three Cs in any given molecule reacted with bisulfite. Interestingly, we found that the GGGCCC duplex was 5-fold more reactive

than GCGCGC, with GGCGCC exhibiting intermediate reactivity. A similar, but less dramatic effect was observed for GGCC over GCGC or CGCG (Figure 7.4 and middle rows of Table 7.1).

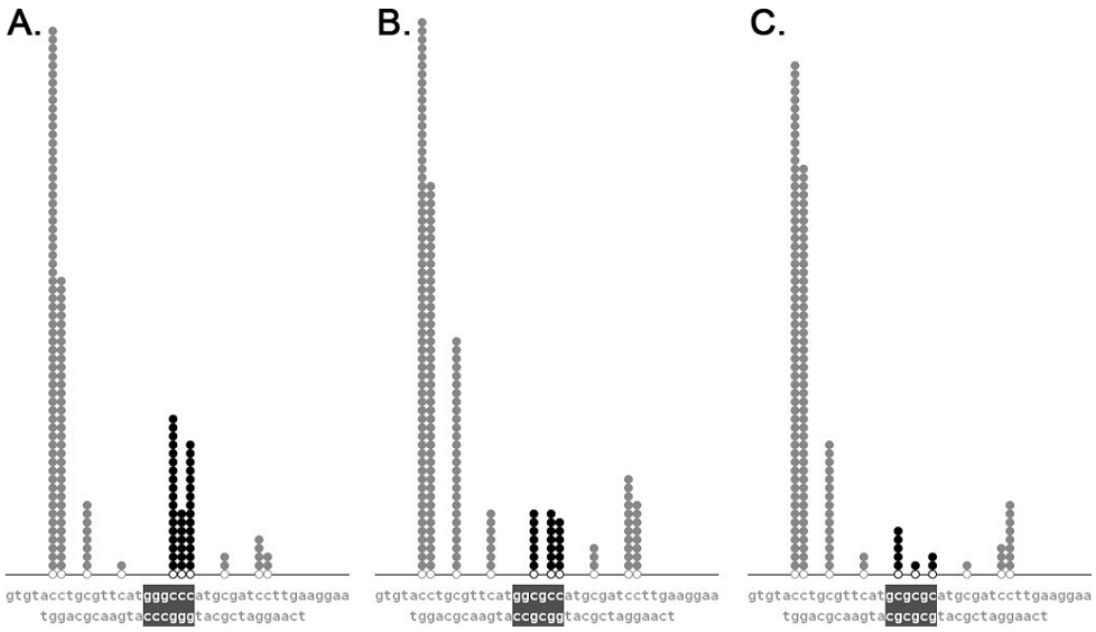


Figure 7.3. Bisulfite reactivity with double-stranded DNA containing GGGCCC, GGCGCC, and GCGCGC sequence elements. Bisulfite reactivities for each cytosine on each duplex DNA substrate are shown in (A–C). Boxed in dark gray are the sequences of interest, and each black circle denotes a bisulfite-catalyzed C to U conversion detected after cloning and sequencing. 30 nt of the long strand sequence have been truncated for space.

Table 7.1. Compilation of bisulfite reactive sites within the central zone of the duplex oligonucleotides.

Sequence	Number of sites reacted	Percentage of sites reacted
$d(GGGCCC)_2$	40	21
$d(GGCGCC)_2$	20	10
$d(GCGCGC)_2$	8	4.2
$d(GGCC)_2$	12	9.4
$d(GCGC)_2$	4	3.1
$d(CGCG)_2$	2	1.6
d(GGGCCC) r(GGGCCC)	1	0.52
d(GCGCGC) r(GCGCGC)	0	0

