PREBIOTIC SYNTHESIS OF NUCLEIC ACIDS

A Dissertation
Presented to
The Academic Faculty

by

Heather D. Bean

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

Georgia Institute of Technology
April 2008

PREBIOTIC SYNTHESIS OF NUCLEIC ACIDS

Approved by:

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

Dr. James C. Powers
School of Chemistry & Biochemistry
Georgia Institute of Technology

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

Dr. David G. Lynn
Schools of Chemistry and Biology
Emory University

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

Dr. Ronald F. Fox
School of Physics
Georgia Institute of Technology

Date Approved: March 29, 2008

To my sisters.

Continue to dream big, work hard, and love what you do.

ACKNOWLEDGEMENTS

experiences of my graduate education. And finally, I thank Profs. Raffaele Saladino and Ernesto Di Mauro for allowing me the opportunity to travel to Italy to conduct research.

I wish to thank my wonderful family. First and foremost, I thank my Mom and Dad, Angela and David Bean, for their steadfast belief in me. The patience, guidance, and love with which they raised my sisters and me is a model I hope to emulate one day. Thanks to my Grandparents, Iva Lee and Bob Bean, and Margaret and Tom Gilbert, for their unwavering pride in their grandchildren. I wish to thank my sisters, Lori and Kimberly Bean, to whom this work is dedicated, for being the best girlfriends that only sisters can be. And I wish to thank Jason Ideker, who may not be related by blood, but is like a brother to me and who lent his ear over countless hours, empathizing with me over the experiences of being a graduate student.

Finally, I wish to thank Jordan Cannon for being brave enough and strong enough to live with me and love me during my temper tantrums, pity parties, and mood swings brought on by the stresses of the final year of graduate school, enabling me to finish my degree with most of my scruples intact. I look forward with great anticipation to the new adventures we will share during our upcoming move together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

*a*NA	Acetal nucleic acid
ANP	Amine-nucleoside polymers
BRSM	Based on recovered starting material
δ	chemical shift (ppm)
DNA	Deoxyribonucleic acid
dOM	1-O-Methyl-2-deoxy-D-ribose
dA	Deoxyadenosine
dC	Deoxycytidine
dG	Deoxyguanosine
ΔG	Change in Gibbs free energy (kcal/mol)
ΔH	Change in enthalpy (kcal/mol)
dT	Deoxythymidine
dT-*ga*-dT	Glyoxylate-linked deoxythymidine dinucleotide
G1	Aldehyde carbon of glyoxylic acid
G2	Carboxylic acid carbon of glyoxylic acid
*ga*NA	Glyoxylate-acetal nucleic acid
HMBC	Heteronuclear multiple bond correlation
HSQC	Heteronuclear single quantum coherence
HPLC	High performance liquid chromatography
ImpA	5′-phosphorimidazolide-adenosine
2-MeImpN	5′-phospho-2-methylimidazolide nucleotides
NMR	Nuclear magnetic resonance

*pa*NA	Pyruvate-acetal nucleic acid
PNA	Peptide nucleic acid
ppm	Parts per million
rA	Adenosine
rC	Cytidine
rG	Guanosine
rU	Uridine
rU-*ga*-rU	Glyoxylate-linked uridine dinucleotide
RNA	Ribonucleic acid
TEAA	Triethylammonium acetate
TNA	Threose nucleic acid

SUMMARY

The origin of the first RNA polymers is central to most current theories regarding the origin of life. However, difficulties associated with the prebiotic formation of RNA have lead many researchers to conclude that simpler polymers, or proto-RNAs, preceded RNA. These earlier polymers would have been replaced by RNA over the course of evolution. A remaining difficulty for this theory is that the *de novo* synthesis of a feasible proto-RNA has not yet been demonstrated by plausible prebiotic reactions. This thesis focuses on two problems associated with prebiotic proto-RNA synthesis: The formation of nucleosides and the necessity of reversible backbone linkages for error correction in nucleic acid polymers.

"The Nucleoside Problem", or the lack of success in forming pyrimidine nucleosides by plausible prebiotic reactions, represents a significant stumbling block to the RNA world hypothesis. Nearly four decades ago Orgel and coworkers demonstrated that the purine nucleosides adenosine and inosine are synthesized by heating and drying their respective bases and ribose in the presence of magnesium, but these reaction conditions do not yield the pyrimidine nucleosides uridine or cytidine from their respective bases. In this thesis a potential solution to The Nucleoside Problem is hypothesized based upon a proposed chemical mechanism for nucleoside formation. This hypothesis is supported by the successful synthesis of 2-pyrimidinone nucleosides by a plausible prebiotic reaction in good yield, demonstrating that pyrimidine nucleosides could have been available in the prebiotic chemical inventory, but that uridine and cytidine were likely not abundant.

Reversible backbone linkages are necessary to provide a mechanism for error correction in non-enzymatic template-directed syntheses of proto-RNAs. In this thesis, acetals are explored as low-energy, reversible linkage groups for nucleosides in polymers. The synthesis of glyoxylate-acetal nucleic acids (*ga*NAs) through simple heating-drying reactions from neutral aqueous solutions is demonstrated, and these linkages are shown to be hydrolytically stable under a considerable range of solution conditions. Computational models demonstrate that the glyoxylate linkage is an excellent electronic and isosteric replacement for phosphate. Molecular dynamics simulations also indicate that a *ga*NA duplex would have structural properties that closely match a phosphate-linked RNA helix, suggesting the possibility for cross-pairing between *ga*NAs and RNAs, allowing for sequence transfer and genetic continuity through the evolution from proto-RNAs to RNA.

The principles illustrated in this thesis by 2-pyrimidinone nucleoside and *ga*NA synthesis can be extended to other prebiotic condensation reactions. Factors affecting condensation yield, such as thermodynamics, kinetics, reactant solubility, and salt effects, are summarized herein.

CHAPTER 1

INTRODUCTION

1.1 THE RNA WORLD – A BRIDGE TO THE ORIGIN OF LIFE?

All known life forms on earth utilize an intricate and interdependent system of biopolymer synthesis and cellular function: genetic information is stored in DNA, which is transcribed into RNA that is then decoded by the ribosome and translated into protein. The universality of this system, the "Central Dogma" of molecular biology, has lead to the hypothesis that all extant biology is descended from a single organism, or pool of interrelated organisms, that utilized DNA for information storage and proteins as reaction catalysts. However, with respect to the origin of life, the hypothesis of a DNA/protein-based common ancestor presents a classic chicken-and-the-egg paradox, "Which came first, DNA or proteins?" DNA codes for protein synthesis, yet proteins are required for nucleic acid synthesis and replication. Proposed in the late 1960s, the RNA world hypothesis provided a solution – the first macromolecule was neither DNA nor protein, but rather RNA (*10-12*). As the intermediary molecule between information storage and protein synthesis, RNA is a likely candidate to be the progenitor of both DNA and protein. RNA is capable of storing genetic information in its base sequence, and catalyzing chemical reactions. In fact, examples of the catalytic role of RNA in extant biology are now known to range from self-splicing mRNA molecules to the ribosomal RNA nucleotides that directly catalyze the coupling of amino acids in a growing peptide chain (*1*).

The synthesis of RNA in extant biology, however, still relies upon the participation of proteins. The protein-free *de novo* synthesis of RNA in a prebiotic reaction has yet to be demonstrated despite several decades of effort (*1*). Many researchers have therefore concluded that RNA was not the first informational polymer of life. Rather, RNA was preceded by an RNA-*like* polymer, or several generations of polymers, termed proto-RNAs, that were structurally and functionally similar to RNA, but easier to assemble. Proto-RNA could have been comprised of different bases, sugars, and linking molecules that were assembled through more thermodynamically and kinetically accessible pathways. Without the constraints of requiring the current four RNA bases, ribose, and phosphate for the construction of an informational polymer, the possible composition of proto-RNAs seems limitless. However, the existence of putative monomer units in the prebiotic chemical inventory for the assembly of proto-RNA would have been dictated by astro- and geochemical processes, paring down the set of molecules from which nature could select.

The origin of life community is compiling an ever-growing list of compounds that are believed to have been available for prebiotic chemical reactions. Astronomers report the compositions of neighboring terrestrial bodies' atmospheres and surfaces, meteorites, comets, and the seemingly empty interstellar medium (*13-15*), which contain a diversity of molecules ranging from molecular hydrogen to ethanol to cyanodecapentayne ($HC_{11}N$) (*16*). Geologists and atmospheric chemists hypothesize about the composition of the early earth's atmosphere, oceans, and land based on mathematical models and minerals comprising the few rocks that have survived the metamorphosis of the planet (*17, 18*). Laboratory chemists use these molecules and proposed environmental conditions to guide

model prebiotic reactions to further expand the prebiotic chemical inventory (*19-26*). Although prebiotic syntheses for many biological building blocks have been demonstrated (e.g., ribose, several amino acids, and the RNA bases), a prebiotic methodology for coupling the RNA monomers into an oligonucleotide has not been demonstrated. Thus, there is currently a gap in our conceptual evolutionary timeline of life from the small molecule world (i.e., the prebiotic chemical inventory) to the appearance of the first proto-RNA molecule.

1.2 THE MISSING LINK

The gap in the prebiotic timeline can be attributed to two distinct challenges, selection and coupling. From the point of view of selection alone, the synthesis of RNA from the prebiotic chemical inventory is a formidable challenge. From the vast prebiotic chemical inventory that likely contained innumerable molecules resembling the building blocks of RNA, how were the Watson-Crick bases and ribose selected? (Figure 1.1) Even if a process existed by which the monomer units could be selected from a complex chemical mixture, how did these building blocks become coupled specifically into β-furanosyl nucleosides, phosphorylated, and incorporated into an RNA polymer in a 3′,5′ orientation so as to assemble extant RNA? (Figure 1.1) Although evoking the existence of a proto-RNA world helps relax some of the restrictions for selection and coupling, these two steps still have a profound influence on the synthesis of the original informational polymers. In the following sections we will elaborate the challenges of selection and coupling as they relate to *de novo* RNA synthesis in the prebiotic environment.