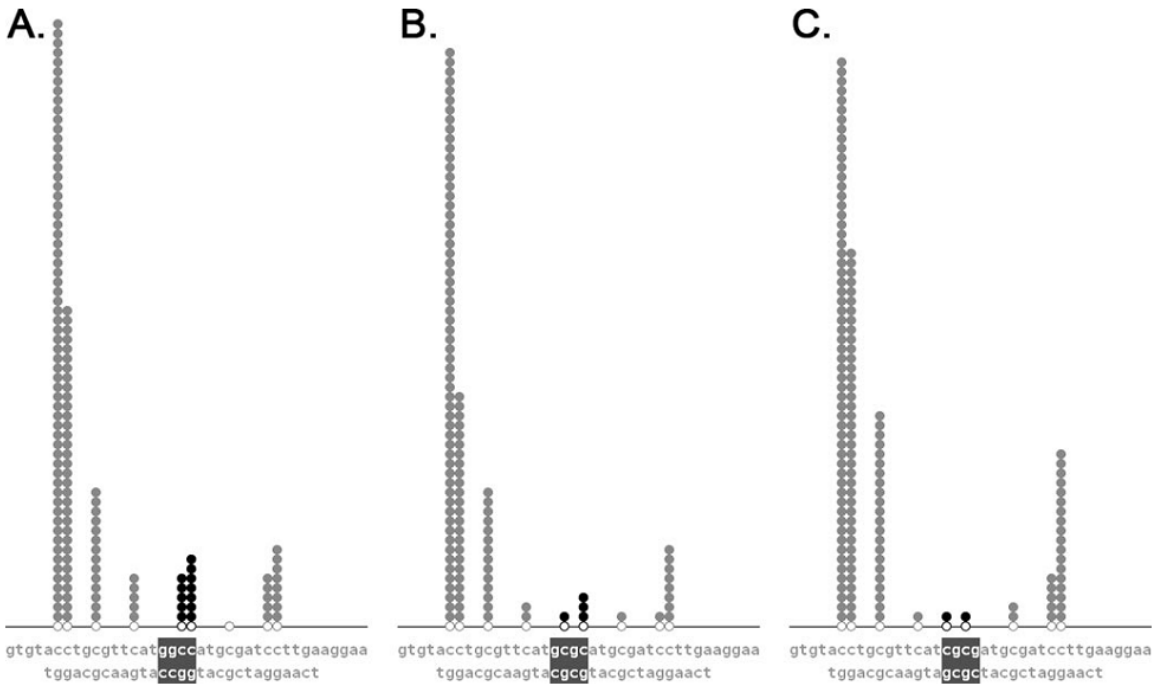


Figure 7.4. Bisulfite reactivity with dsDNA containing GGCC, GCGC, and CGCG sequence elements. Bisulfite reactivities for each cytosine on each duplex DNA substrate are shown in A–C. Boxed in dark gray are the sequences of interest, and each black circle denotes a bisulfite-catalyzed C to U conversion detected after cloning and sequencing. 30 nt of the long strand sequence have been truncated for space.

Clearly, bisulfite detects one or more subtle structural differences between these duplexes. Circular dichroism showed the series d(CGCATATATGCG)$_2$ (B-form DNA), d(catGGGCCCatg)$_2$, d(catGGCGCCatg)$_2$, d(catGCGCGCatg)$_2$, and r(cauGCGCGCaug)$_2$ (A-form RNA) exhibits increasing A-form character (Figure 7.5). This effect was present at two markedly different salt concentrations. Consistent with these data, a B/A-intermediate structure of d(catGGGCCCatg)$_2$ (PDB ID 1DC0) was previously observed by X-ray crystallography.[16] The bisulfite reactivity of a corresponding A-form DNA/2'-O-methyl RNA heteroduplex was lower than that for B-form DNA (Figure 7.6 and bottom rows of Table 7.1), suggesting a correlation between structure and reactivity of B/A > B-DNA > A-DNA.