1.2.1 Nucleoside Formation

As discussed above, researchers have defined a varied and ever-growing prebiotic chemical inventory, from which adenine, guanine, cytosine, uracil and D-ribose would
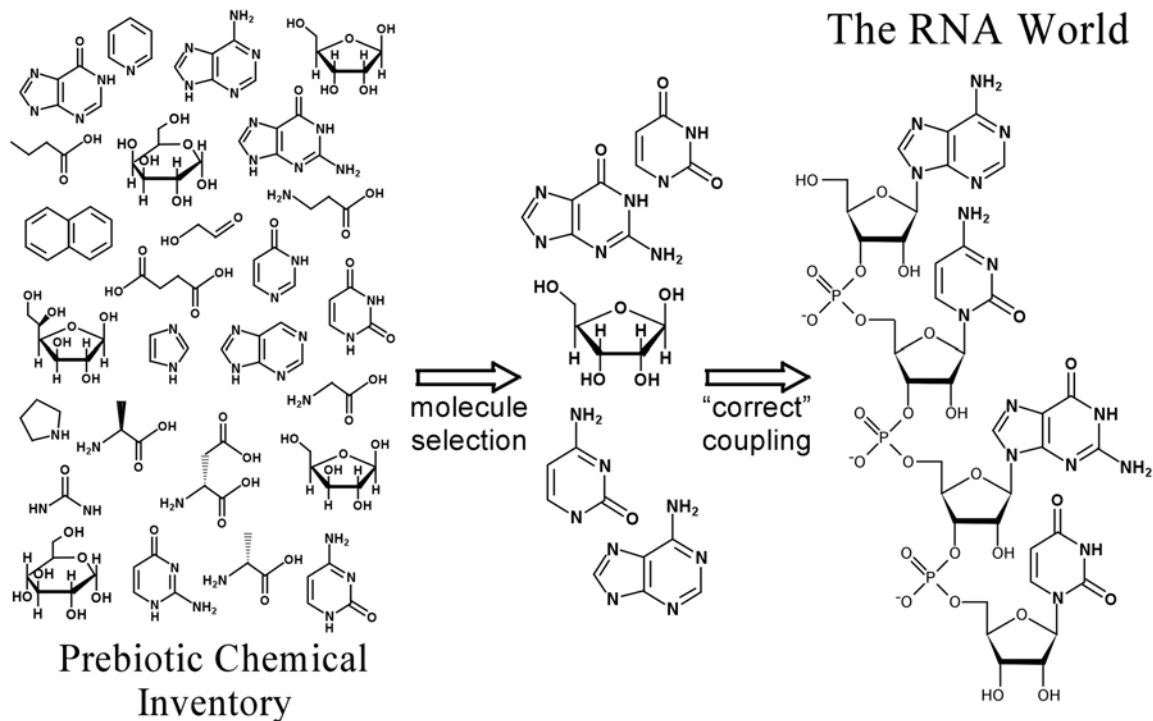
Figure 1.1: From the small molecule world to the RNA world. The two primary challenges for assembling RNA (or proto-RNA) from the prebiotic chemical inventory are the selection of the correct RNA building blocks (adenine, uracil, guanine, cytosine, and ribose), and coupling of these molecules in the correct way into nucleosides, nucleotides, and oligonucleotides.

have had to have been selected for RNA synthesis. On what criteria were these selections made? One possible answer is that the selection was primarily based on high concentrations of RNA precursors relative to similar molecules in the prebiotic environment. However, several lines of evidence suggest that this mode of selection was not applicable to prebiotic RNA synthesis, as demonstrated by the case of D-ribose. The formose reaction famously produces an incredible range of sugar products from

glyceraldehyde to hexoses, branched and straight-chain, with every possible combination of stereoisomer and handedness (D and L) observed in the mixture (*27*). It could be argued that the lack of selectivity in ribose production could have been overcome if ribose was significantly more stable than the other sugars produced by the formose or other reaction, thereby concentrating ribose in the mixture over time. However, the Miller laboratory demonstrated that the half-life of ribose in aqueous solution is within an order of magnitude of the half-lives of other aldopentoses and aldohexoses, and not even among the most stable (*28*). Therefore, it could not have been the concentration of ribose in solution, either through favored synthesis or accumulation through greater stability, that selected it for inclusion in RNA molecules.

The argument of selective synthesis in regards to the incorporation of β-furanosyl nucleosides as the sole building monomer unit in RNAs is also without support. When coupling bases and ribose together in a simple heating-drying reaction (*8, 29*), four nucleoside anomers are possible, β-furanosyl, β-pyranosyl, α-furanosyl, and α-pyranosyl ribonucleosides (Figure 1.2), and the β-furanosyl nucleoside is produced without selectivity (*8, 29-31, 7*). This phenomenon is not restricted to glycosides of ribose, but is common for any sugar that adopts multiple conformations in solution (*32-34, 31*). It has also been demonstrated that the distribution of nucleoside anomers produced in acidic aqueous solutions are kinetically controlled on short time scales, but will reach an equilibrium distribution that minimizes steric interactions between the base and sugar, which for aldopentoses are the α- and β-pyranosyl forms (*31*). Therefore, it would not have been the selective production of β-furanosyl nucleosides that influenced their incorporation into prebiotic nucleic acids.

Eschenmoser and co-workers have sought to understand if there is a structural or chemical advantage that β-furanosyl ribose imposes on RNA that selected it from other anomers of ribose and other sugars for the construction of the backbone. Their data show

β-furanosyl   α-furanosyl

β-pyranosyl   α-pyranosyl
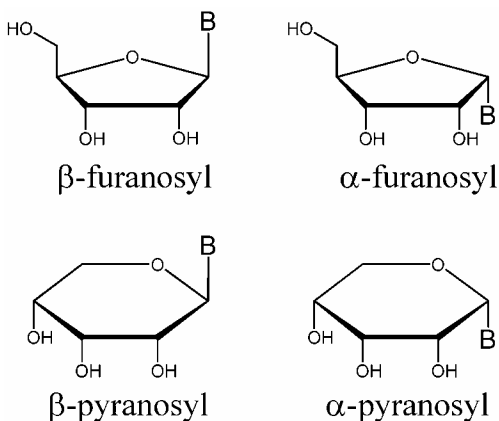
Figure 1.2: Four anomers of ribosyl nucleosides. "B" represents the nucleoside base (uracil, cytosine, adenine, or guanine in the case of RNA).

that there are several sugars (tetroses, pentoses, hexoses) and alternative conformations of ribose (e.g., β-pyranosyl) that are capable of forming double-stranded structures when incorporated into nucleic acid polymers; some of the modified duplexes are even more stable than the RNA duplex (*35-39*). Taken together, these data support the notion that the selective formation of a *ribo*nucleoside out of the prebiotic chemical inventory is not likely, and that proto-RNA backbones may have been constructed from a sugar (or a mixture of sugars) other than ribose.

1.2.2 Coupling and the Nucleoside Problem

In 1972, Fuller and Orgel reported the first prebiotic synthesis of ribonucleosides, coupling the preformed bases with ribose in a drying-heating reaction (*8*). This method of

synthesis was successful for the formation of adenosine and inosine from their concomitant purine bases, adenine and hypoxanthine, but did not work for the synthesis of uridine or cytidine (*8, 29*). Two years earlier Sanchez and Orgel had demonstrated the stepwise synthesis of cytidine monophosphate, beginning with 5′-phosphorylated ribose, followed by multiple steps, including chemical addition, UV anomerization, and hydrolysis (*40*). Despite three decades of experimentation by Orgel and others to find a more plausibly prebiotic route to uridine and cytidine, no simpler methods have arisen (*41*). The lack of pyrimidine nucleoside synthesis, known as The Nucleoside Problem, has even lead some scientists to hypothesize that pyrimidines were absent from the earliest nucleic acid polymers (*11, 12, 42-44*). In Chapter 2 a potential solution to The Nucleoside Problem is hypothesized, which has its roots in a proposed chemical mechanism that gives rise to purine nucleosides but inhibits uridine and cytidine synthesis.

1.2.3 The Watson-Crick Base Pairs

Selection of the nucleic acid bases (A, G, C, U) for prebiotic RNA synthesis also poses a barrier. Although the Watson-Crick base pairs are energetically favored within a duplex, individual bases will not form hydrogen bonds in aqueous solution, nor will their nucleosides or nucleotides, but rather will form columnar stacks (*45*). In aqueous solution stacking is thermodynamically favored over base pairing as it reduces the solvent-exposed surface area of the largely hydrophobic bases, and the hydrogen donor and acceptor groups that participate in base pairing can be satisfied through H-bonding with water (*45*). Despite the fact that Watson-Crick base pairs are essential to information transfer and storage in all organisms, and therefore were likely selected by biology's

earliest ancestors explicitly because of this pairing ability, the free bases do not form base pairs in solution. This realization elicits the *Paradox of Base Pairing* – how would the bases have been selected for inclusion in the first proto-RNA molecules for the purpose of Watson-Crick base pairing if the bases themselves did not form base pairs without having first been linked by a backbone.

Even when incorporated into a polymer, pairing of the four RNA bases is not restricted to the Watson-Crick faces. For example, the guanine-uracil wobble base pair is relatively stable compared to the adenine-uracil base pair and well represented in ribonucleic acid secondary structures, and non-Watson-Crick (e.g., Hoogsteen) pairing occurs in triplex and quadruplex nucleic acid structures. As with the synthesis of sugars, the prebiotic formation of the bases was most likely not limited to the two purines and two pyrimidines of RNA. Rather, a myriad of similar heterocyclic molecules were potentially available in the prebiotic chemical inventory (*30, 25*). Many examples exist of modified nucleic acid base pairs that function within a ribonucleic acid duplex, including the modified bases of tRNA, the size-expanded duplexes of Kool and co-workers, and even purine-purine duplexes (*46, 44, 47-49*). With such a diversity of nucleic acid base-like molecules with multiple pairing conformations available, it is difficult to imagine how the earliest proto-RNAs would have spontaneously incorporated bases that participate in the formation of Watson-Crick base pairs without there having first existed a mechanism that selected the bases that can form a specific pairing geometry from the myriad of molecules that cannot.

1.2.4 Nucleotide and Oligonucleotide Synthesis

The RNA backbone consists of repeating units of ribose and phosphate, linked through phosphodiester bonds. The advantages that phosphate imparts to RNA are numerous, not the least of which are hydrolytic protection, aqueous solubility, maintenance of an extended backbone conformation, and genotypic-phenotypic decoupling, all provided by the negative charge of phosphate at neutral pH (*50, 51*). However, the ubiquity of phosphate nucleotides in biology, not just as nucleic acid monomers, but also as the energy currency of the cell (i.e., ATP), derives from the juxtaposition of thermodynamic instability and kinetic stability of phosphoesters and pyrophosphates (*50*). Many possible nucleic acid linking molecules possess an acid group that could provide negative charges (*50, 6*), but it is the thermodynamic and kinetic properties that likely drove the evolutionary selection of phosphate, yet may have also excluded phosphates from early proto-RNA molecules.

The hypothesis that phosphate was not the linker in early predecessors of RNA is not new, and is partly driven by the lack of phosphate in available forms on the early earth (*52*). Orthophosphate, the predominant phosphorous-containing moiety in extant biology, is a limiting reagent for living organisms in the environment because of its low solubility. The vast majority of phosphate on the earth is mineralized, and the hypothesized atmospheric and oceanic conditions of the early earth would have also resulted in low abundances of available orthophosphate (*52*). Although recent studies have demonstrated the ability of phosphate minerals to phosphorylate nucleosides in formamide (*53*), and new hypotheses are arising for the availability of soluble reduced

phosphorous compounds on the early earth (*54*), phosphoester bond formation is still an enthalpically unfavorable process.