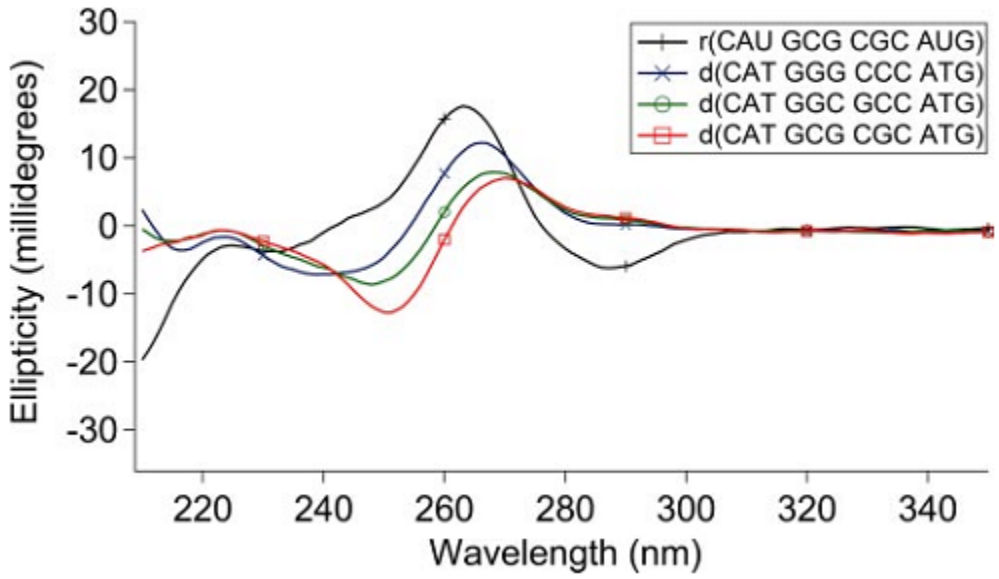


Figure 7.5. Circular dichroism of nucleic acids containing GGGCCC, GCGCGC, and GGCGCC sequence elements. Circular dichroism spectra of B-form DNA (catGCGCGCatg), A-form RNA (cauGCGCGCaug), and two B/A-form intermediate sequences (catGGCGCGatg and catGGGCCCatg) in 100mM Na^+ are shown.

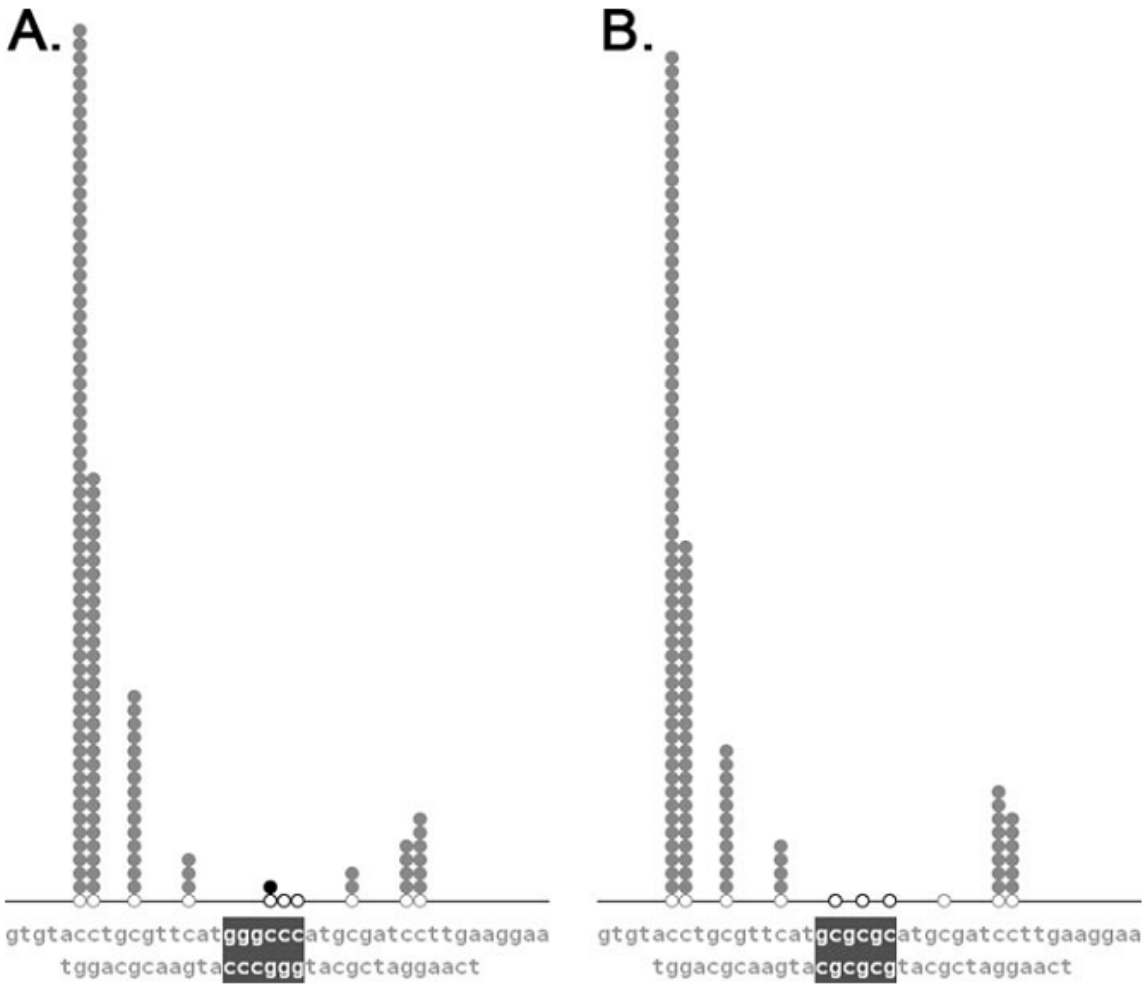


Figure 7.6. Bisulfite is minimally reactive with DNA-2'-O-methyl RNA hybrid duplexes. Bisulfite reactivities for each cytosine on each hybrid duplex DNA-2'-O-methyl RNA substrate are shown. Notations are as in Figure 7.3.

7.5. Electrostatic Simulations

Given that the sequence-specific electrostatic potential of the major groove may also enhance the bisulfite-reactivity of consecutive cytosine bases, we used the non-linear Poisson-Boltzmann equation [21] to estimate the local occupancy of the major groove with bisulfite for the duplexes d(catGGGCCCatg)$_2$ and d(catGCGCGCatg)$_2$ in the canonical B-form helix. We found a contiguous electropositive "patch" in the major groove of d(catGGGCCCatg)$_2$ not present in d(catGCGCGCatg)$_2$ (Figure 7.7). Such electropositive patches, composed of consecutive cytosines, could localize the bisulfite anion and, thereby, increase its reactivity with those cytosines. This proposed sequence-specific localization of bisulfite ion at an electropositive region of DNA is analogous to the well-documented sequence-specific localization of cations along electronegative regions of DNA formed by consecutive guanines (in the major groove) or A/T base pairs (in the minor groove).[5]

The sequence-dependent dynamics of base pairing may also contribute to the sequence-dependent bisulfite reactivity reported here for duplex DNA. Imino proton exchange NMR shows that the central six base pairs in d(catGGGCCCatg)$_2$ are overall more dynamic (i.e., prone to base pair opening) than those in d(catGGCGCCatg)$_2$, and the central four in d(catGGCCatg)$_2$ are more unpaired than those in d(catGCGCatg)$_2$ or d(catCGCGatg)$_2$.[17] Dornberger et al. have suggested that sequences with consecutive cytosines adopt a variation of the B-form structure which causes the base pairs in these sequences, when flanked by canonical B-form helices, to breathe more than sequences alternating between guanines and cytosines. This conclusion is consistent with our observation that d(catGGGCCCatg)$_2$ exhibits a CD spectrum that is indicative of a B/A-

form structural intermediate, as this CD could result from the GGGCCC region adopting a more A-form-like structure, while the flanking three base pairs maintain a B-form helix.