The thermodynamic impediment of phosphodiester formation that is imposed upon nucleotide polymerization and replication restricts the ability to test the hypotheses regarding the driving factors of this phenomenon (e.g., sequence effects, stacking interactions, surface adhesion). Therefore, many researchers choose to use one of a variety of available activating chemistries to drive phosphodiester bond formation, removing the activation step as the limiting factor of oligonucleotide synthesis or replication. However, the use of activating agents in phosphodiester bond formation leads to backbones with mixed 3′,5′ and 2′,5′ linkages, misincorporated (i.e., non-W-C pairing) bases, and truncation products resulting from strand cyclization (*55-57*). In the absence of an editing mechanism, such as the enzymatic processes involved in cellular DNA replication, these undesirable products represent kinetic traps. As a result, particularly in the case of cyclized products, the yield of long oligonucleotides is greatly affected. For example, Ferris and coworkers have observed that using 1-methylimidazole activated adenine mononucleotides for oligomerization on mineral surfaces results in a large proportion of cyclic dimer formation (*57*). And the problem of cyclization is not restricted to non-templated reactions and very short polymers (i.e., dinucleotides), but was also observed as a major side product in the carbodiimide-facilitated ligation of tiling half-complementary sixmer oligonucleotides (*56*). Not simply a nuisance in the lab, kinetically-trapped byproducts would also halt polymerization and replication of nucleotides in the prebiotic environment if activating agents were employed to promote phosphodiester, or other high-energy bonds.

Nucleotide sequence effects also present a significant barrier to non-enzymatic template directed replication of nucleic acids. As detailed above with the paradox of base pairing, hydrogen bonding of the template strand with the complementary monomer in water is not a sufficient driving force for monomers to associate with the primer and template strands of a replicating complex. Rather, monomer association is dominated by base stacking. As the pyrimidines have poor stacking interactions compared to those of purines, they are not efficiently incorporated into product strands during template directed synthesis. In fact, two or more sequential adenines in an oligonucleotide template strand will halt replication as the self-association constant for thymidine is too low for tandem incorporation into the product strand (*58*). Other problems such as the formation of G-quadruplexes by G-rich template strands, and non-Watson-Crick base pairings (e.g., G-U wobble pairing) further inhibit high-fidelity sequence transfer in template-directed synthesis. These problems, if not overcome by some mechanism, would greatly restrict the sequence space for evolving ribozymes and proto-ribozymes as newly mutated sequences that possess new catalytic properties may not be replicable (i.e., heritable).

Non-enzymatic template-directed synthesis is also plagued by the low association constants of monomer nucleotides or short oligonucleotide sequences for the template stand. Mineral surfaces have long been touted as a means of locally concentrating the monomers of biomolecules on the early earth (*59*), and they have been demonstrated to enhance the polymerization of activated nucleotide monophosphates (*60*). However, these mineral surfaces have never been demonstrated to be selective for the RNA bases, and would likely adsorb a range of compounds. The prebiotic chemical inventory is

hypothesized to have contained a wide variety nucleic acid base-like heterocycles of various sizes, shapes, and proton donor and acceptor patterns. The non-selective adsorption of these molecules onto a mineral surface, followed by their inclusion into a polymer, would likely result in an oligonucleotide that is unable to basepair, and therefore unable to replicate. Other characteristics of mineral-catalyzed polymerization, such as irreversible product strand binding and an oligonucleotide binding orientation that inhibits base pairing (i.e., replication), make mineral surfaces of questionable utility as promoters of oligonucleotide synthesis and replication in the prebiotic environment.

1.3 BRIDGING THE GAP

As outlined above, bridging the gap between the small molecule world and the RNA world (or proto-RNA world) requires a means of selecting and coupling base pairs into an oligonucleotide. Two complementary solutions have been proposed to address the quandaries of selection and coupling for the synthesis of proto-RNAs, 'molecular midwives' (*2, 5*) and reversible backbone linkages (*61, 2, 5*), which are elaborated below.

1.3.1 Molecular Midwives: Base Selection and Preorganization

It has been proposed that there was a template upon which the nucleic acid base pairs were preorganized prior to their incorporation into oligonucleotides to form proto-RNAs. These templates would have been heterocyclic molecules available in the prebiotic chemical inventory, produced by reactions and starting materials similar to those that made the nucleic bases in the small molecule world. The term 'molecular midwife' has been coined to describe any such molecule that facilitated the formation of proto-RNA, but was not covalently linked to the polymer (*2*). These midwife molecules would have aided in the 'birth' of a proto-RNA, but would have no longer been necessary

for RNA synthesis and replication once evolution had selected a superior means for carrying out these reactions (e.g., catalysis via ribozymes or protein enzymes).

One can imagine that a molecular midwife would select base-pairing molecules from the mixture of heterocyclic compounds by providing a nanometer scale template upon which the base pairs could stack in solution (Figure 1.3). The midwife molecules
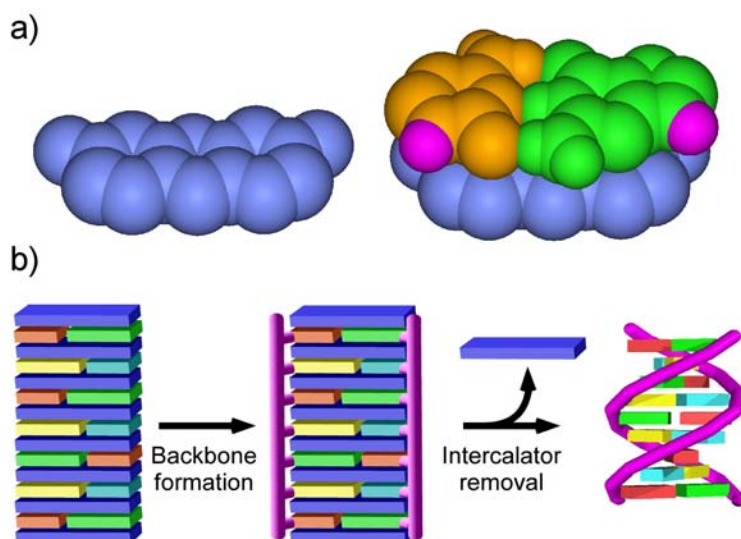
Figure 1.3: Schematic representation of the 'molecular midwife' hypothesis (*2*). a) A midwife molecule, proflavine (in blue) acts as a template for the formation of a Watson–Crick base pair. H-atoms of the nucleoside bases that will be replaced by C-atoms upon backbone formation are shown in magenta. b) In the case of molecular midwives with a greater association constant for base assembly than for self association, columnar stacks containing alternating midwife molecules and base assemblies spontaneously form under certain conditions. These columnar stacks preorganize the bases such that the introduction of a linkage chemistry leads to the formation of RNA-like polymers with a backbone length of 6.8 Å. Because the midwife molecules are only associated with the resulting polymers through non-covalent interactions, changes in solution conditions can lead to the removal of the intercalating midwives. Figure reproduced directly from (*5*).

would have been similar in size and shape to the small planar molecules known to intercalate the bases of extant RNA and DNA (*62*). Through the formation of a columnar stack of alternating midwives and base pairs, the bases would be locally concentrated and

oriented in the correct position for joining by the formation of a polymer backbone. As the bases would be spaced at 6.8 Å by the intervening midwife molecules, the backbone that stitched the bases together would resemble the length of the extant RNA backbone, allowing the nascent proto-RNA polymer to form a helical duplex upon removal of the intercalating midwife molecules, and would allow the single strands of proto-RNA the flexibility to form folded structures that are a common characteristic of ribozymes. Additionally, the selection and organization of the bases as base pairs ensures that the proto-RNA polymer can be replicated via the synthesis of a complementary strand, and also ensures the synthesis of a duplex of homogeneous width, required for duplex stability (*47*). Replication facilitated by molecular midwives would also alleviate some of the template sequence effects that plague non-enzymatic template directed replication, such as the inability to replicate purine-rich sequences due to the low self-association constants of pyrimidine nucleotides. As the midwife molecules are not covalently linked to the newly synthesized polymers, their dissociation (and subsequent reassociation for a new synthesis/replication cycle) is easily facilitated through changes in solution conditions, such as concentration, ionic strength, or temperature (Figure 1.3).