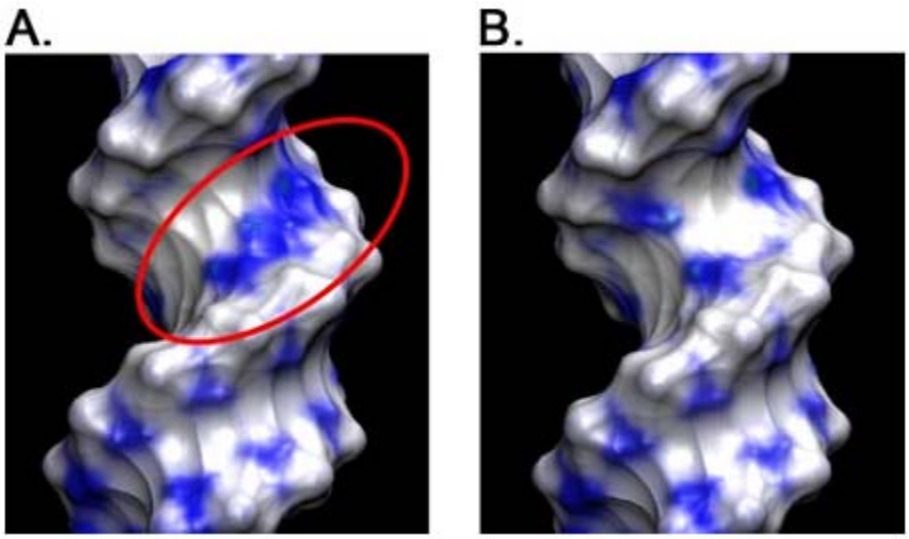


Figure 7.7. Calculated localization of bisulfite ions on double-stranded DNA containing GGGCCC (A) and GCGCGC (B) sequences. Red ellipse in A indicates the electropositive patch that is observed along the major groove edge of the three consecutive cytosine bases.

Finally, supercoiling is known to destabilize duplex DNA and can induce structures which underwind DNA to mitigate the supercoiling. With bcl-2 MBR and bcl-1 MTC sequences, supercoiling enhances overall reactivity, including that at stretches of consecutive cytosines (Figures 7.1-2).

7.6. Conclusion

Lieber and colleagues have previously shown that the bcl-2 MBR has a relatively high degree of bisulfite reactivity which correlates with its recombination efficiency (as measured using a transfection assay which recapitulates many aspects of the bcl-2 translocation, t(14;18)).[13] A three-base mutation disrupting a set of consecutive cytosines reduced both the bisulfite reactivity and the recombination efficiency. These recombination events were also shown to be dependent on the RAG endonuclease. Since then, we found strings of consecutive cytosines to be key features at the bcl-1 and c-myc intron 1-exon1 regions as well.[26, 27]

The current study examines the precise structural conformations reflected by bisulfite reactivity. Bisulfite appears to be exquisitely sensitive to very subtle perturbations of DNA structure which increase the steric accessibility of cytosine and alter the electrostatic potential of the major groove. While some effect on bisulfite reactivity was expected from fraying and supercoiling, the dramatic effect observed simply by switching two base pairs – from G<u>G</u>GC<u>C</u>C to G<u>C</u>GC<u>G</u>C – was entirely unexpected and is indicative of significant sequence-dependent structural alterations. Bisulfite reactivity at consecutive cytosines appears to be invariant with length; it is observed in large duplexes (e.g., plasmids and genomic DNA) and short oligonucleotide substrates.

In light of these results, we suggest the possibility that consecutive cytosines induce structural and electrostatic changes that predispose to nucleases, such as the RAG complex, which is essential for recombination *ex vivo*. While the precise structural mechanisms remain speculative, both the RAG complex and the Artemis:DNA-PK_{CS} complex are known to cleave at bubble structures (single-stranded regions) as small as 1 to 3 nt.[26-28] A study of c-myc chromosomal translocations in sporadic Burkitt lymphomas – known to depend on a single-stranded cytosine deaminase, AID, identified a consensus motif of 5'-CTCCTCCCC-3' for translocation breakpoints, consistent with translocation zones being GC-rich and rich in consecutive cytosines.[29] To the extent that bisulfite reactivity reflects a tendency toward B/A-intermediate structure or single-stranded character, bisulfite-reactive regions with stretches of consecutive cytosines, such as these translocation cluster regions, appear to be targets for such DNA double-strand break mechanisms.

7.7. References

1. Saenger, W., *Principles of Nucleic Acid Structure*. Springer Advanced Texts in Chemistry, ed. C. Cantor. 1984: New York: Springer-Verlag. 556.

2. Calladine, C.R., et al., *Understanding DNA: The Molecule and How it Works*. 3rd ed. 2004, San Diego: Elsevier Academic Press.

3. Sinden, R.R., *DNA Structure and Function*. 1994, San Diego: Academic Press. 398.

4. Mirkin, S.M., *DNA structures, repeat expansions and human hereditary disorders.* Curr Opin Struct Biol, 2006. **16**(3): p. 351-8.

5. Hud, N.V. and J. Plavec, *A unified model for the origin of DNA sequence-directed curvature.* Biopolymers, 2003. **69**(1): p. 144-159.