The association constants of intercalators for DNA or RNA generally range between $10^5 – 10^6$ M^{-1} (*63*). Therefore, if the concentration of intercalator were increased above approximately 10 μM in the presence of nucleic acid bases, the equilibrium amount of the bases that are assembled with intercalators (versus free in solution) would increase, as predicted by the law of mass action. Templating could occur not only for duplexes, but also for triplexes or quadruplexes, given a molecular midwife of the appropriate size and shape (Figure 1.4). The Hud laboratory has demonstrated the power
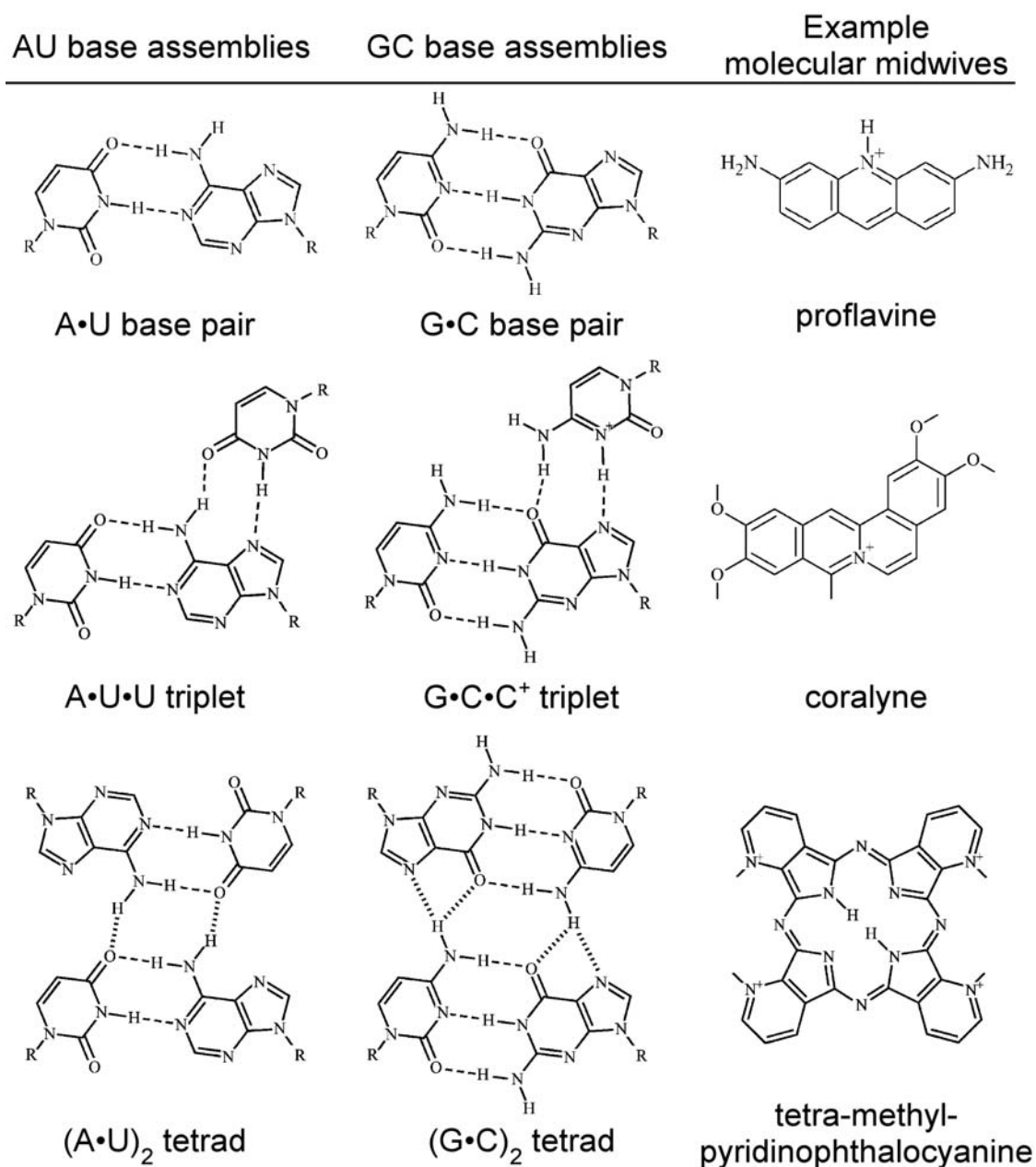
Figure 1.4: Nucleoside base assemblies and heterocyclic molecules with similar shapes. Proflavine and coralyne have already been shown to act as midwifes in ligation assays that include duplex or triplex formation between the template and substrate strands, respectively. The phthalocyanine analogue with a similar shape to the base tetrads is shown for illustrative purposes, as a molecular midwife has not yet been demonstrated for a tetrad system. Figure reproduced directly from (*5*).

of intercalators to promote duplex- and triplex-based template-directed synthesis in previous reports (*64, 5*). In these works, proflavine and coralyne were shown to facilitate up to a 1000-fold enhancement in the coupling of dT_3 and dT_4 phosphorothioate substrate strands into a dT_7 product in the presence of an existing template strand (proflavine, Figure 1.5a) or an existing hairpin template (coralyne, Figure 1.5b), respectively.
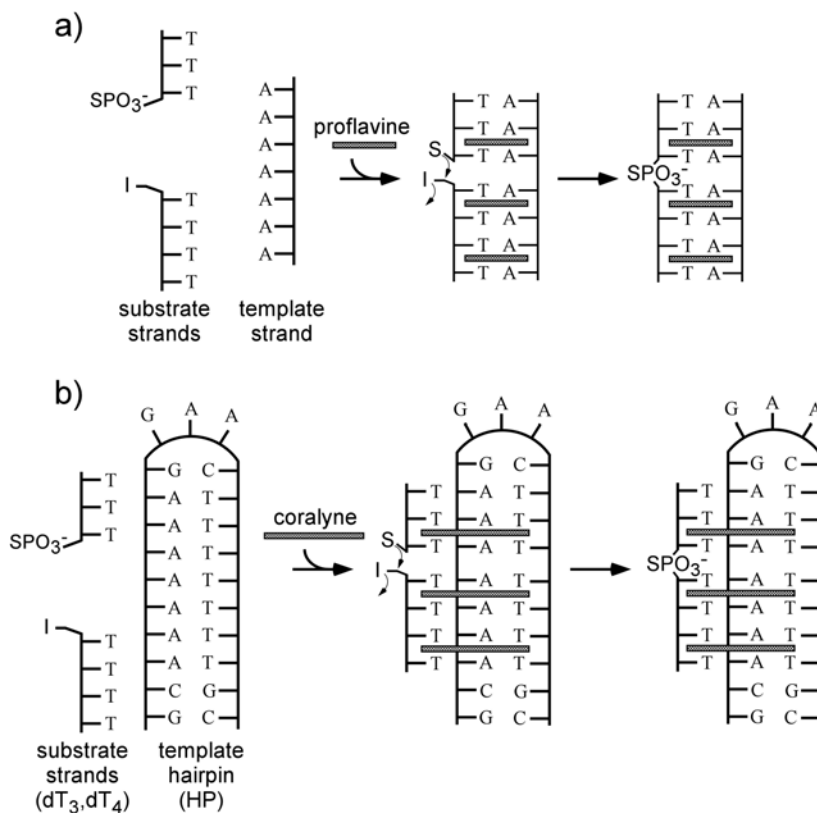
Figure 1.5: Experimental system used to explore the ability of intercalators to act as midwifes in a template directed ligation reaction including Watson–Crick base pairs or base triples. a) 3′-Phosphorothioate-$(dT)_3$ and 5′-iodo-$(dT)_4$ substrate strands are too short to form a stable complex with the $(dA)_7$ template strand. The addition of proflavine promotes the assembly of a ligation-active complex, which leads to the formation of the $(dT)_7$ product within an internal phosphorothioate linkage. b) Coralyne facilitates the association of 3′-Phosphorothioate-$(dT)_3$ and 5′-iodo-$(dT)_4$ substrate strands with a hairpin template to form a triplex ligation-active complex. Proflavine, which is specific for binding duplexes, is not active in promoting ligation in the presence of the hairpin template (*5*). Figure adapted from (*5*).

Assembly of the substrates strands on the template oligonucleotides was also shown to be shape-specific, as proflavine, with the shape of a Watson-Crick base pair is able to promote template-directed synthesis in a duplex-based system but not in a triplex-based system, while the opposite is true for coralyne, which has the shape of a pyrimidine triplet (Figure 1.4) (*5*). Although the coupling chemistry used in these ligation experiments is not considered to be prebiotically plausible, the results demonstrate that the assembly of an otherwise unstable template and substrate complex is promoted by intercalation, and that backbone coupling is possible in an intercalation-stabilized assembly.

Intercalators could have also provided selective pressure toward the formation of a more uniform backbone. As discussed above, non-enzymatic template directed ligations produce a product strand of mixed 2′,5′ and 3′,5′ linkages, often with an enrichment in 2′,5′ linkages. Recent work in the Hud laboratory has demonstrated that proflavine has a 25-fold higher affinity for 2′,5′-linked versus 3′,5′-linked RNA (*65*). However, other intercalators, such as ethidium bromide, favor the 3′,5′-linked RNA duplexes (*65*). It can therefore be postulated that if a given midwife molecule has a preferential association with a 3′,5′-linked RNA backbone, then it might also promote the formation of this backbone linkage over the 2′,5′-linked alternative, particularly if reversible backbone linkage chemistries are employed (elaborated below).

1.3.2 Thermodynamically Controlled Polymerization

Although molecular midwives could have facilitated the selection of the bases and the backbone length for the first proto-RNA polymers, the thermodynamic barriers to the formation of an RNA phosphodiester backbone still represent a formidable challenge to

prebiotic synthesis. Artificial chemical activation to promote the formation of phosphodiester bonds, such as 1-methylimidazole or pyrophosphate activation of phosphates, results in low replication fidelity due to strand cyclization, base misincorporation, and premature product chain termination (*55-57*). These difficulties can be circumvented by using low energy, reversible bonds for coupling nucleosides, facilitating the selection of the most thermodynamically-favored base pairing and backbone conformation.

Reversible bond formation is utilized by extant biology during the first step of replication as a primary means of proofreading via the thermodynamic selection of the lowest energy base pair (i.e., correct Watson-Crick complement) (Figure 1.6). The reversible condensation of the nucleoside triphosphate is followed by the hydrolysis of pyrophosphate, kinetically trapping the newly formed phosphodiester bond (Figure 1.6). In emulating this two-step process through the use of low energy bonds, it has been
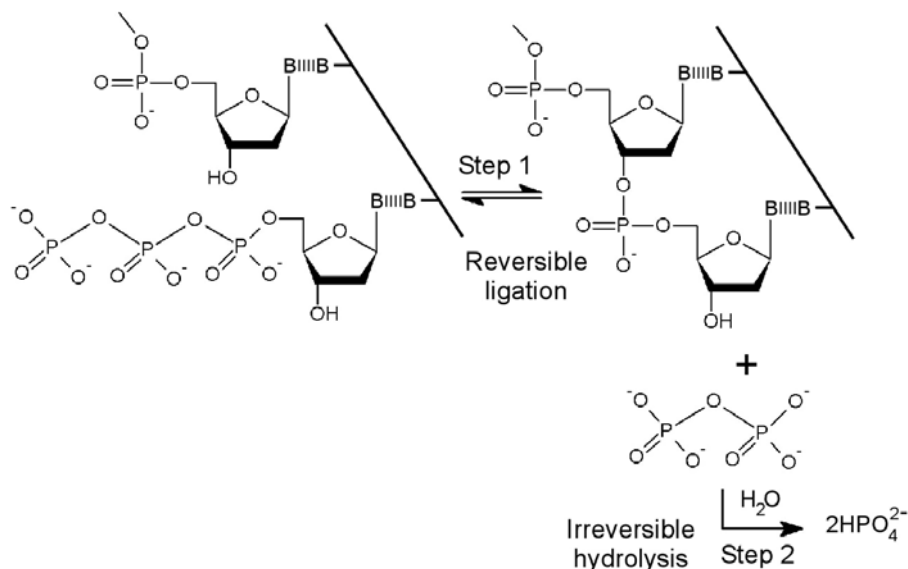
Figure 1.6: The biological synthesis of oligonucleotides proceeds via two steps, the first a reversible ligation of the nucleotide triphosphate, the second an irreversible hydrolysis of pyrophosphate. Figure adapted from (*5*).

18

hypothesized that both the initial step that exploits the stability of the template association, and the subsequent reaction that traps the thermodynamically-favored product are necessary for high-fidelity replication *(5)*.