6. Wahl, M.C. and M. Sundaralingam, *Crystal structures of A-DNA duplexes.* Biopolymers, 1997. **44**(1): p. 45-63.

7. Hays, F.A., et al., *How sequence defines structure: a crystallographic map of DNA structure and conformation.* Proc Natl Acad Sci U S A, 2005. **102**(20): p. 7157-62.

8. Svozil, D., et al., *DNA conformations and their sequence preferences.* Nucleic Acids Res, 2008. **36**(11): p. 3690-706.

9. Vargason, J.M., K. Henderson, and P.S. Ho, *A crystallographic map of the transition from B-DNA to A-DNA.* Proc. Natl. Acad. Sci. USA, 2001. **98**: p. 7265-7270.

10. Lu, X.J., Z. Shakked, and W.K. Olson, *A-form conformational motifs in ligand-bound DNA structures.* Journal of Molecular Biology, 2000. **300**(4): p. 819-840.

11. Gough, G.W., K.M. Sullivan, and D.M. Lilley, *The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite.* EMBO J., 1986. **5**: p. 191-196.

12. Yu, K., et al., *R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells.* Nature Immunol., 2003. **4**: p. 442-451.

13. Raghavan, S.C., et al., *A non-B-DNA structure at the bcl-2 major break point region is cleaved by the RAG complex.* Nature, 2004a. **428**: p. 88-93.

14. Clark, S.J., et al., *High sensitivity mapping of methylated cytosines.* Nucleic Acids Res., 1994. **22**: p. 2990-2997.

15. Raizis, A.M., F. Schmitt, and J.-P. Jost, *A bisulfite method of 5-methylcytosine mapping that minimizes template degradation.* Anal. Biochem., 1995. **226**: p. 161-166.

16. Ng, H.L., M.L. Kopka, and R.E. Dickerson, *The structure of a stable intermediate in the A <--> B DNA helix transition.* Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2035-9.

17. Dornberger, U., M. Leijon, and H. Fritzsche, *High base pair opening rates in tracts of GC base pairs.* J. Biol. Chem., 1999. **274**: p. 6957-6962.

18. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* . Comp. Phys. Commun., 1995. **91**: p. 1-41.

19. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W665-7.

20. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.

21. Lamm, G., *The Poisson-Boltzmann Equation*, in *Reviews in Computational Chemistry*, K.B. Lipkowitz, R. Larter, and T.R. Cundari, Editors. 2003, Wiley-VCH, John Wiley & Sons, Inc.: Berlin. p. 147-365.

22. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics and Modeling, 1996. **14**(1): p. 33-38.

23. Shapiro, R., V. DiFate, and M. Welcher, *Deamination of cytosine derivatives by bisulfite: mechanism of the reaction.* J. Am. Chem. Soc., 1974. **96**: p. 906-912.

24. Sono, M., Y. Wataya, and H. Hayatsu, *Role of bisulfite in the deamination and the hydrogen isotope exchange of cytidylic acid.* J. Am. Chem. Soc., 1973. **95**: p. 4745-4749.

25. Bertoni, F., E. Zucca, and F.E. Cotter, *Molecular basis of mantle cell lymphoma.* Br J Haematol, 2004. **124**(2): p. 130-40.

26. Raghavan, S.C., et al., *Double-strand break formation by the RAG complex at the bcl-2 Mbr and at other non-B DNA structures in vitro.* Mol. Cell. Biol., 2005a. **25**: p. 5904-5919.

27. Raghavan, S.C., et al., *The structure-specific nicking of small heteroduplexes by the RAG complex: implications for lymphoid chromosomal translocations.* DNA Repair (Amst), 2007. **6**(6): p. 751-9.

28. Ma, Y., K. Schwarz, and M.R. Lieber, *The Artemis:DNA-PKcs Endonuclease Can Cleave Gaps, Flaps, and Loops.* DNA Repair, 2005. **4**: p. 845-851.

29. Abeysinghe, S.S., et al., *Translocation and gross deletion breakpoints in human inherited disease and cancer I: nucleotide composition and recombination-associated motifs.* Human Mut., 2003. **22**: p. 229-244.