1.3.2.1 Reductive Amination as a Test Case for the Utility of Reversible Backbone Linkages

Lynn and co-workers have developed a nucleoside ligation chemistry, reductive amination of amines and aldehydes, which maintains the functional features of the two-step reversible coupling and irreversible trapping in a template-directed synthesis (Figure
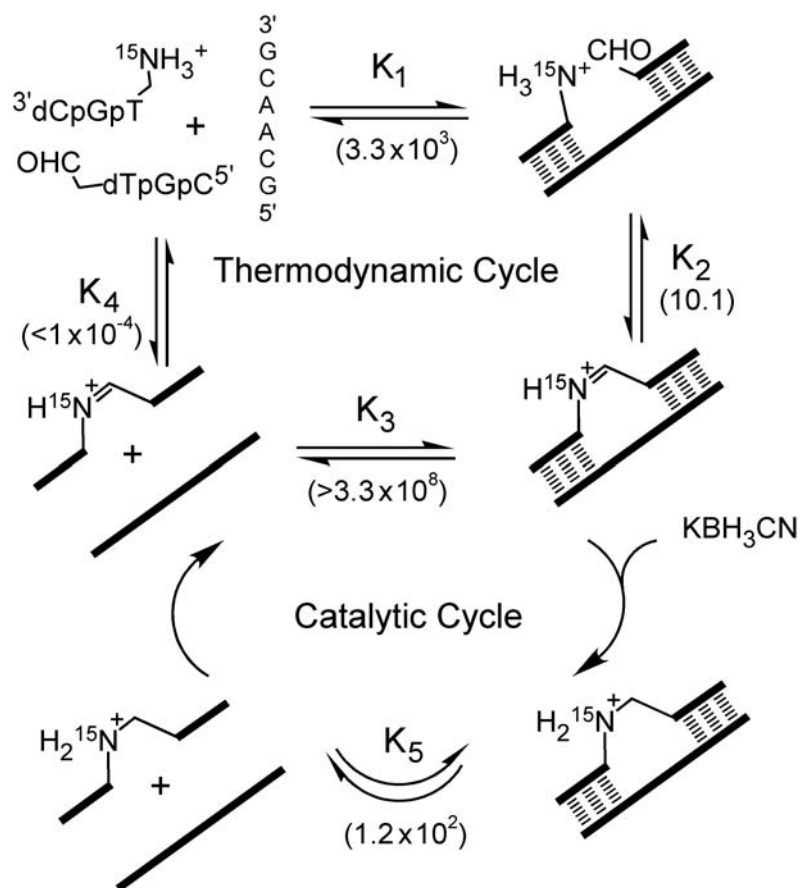
Figure 1.7: Catalytic activity of a simple DNA template, d(GCAACG), on the reductive amination of 5′-NH$_2$-d(TGC) and d(CGT)-3′-CH$_2$CHO. The equilibrium constants were determined by NMR. Figure reproduced directly from (*5*).

1.7). The first step of the ligation is the formation of an imine, a reversible reaction as indicated in the top panel of Figure 1.7. Though reversible, the duplex formed by the association of the imine-linked polymer with the DNA template is highly favored (K_3, Figure 1.7) (*66-68*). Surprisingly, only imine, and not the hydrated hemi-aminal intermediate, was observed in the associated complex contributing to the equilibrium. Therefore, even in water, where hydration is expected to contribute significantly to the equilibrium, the imine more significantly stabilizes the duplex over the hemi-aminal and functions as a good isosteric replacement of the phosphodiester bond.

The second, irreversible step of bond formation is the chemical reduction of the imine to the amine to form amine-nucleoside polymers (ANPs) (lower panel in Figure 1.7). This simple chemical change reduces duplex stability by a factor of nearly 10^6, which is attributed to the increased flexibility in the amine backbone arising from the removal of two conformationally restricted sp^2-hybridized centers (N and C) (*66, 67, 69*). This change in stability dramatically decreases product inhibition, which generally plagues template-directed oligonucleotide synthesis. Here, the product concentration must reach a 10^6-fold excess over the substrate concentration before it would significantly inhibit the ligation reaction by competing for catalytic sites along the template. More generally, these studies clearly establish that alternative backbone linkages can dramatically regulate the thermodynamics of template association, and this thermodynamic control can be exploited for catalytic cycles involving template-directed synthesis (Figure 1.7).

In order to demonstrate the utility of the imine/amine chemistry in a non-enzymatic template directed replication, Lynn and coworkers prepared mononucleotide

and dinucleotide substrates with a 5′ amine and a 3′ aldehyde moiety (Figure 1.8). A general characteristic of bifunctional monomers is the intramolecular reaction to produce cyclic nucleotides and small cyclic oligonucleotides, severely reducing the active substrate concentration and diminishing product formation. While the T_1, T_NT and A_NT nucleotides can form intramolecular imines, no reaction was detected under mild reducing conditions in an aqueous environment, an observation attributed to the low imine concentration in the absence of a template (*70*). In the presence of a template, these monomers condense to yield sequence and chain-length specific products.
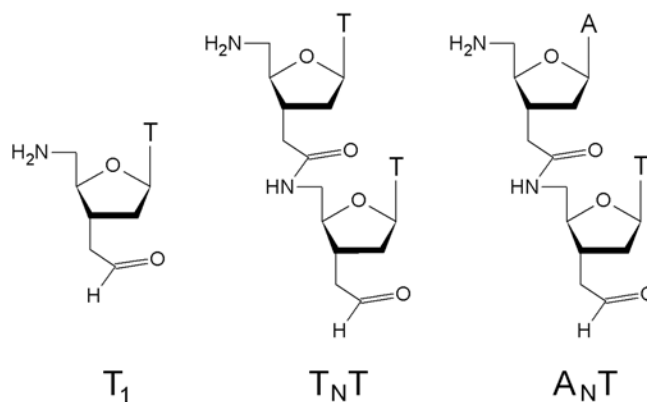
Figure 1.8: Bifunctional amine-nucleotide monomers used in the non-enzymatic template-directed replication of DNA templates.

As demonstrated by the product distribution of template-directed reductive polymerization of T_1 on a $(dAp)_8$ template in the presence of sodium cyanoborohydride, the polymerization proceeds through a step-growth mechanism (Figure 1.9). This polymer growth is a result of the reduced affinity of the amine products for the template strand versus the reactants or imine intermediates (Figure 1.7). That is, as the association constant for the imine dimer of T_1 and the template is significantly greater than that of the product T_2, the rate of monomer reduction is greater than dimer, which again is greater
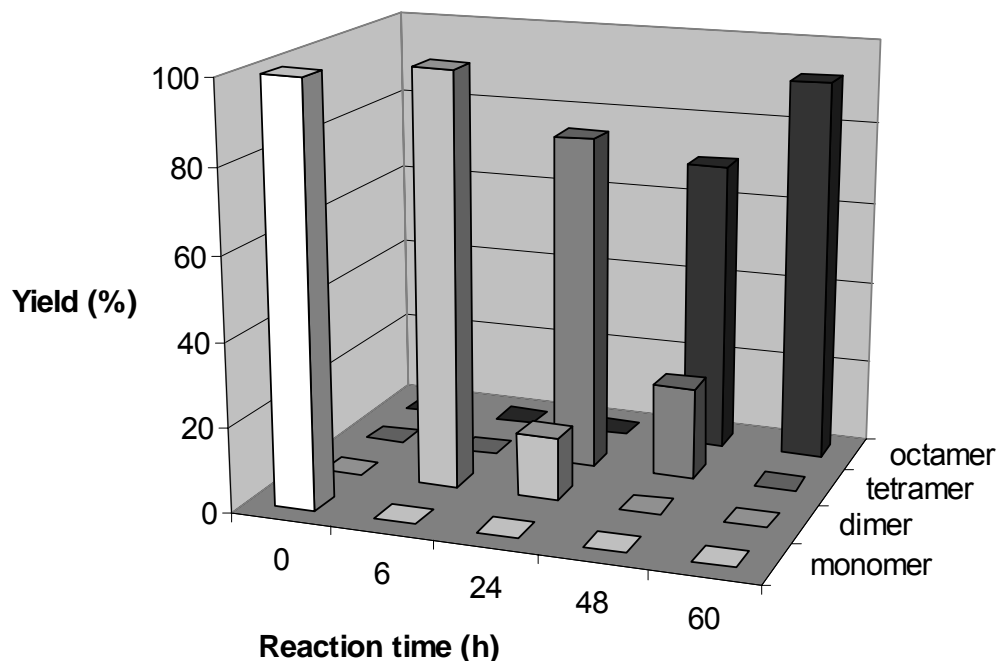
Figure 1.9: Product distribution upon reductive polymerization of T_1 on a $(dAp)_8$ template. The reactions contained 20 mol-equivalents of $NaBH_3CN$ in H_2O at 24 °C. At 48h additional reducing agent was added to complete the reaction. Figure adapted from (*5*).

than tetramer. Once the monomer supply is exhausted, T_2-T_2 condensation proceeds again at a higher rate than T_2-T_4 or T_4-T_4 couplings, and so on. The result is that no T_3, T_5, T_6, or T_7 intermediates are observed, and chain-length specific products are produced in high yield (Figure 1.9).

The equilibrium control of replication inherent to the imine/amine system not only enables the synthesis of full-length products, but also favors high-fidelity sequence translation. To demonstrate this, a mixed sequence 32mer DNA template strand, 5′-dAAAAAT(AAAAAAAT)$_3$AA-3′, was replicated using T_1 and the amide-linked dimers T_NT and A_NT (Figure 1.8). The 32mer ANP complementary strand was synthesized as the sole product (>99%). Whether using a combination of T_1 and A_NT, or T_NT and A_NT

as the substrates, no truncation products resulting from strand cyclization, nor misincorporated bases were detected by HPLC or mass spectrometry. There is now evidence that the transfer of sequence information can also be achieved in the reverse direction, from ANPs to DNA (*69*). These data clearly establish that the use of such two-stage coupling chemistries can achieve high-fidelity information transfer and template-defined chain length products, thereby eliminating cyclized side products. This example underscores the utility of equilibrium-controlled polymerization through reversible two-step non-enzymatic template-directed syntheses.

The ANP backbone is but one example of a reversible two-step ligation chemistry, and other backbone chemistries that facilitate equilibrium control of polymerization should be explored, which is a primary objective of this thesis. Chapter 3 describes the development and characterization of acetals as low-energy nucleotide backbone linkers. The data presented in this chapter establish that dinucleotides linked by glyoxylate (glyoxylate-acetal nucleic acids, or *ga*NAs) can be synthesized through simple heating-drying reactions from neutral aqueous solutions containing metal chlorides. The *ga*NA acetal backbone is demonstrated to be stable to hydrolysis under a range of aqueous conditions and is calculated to be a good isosteric replacements to the phosphodiester bond.

1.4 CONDENSATION REACTIONS AND BIOPOLYMERS

Condensation reactions are the hallmark of biopolymer formation. The growth of a polypeptide chain occurs with the release of a water molecule with each amino acid monomer addition. Polysaccharide synthesis from sugar monomers is also a condensation reaction. The synthesis of RNA polymers from free bases, ribose, and phosphate releases

water at three stages – at the synthesis of nucleosides from bases and ribose, the formation of nucleotides by the addition of a phosphate to a nucleoside, and with the coupling of nucleotide units into a nucleic acid polymer. It may seem contradictory that condensation reactions are utilized in the construction of these macromolecules, which constitute the majority of organic matter on earth, since life exists in an aqueous environment where depolymerization via hydrolysis would be energetically favored. However, hydrolysis is the cornerstone of ATP metabolism, the energy currency of the cell, and it also facilitates the recovery and reuse of amino acid, sugar, and nucleotide monomers from the degradation of biopolymers. Therefore, hydrolysis enables biopolymer formation to be reversible, which in the case of nucleic acid synthesis and replication, can allow for error correction through thermodynamic control.

Because of the tight coupling of condensation reactions and biopolymer synthesis, condensation is considered a prebiotically plausible process. In cells, condensation reactions are promoted through the aid of protein enzymes. In prebiotic processes, in the absence of enzymes, condensation can be achieved by the physical removal of water from the reaction medium, i.e., through drying. In the following chapters heating-drying reactions are used to explore potential solutions to two of the problems of prebiotic RNA synthesis. First, The Nucleoside Problem is addressed through the prebiotic synthesis of zebularine, a pyrimidine nucleoside (Chapter 2), followed by the synthesis of an energetically-favored alternative to the phosphodiester backbone, *ga*NAs (Chapter 3).

CHAPTER 2

ADDRESSING THE NUCLEOSIDE PROBLEM WITH

ZEBULARINE[†]

2.1 INTRODUCTION

Glycosylation of pyrimidine bases to give nucleosides by a plausible prebiotic reaction has not been achieved, and this represents a serious challenge to the "RNA world" and to many of its proposed precursors (*43*). Three decades ago, Orgel and co-workers demonstrated that adenine and hypoxanthine form glycosidic bonds with D-ribose when dried and heated together, although adenine gives mainly the product of attack at the NH_2 group (*8*). As much as 20% of the nucleoside was reported from hypoxanthine (BRSM) (*8*). However, neither cytosine nor uracil gives rise to nucleosides under these conditions, nor does guanine, a fact attributed to its low solubility (*43*). Some researchers have taken to exploring completely different pathways to prebiotic pyrimidine nucleoside formation, including the construction of the cytosine base on a sugar phosphate (*41*). Suggestions that the original bases may have differed from the canonical AUCG leave open the possibility that some of these bases might form glycosides under prebiotic conditions (*30, 71, 2, 48, 49*). Indeed, urazole, a five-membered ring derived from hydrazine, readily forms glycosides (Figure 2.1) (*30*). In

search of alternative pyrimidine bases, we have explored the condensation of ribose with 2-pyrimidinone (Figure 2.2) by both experiment and calculation.
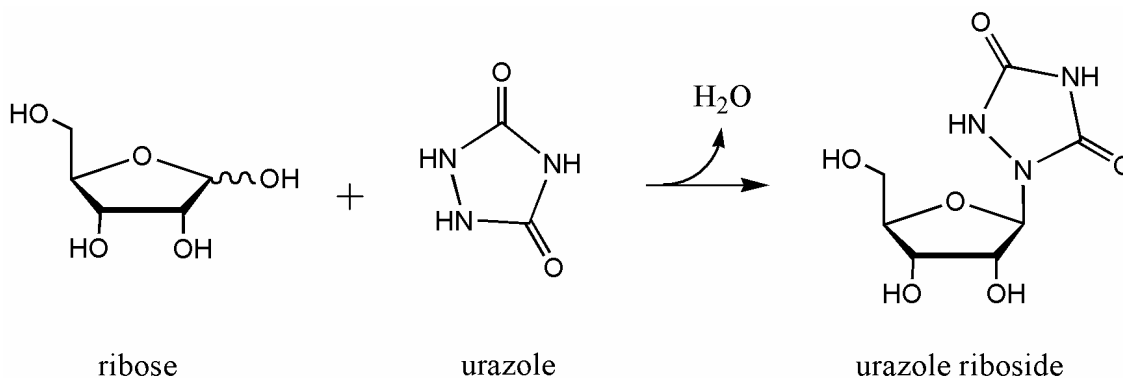
Figure 2.1: Urazole condenses with ribose to form a pyrimidine-like nucleoside (*30*).
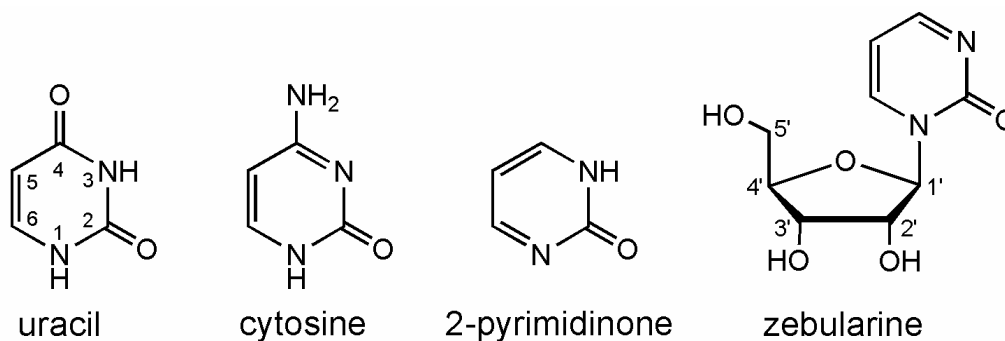
Figure 2.2: Chemical structures of pyrimidine bases and zebularine, the β-furanosyl ribonucleoside of 2-pyrimidinone.

2.2 MATERIALS AND METHODS

2.2.1 Prebiotic Syntheses of Purine and Pyrimidine Nucleosides

β-ribonucleoside synthesis with adenine (adenosine), hypoxanthine (inosine), uracil (uridine), and 2-pyrimidinone (zebularine) was carried out using three reaction conditions: pH 2.1, pH 6.3, and the Orgel reaction conditions (i.e., 50 mM $MgSO_4$ and 25 mM $MgCl_2$). The pH 2.1 reaction contained 5 mM base and 10 mM D-ribose in 200 μl,

with pH adjustments made by titration with HCl. The pH 6.3 reaction contained 5 mM base, 10 mM D-ribose, and 50 mM sodium cacodylate in 200 μl. The Orgel reaction contained 5 mM base, 75 mM D-ribose, 50 mM MgSO$_4$, and 25 mM MgCl$_2$ (pH 6, unbuffered), as described by Fuller and Orgel (8), in 200 μl. Sample pH was determined using a standard glass combination electrode. All reactions were carried out in triplicate by drying the reaction mixtures on 2″ watch glasses in an oven at 75 °C for 2 h, except for the Orgel reaction conditions, for which samples were dried at 100 °C for 2.25 h. After cooling to room temperature, all samples were resuspended in 200 μl dH$_2$O and analyzed by HPLC (see below) at the wavelength of maximum absorbance for each base (i.e., adenine, 260 nm; hypoxanthine, 249 nm; uracil, 259 nm; and 2-pyrimidinone, 303 nm). Identification of the β-furanosyl nucleoside of each base was made by co-injection onto the HPLC of the authentic standard with the reaction products. Nucleoside formation results, reported as percent yield, were calculated using Equation 2.4. Further

$$I_N \sim bc_N\varepsilon_N \qquad \text{Equation 2.1}$$
$$I_B \sim bc_B\varepsilon_B \qquad \text{Equation 2.2}$$

$$\% \text{ Yield} = \frac{c_N}{c_B + c_N} \times 100 \qquad \text{Equation 2.3}$$

$$\% \text{ Yield} = \frac{\dfrac{I_N}{\varepsilon_N}}{\dfrac{I_B}{\varepsilon_B} + \dfrac{I_N}{\varepsilon_N}} \times 100 \qquad \text{Equation 2.4}$$

N = β-ribofuranosyl nucleoside
B = base
I = integrated area of HPLC peak
b = pathlength
c = concentration of analyte
ε = molar absorptivity coefficient

confirmation of the β-furanosyl nucleoside of 2-pyrimidinone was made by NMR (see Section 2.2.3). NMR analysis revealed two species with the same relative retention times in the HPLC, and these two species were quantified together from the HPLC. The yield of the β-furanosyl nucleoside in the HPLC peak was then adjusted according to the ratio of the two species as determined by NMR.

2.2.2 Degradation of Purine and Pyrimidine Nucleosides

Studies of nucleoside glycosidic bond cleavage were performed under the same reaction conditions as given above for nucleoside synthesis, except with the base and ribose in each solution being replaced with 5 mM nucleoside (adenosine, inosine, uridine, or zebularine). Degradation samples were also dried at the same temperature and for the same durations as the synthesis reactions. After drying, samples were resuspended in 200 μl dH$_2$O and analyzed by HPLC (see Section 2.2.3). Results are reported as percent nucleoside degradation, where the integral of the free base is divided by the sum of the integrals for the free base and the nucleoside, with adjustments made for the difference in molar extinction coefficients for the free base and its nucleoside.

2.2.3 Mass Spectrometry, HPLC, and NMR Spectroscopy

The mass spectrum of the 2-pyrimidinone nucleoside reaction products was collected in negative ion mode on a Micromass Quattro LC mass spectrometer and ionized under electrospray conditions (capillary voltage, 2.5 kV; cone voltage, 40 V).

HPLC separation of nucleoside products was achieved by either isocratic or gradient elutions. A Phenomenex ODS analytical column (4.6 × 250 mm, 5 μ) was used for both methods. The separation of pyrimidine nucleoside products was achieved using an isocratic flow of 0.1 M triethylammonium acetate (TEAA) at 1.0 ml/min. Separation

of the purine nucleosides required a gradient of 10–30% B (0–5 min), 30–50% B (5–30 min), flow rate = 1.0 ml/min, where mobile phase A = 0.1 M TEAA, and B = 25% (v/v) acetonitrile in 0.1 M TEAA. The β-furanose ribonucleoside product of each synthesis was identified in the HPLC chromatograph by co-injection of an authentic standard (Sigma-Aldrich) with the nucleoside product mixture. The 2-pyrimidinone nucleoside samples used in the NMR study were purified by the isocratic method on a Phenomenex ODS semipreparative column (10 × 250 mm, 5 µ). NMR spectra of the 2-pyrimidinone nucleosides were recorded at 500 MHz in D_2O at 25 °C.