CHAPTER 8

CONCLUSION[g]

8.1. Future Directions

While great progress has been made towards a robust, high-fidelity, and high-yielding scheme for the synthesis of RNA from self-assembled monomers components, this goal has not yet been realized. However, several potential solutions to these challenges become available if we accept the possibility that one or more proto-RNAs predated RNA. More than a mere academic exercise, the discovery of a self-assembling RNA-like polymer would have tremendous potential for applications in biotechnology, medicine and materials science. The results described herein, in addition to several promising recent reports, suggest that such a reinvention may occur in the very near future.

Acetal linkages form easily and reversibly, thereby providing the benefits demonstrated by the reversible imine chemistry described by Lynn. Phosphorus (V) esters and acetals are both tetrahedral and approximately isosteric, and it is known that aNAs can base pair with RNAs.[1] With the proper choice of aldehyde or ketone, these linkages can provide a negative charge that has proven critical to the phosphodiester backbone. For these reasons, acetal-linked nucleic acids (*a*NAs) are an intriguing potential ancestral polymer.[2] Acetal formation occurs via a hydrolytically labile

---

[g] This chapter was adapted from previously published work. (Engelhart, AE; Hud, NV. "Primitive Genetic Polymers." Cold Spring Harb. Perspect. Biol. (2010) DOI: 10.1101/cshperspect.a002196.). It is reproduced with permission. © Cold Spring Harbor Laboratory Press.

hemiacetal intermediate, providing a two-step error-correcting scheme during replication. Additionally, bond formation is favored over hydrolysis upon water removal. The concentration of substrates that occurs during dry-down, driving bond formation, would also favor the hybridization of short, complementary sequences, providing an accessible means by which to drive polymerization.[3] The functionalization of an acetal with a negatively charged group, such as carboxylate group in the case of the aldehyde glyoxylate, would provide a localized charge near the backbone acetal linkage, thereby providing many of the favorable electrostatic properties observed in the case of phosphate. Glyoxylate has been shown to form in a prebiotically plausible reaction from glycolaldehyde.[4] We have speculated that glyoxylate could have preceded phosphate in an early proto-RNA, in the form of glyoxylate-acetal nucleic acids (*ga*NAs).[5] Based on energy minimized molecular models, it appears that a *ga*NA duplex would exhibit helical properties very similar to RNA, suggesting that information transfer could occur between the two polymers.[5] The formation of *ga*NA dinucleotides has been demonstrated upon heating and drying of a neutral solution containing nucleosides, glyoxylate and mono- or divalent cations. These acetal linkages formed by glyoxylate and nucleosides are surprisingly stable to hydrolysis – no detectable decomposition occurs after two weeks at room temperature and neutral pH. Importantly, and unlike in the case of phosphate linkages, *ga*NAs can be hydrolyzed at moderately elevated temperature in the presence of salt (60°C, 1 M $MgCl_2$). An accessible means to modulate polymer stability (in this case, temperature and salt concentration) provides an attractive means by which to "turn on and off" thermodynamic control of polymerization – a crucial feature in the chemical synthesis of an early informational polymer. It is possible that other, similar, molecules

could form such linkages. For instance, pyruvate, like glyoxylate, provides a negative charge and tetrahedral geometry, and ketal bond formation is still expected to be more facile than phosphodiester bond formation. Further, pyruvate is ubiquitous in life and formed in excellent yield in model prebiotic reactions.[4]

Another reduced phosphorus species, phosphite, is particularly attractive as a potential backbone component. It exhibits enhanced solubility relative to phosphate (which tends to form mineral precipitates in the presence of divalent ions). Additionally, it enjoys enhanced reactivity towards esterification relative to phosphate. Schwartz and coworkers reported high yields of uridine 5′-H-phosphonate (as high as 44% in the presence of urea) in 60 ºC drying reactions of aqueous ammonium phosphite, uridine and, optionally, urea. When dried at 110 ºC, these authors reported production of small amounts of putative higher-order products, the most abundant of which was produced in 4% yield, which they tentatively assigned as the 5′,5′ H-phosphonate-linked dinucleotide. This species could be oxidized with iodine to a second species, assigned as the phosphate-linked dinucleotide.[6] Like acetals, H-phosphonate diesters are conditionally reversible, low-energy linkages. While they are uncharged, upon oxidation, they become negatively charged phosphate diesters, making phosphonates another attractive, conditionally reversible backbone linkage. These linkages are particularly appealing, since they generate the contemporary phosphodiester linkage upon oxidation.