2.3 RESULTS AND DISCUSSION

2.3.1 Optimization of Acid-Catalyzed Reaction Conditions

The optimal drying temperature and relative concentrations of 2-pyrimidinone and ribose for the zebularine formation reaction were determined by a series of experiments in which each of several variables were systematically changed with respect to the other. The 2-pyrimidinone:ribose ratios investigated were 1:1, 1:2, and 1:4. The drying temperatures tested were 65, 75, 85, and 95 °C. Zebularine yields were quantified as described by Equation 2.4. Yields are shown in Figure 2.3 for a series of samples at pH 2.3 for which temperature and base:ribose ratios were varied. Based upon these results, a two-fold excess of ribose to base was selected for additional reactions, as higher ratios did not result in a substantial increase in zebularine yield. These data also revealed an increase in zebularine yield with temperature, with the greatest increase (300%) observed between 65 and 75 °C. However, the variability between replicates also increased significantly above 75 °C, possibly due to a greater amount of insoluble material after

drying at these higher temperatures. As a result, 75 °C was chosen as the drying

temperature for the 2-pyrimidinone nucleoside reaction.
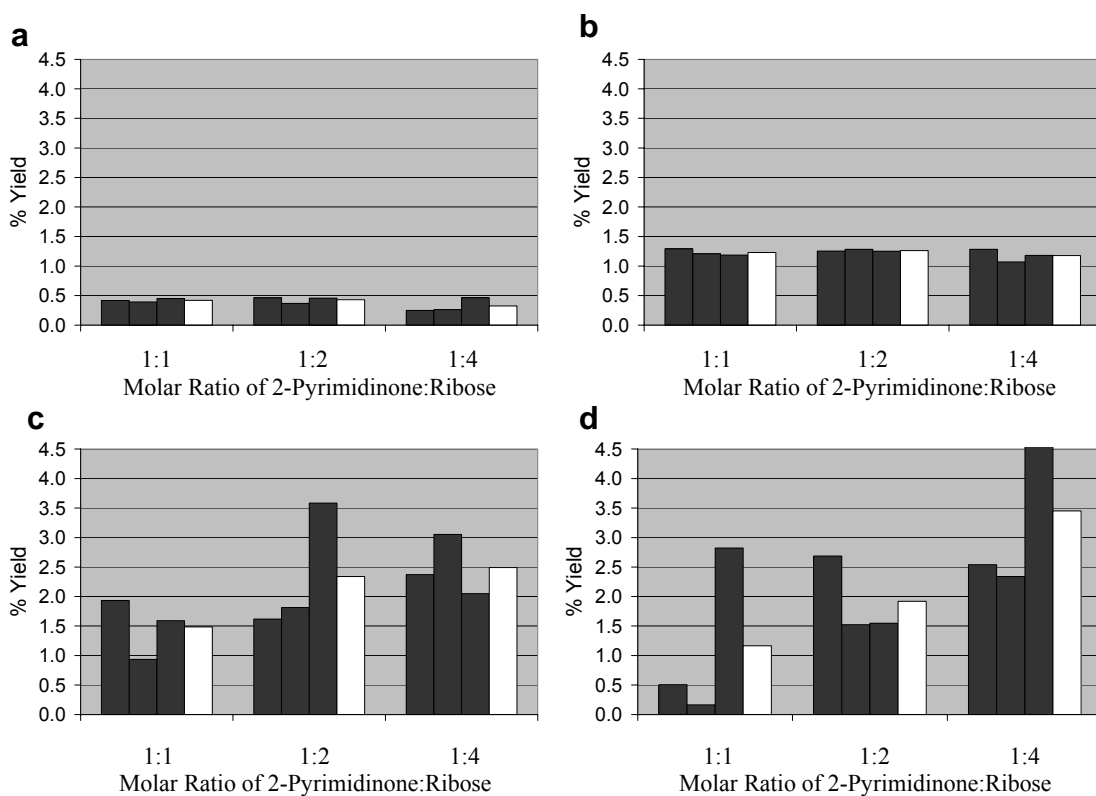
Figure 2.3: Percent zebularine yield as a function of 2-pyrimidinone:ribose ratio and drying/reaction temperature. a) Reaction mixtures dried at 65 °C. b) Reaction mixtures dried at 75 °C. c) Reaction mixtures dried at 85 °C. d) Reaction mixtures dried at 95 °C. Each experiment was performed in three replicates. All samples, which contained 5 mM 2-pyrimidinone and 5 – 20 mM ribose at pH 2.3, were dried from 200 μl for 2 h on a 2″ watch glass, resuspended in 200 μl of dH$_2$O, and analyzed by HPLC as described above. Dark bars represent yields of individual sample replicates, white bars are averages of the three replicates.

2.3.2 Characterization of the 2-Pyrimidinone Nucleoside Reaction Products

Fuller and Orgel demonstrated that nucleoside formation by purines and ribose

under drying conditions is an acid-catalyzed reaction that can also be promoted by

divalent metal ion salts at neutral pH (*8*). 2-pyrimidinone and ribose were heated and

dried under neutral conditions in the presence of magnesium salts (Orgel's reaction

conditions) and from a slightly acidic solution of pH 2.1. The actual pH at reaction is not known, but is likely to be less than 1 because of the drying conditions. 2-pyrimidinone nucleosides were formed under both reaction conditions, as indicated by the appearance of new peaks in the HPLC (Figure 2.4). Buffering the sample pH to 6.3 reduces 2-pyrimidinone nucleoside production to an unobservable level, although other reaction products are still observed (Figure 2.4). Thus, our results indicate that the zebularine formation is similar to the purine glycosylation reaction in that it is both acid-catalyzed and is promoted by Mg^{2+} near neutral pH.
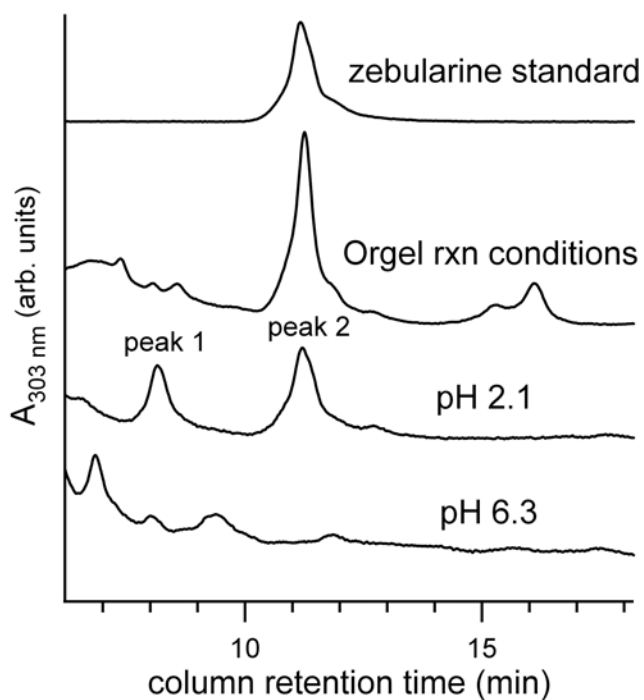
Figure 2.4: HPLC chromatographs of an authentic zebularine standard with products of the drying-heating reaction of 2-pyrimidinone with ribose. The Orgel reaction conditions: fifteen-fold excess of ribose, 25 mM $MgCl_2$, 50 mM $MgSO_4$, pH 6.2, dried and heated at 100 °C for 2.25 h. For the two other chromatographs shown: twofold excess of ribose, dried and heated at 75 °C for 2 h. All reactions were initially 5 mM in 2-pyrimidinone. The α-furanosyl nucleoside elutes as peak 1, the two β-anomers co-elute as peak 2. The shoulder peak observed in the zebularine standard is an artifact of the HPLC gradient.

The products of the 2-pyrimidinoneribose drying-heating reaction at pH 2.1 were further characterized by mass spectrometry, UV-vis spectrophotometry, and NMR spectroscopy. For mass spectrometry, the reaction mixture was dried as described in Section 2.2.1, and then resuspended in 200 μl MeOH. The ESI- spectrum shows two peaks: m/z = 113, 227 (Figure 2.5). The higher mass peak matches the nominal mass of a 2-pyrimidinone nucleoside minus one proton. The lower mass peak is tentatively assigned to the hydrate of the unreacted 2-pyrimidinone base.
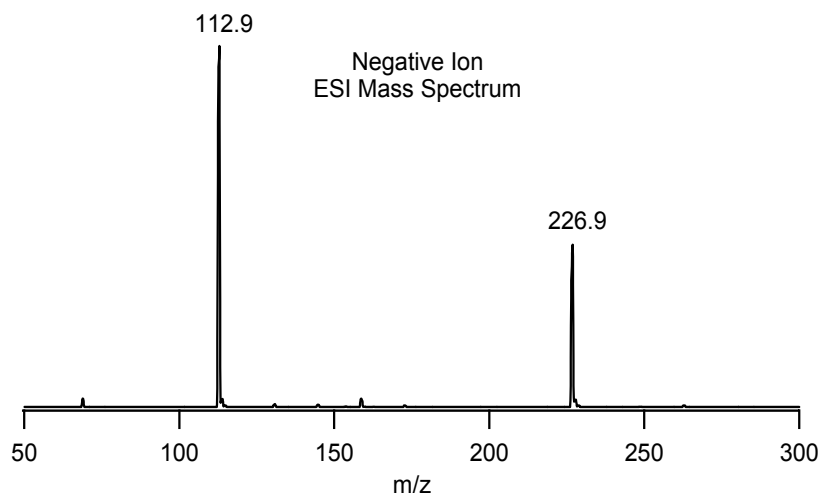
Figure 2.5: Negative ion ESI mass spectrum of the methanol-soluble products of the 2-pyrimidinone nucleoside synthesis at pH 2.1. The chemical formula and molecular weight for zebularine are $C_9H_{12}N_2O_5$, 228 g/mol.