While our current examples of self-assembling polymers with reversible linkages are limited, the properties of these polymers are impressive and indicative of the promise held by the development of this area of polymer science. For example, dynamic covalent assemblies have been examined by the Lehn laboratory in the form of guanosine

hydrogels, and they have shown that these self-assembling materials can act as controlled release media for bioactive small molecules.[7, 8] Similarly, these investigators, as well as Leibler and coworkers, have also shown that polymers with the capacity for constitutional reorganization exhibit a number of intriguing material properties, including self-healing and chemoresponsive characteristics.[9, 10]

Most recently, Ghadiri and coworkers have demonstrated the ability for peptide nucleic acid polymer with reversible linkages between a polycysteine backbone and thioester acyclonucleosides to undergo a form of evolution in response to the presence of different DNA templates.[11] Almost simultaneously, the Liu group reported a similar system, involving amidation or reductive amination of a PNA backbone with acyclonucleosides containing carboxylate and aldehyde functional groups.[12] One can conceive of the development of RNA-like polymers following a similar theme that create functional polymers that are environmentally responsive, perhaps even evolving new functions in response to unanticipated environmental signals, which is completely feasible for polymers with reversible linkages that can assume the thermodynamically most favored state, as defined by their local environment.

The prospect of proto-RNA is reason for optimism that we will yet discover RNA-like polymers with self-assembling properties that are superior to any currently known. The chemical space that still must be searched is vast, but it is not infinite. The systems discussed above and in Chapter 1 are just a few examples of polymerization systems that have provided clues as to what such polymers might look like – as well as, perhaps, to the origin of RNA.

8.2. References

1. Matteucci, M.D. and N. Bischofberger, *Sequence-Defined Oligonucleotides As Potential Therapeutics.* Annual Reports in Medicinal Chemistry, 1991. **26**: p. 287-296.

2. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* Journal of Theoretical Biology, 2000. **205**(4): p. 543-562.

3. Wiberg, K.B., K.M. Morgan, and H. Maltz, *Thermochemistry of carbonyl reactions. 6. A study of hydration equilibria.* J. Am. Chem. Soc., 1994. **116**: p. 11067-11077.

4. Weber, A.L., *The sugar model: catalysis by amines and amino acid products.* Orig. Life Evol. B., 2001. **31**: p. 71-86.

5. Bean, H.D., et al., *Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA.* Origins of Life and Evolution of the Biosphere, 2006. **36**(1): p. 39-63.

6. De Graaf, R.M. and A.W. Schwartz, *Thermal synthesis of nucleoside H-phosphonates under mild conditions.* Origins of Life and Evolution of the Biosphere, 2005. **35**(1): p. 1-10.

7. Sreenivasachary, N. and J.M. Lehn, *Gelation-driven component selection in the generation of constitutional dynamic hydrogels based on guanine-quartet formation.* Proc. Natl. Acad. Sci. USA, 2005. **102**(17): p. 5938-5943.

8. Sreenivasachary, N. and J.M. Lehn, *Structural selection in G-quartet-based hydrogels and controlled release of bioactive molecules.* Chemistry-an Asian Journal, 2008. **3**: p. 134-139.

9. Sreenivasachary, N., et al., *DyNAs: Constitutional dynamic nucleic acid analogues.* Chemistry-a European Journal, 2006. **12**(33): p. 8581-8588.

10. Cordier, P., et al., *Self-healing and thermoreversible rubber from supramolecular assembly.* Nature, 2008. **451**: p. 977-980.

11. Ura, Y., et al., *Self-assembling sequence-adaptive peptide nucleic acids.* Science, 2009. **325**(5936): p. 73-77.

12. Heemstra, J.M. and D.R. Liu, *Templated Synthesis of Peptide Nucleic Acids via Sequence-Selective Base-Filling Reactions.* Journal of the American Chemical Society, 2009. **131**(32): p. 11347-11349.