Analysis by HPLC of the pH 2.1 2-pyrimidinone nucleoside reaction revealed two primary products with retention times greater than that of the unreacted 2-pyrimidinone base (Figure 2.4). Peak 2 was initially assigned as the β-furanosyl nucleoside, zebularine, in part, by co-injecting the sample with an authentic zebularine standard (Figure 2.6). Both peaks 1 and 2 were subsequently isolated by semi-preparative HPLC

chromatography (Section 2.2.3) and analyzed by UV-vis spectrophotometry and NMR spectroscopy. The UV-vis spectrum of the 2-pyrimidinone starting material shows a UV absorbance maximum at 298 nm, whereas the zebularine standard has an absorbance maximum at 303 nm (Figure 2.7). The spectra for peaks 1 and 2 both exhibit UV maxima at 303 nm, providing support that both are nucleoside products (Figure 2.7).
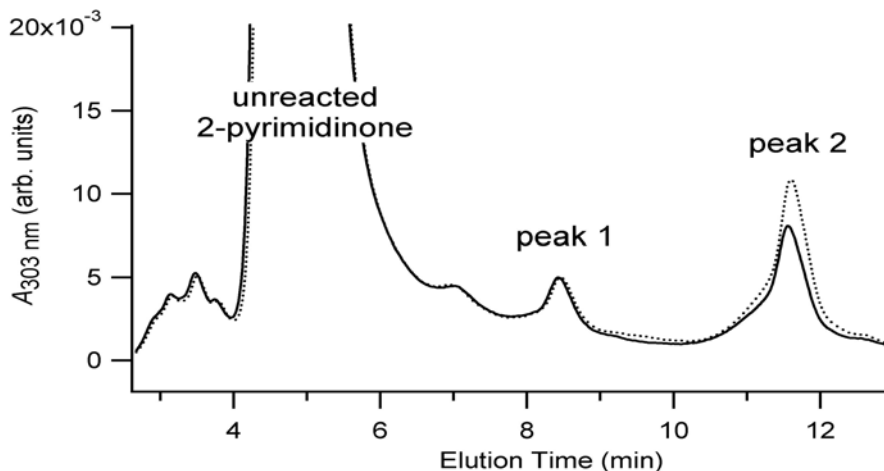
Figure 2.6: HPLC chromatographs of products that result from the drying and heating of 2-pyrimidinone with ribose at pH 2.1 (solid trace). Co-injection with an authentic zebularine standard is shown as dotted trace.
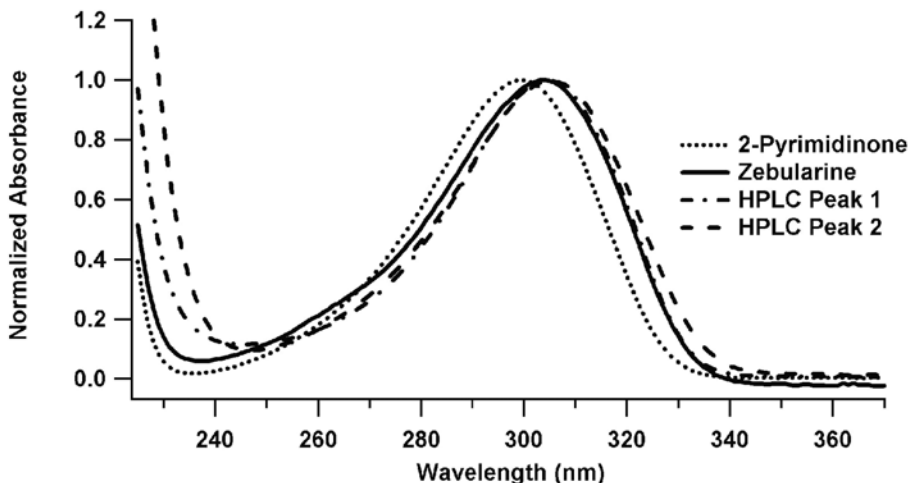
Figure 2.7: Normalized UV-vis absorbance spectra of the 2-pyrimidinone starting material, the zebularine standard, and the 2-pyrimidinone nucleoside reaction products HPLC peaks 1 and 2.

HPLC peaks 1 and 2 were also analyzed by NMR spectroscopy. The ^{1}H spectra of these products are shown in Figure 2.8, along with the spectrum of an authentic zebularine standard. The proton resonances were assigned by comparison with published chemical shifts for zebularine and confirmed by ^{1}H-^{1}H 2D NMR correlational spectroscopy (Figures 2.8 – 2.10). The labels for the ^{1}H resonances of the non-exchangeable sugar protons follow standard nucleoside convention. The non-exchangeable base protons are labeled as designated in Figure 2.2. The NMR spectra clearly show the characteristics of nucleosides, with chemical shifts and coupling constants that are consistent with other published ribonucleosides (*72, 30, 73*). Additionally, the COSY spectra for the isolated products show the cross peaks that are expected for nucleosides (Figures 2.8 – 2.10). Taken together, these spectra demonstrate that the heating-drying reaction employed produces more than one of the four possible 2-pyrimidinone nucleosides (i.e., β-furanosyl, β-pyranosyl, α-furanosyl, or α-pyranosyl ribonucleosides).

Comparison of the 1D NMR spectra of the authentic zebularine standard (Figure 2.8c) with that of the major species in HPLC peak 2 (Figure 2.8b) reveals a one-to-one correspondence for all protons, confirming the synthesis of the 2-pyrimidinone β-furanosyl ribonucleoside. The minor species in HPLC peak 2 is tentatively assigned the β-pyranose anomer of zebularine, based upon the large coupling constant between H1′ and H2′, the upfield shift of H2′, and the downfield shift of H3′, which have also been reported for other β-ribopyranosyl nucleosides (*30, 73*). The NMR spectrum of this minor species is also consistent with the reported chemical shifts for the β-ribopyranosyl 2-
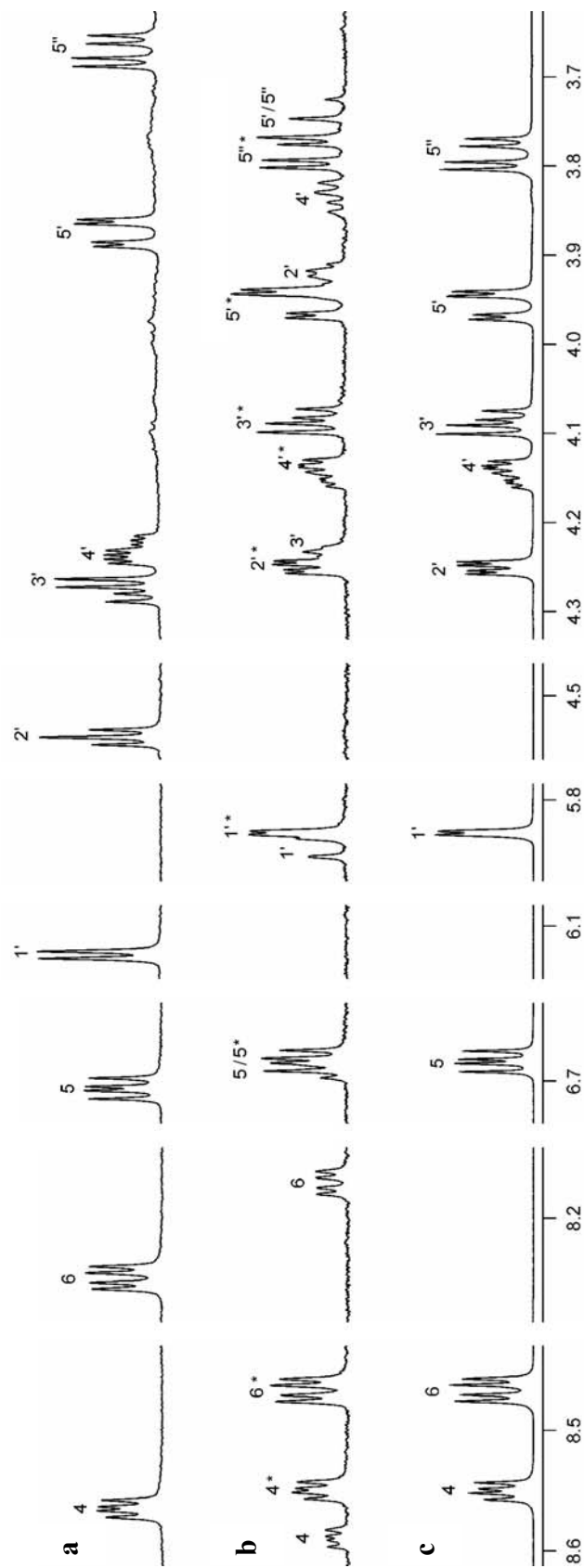
Figure 2.8: NMR spectra of 2-pyrimidinone ribonucleoside products and the zebularine standard. a) Spectrum of HPLC peak 1, which is tentatively assigned to the α-furanosyl ribonucleoside. b) Spectrum of HPLC peak 2. Two nucleoside products are isolated as peak 2, of which the major product is the β-furanosyl nucleoside, zebularine (peaks labeled with asterisks) and the minor product is tentatively assigned to the β-pyranosyl nucleoside. c) NMR spectrum of the authentic zebularine standard.
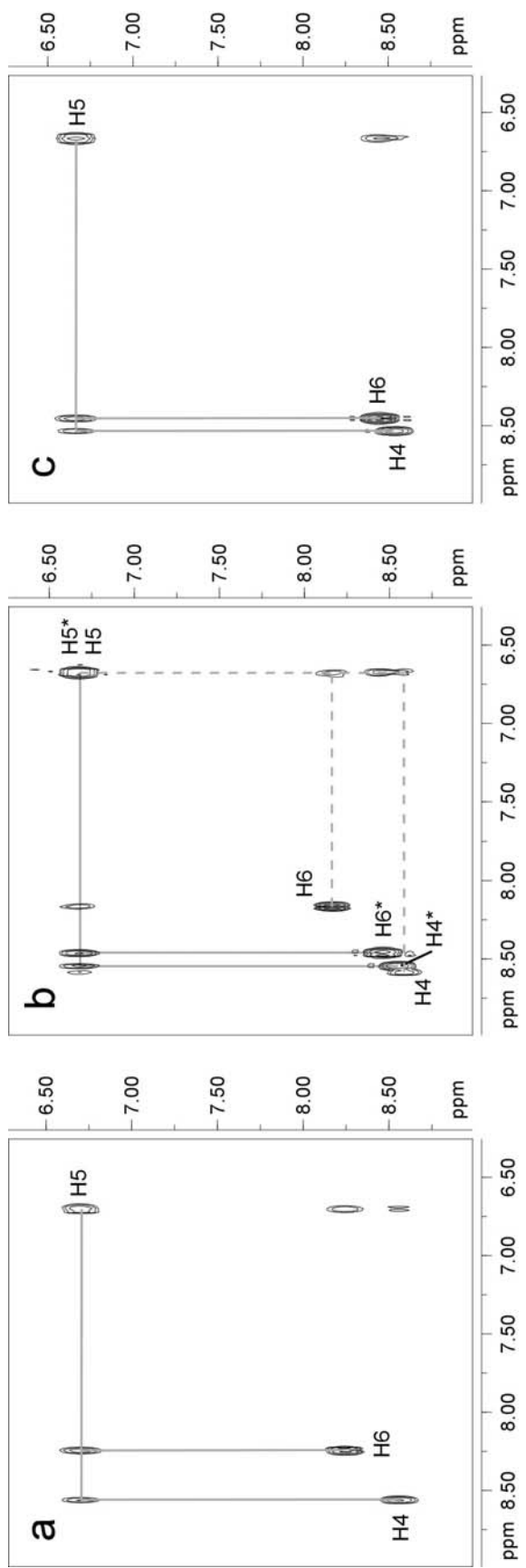
Figure 2.9: A comparison of cross peaks in the aromatic-aromatic region of 2D COSY spectra of the 2-pyrimidinone nucleoside products and the zebularine standard. a) HPLC Peak 1. b) HPLC peak 2, which contains two species. The protons of the major species, zebularine, are denoted with asterisks. c) The zebularine standard.

36
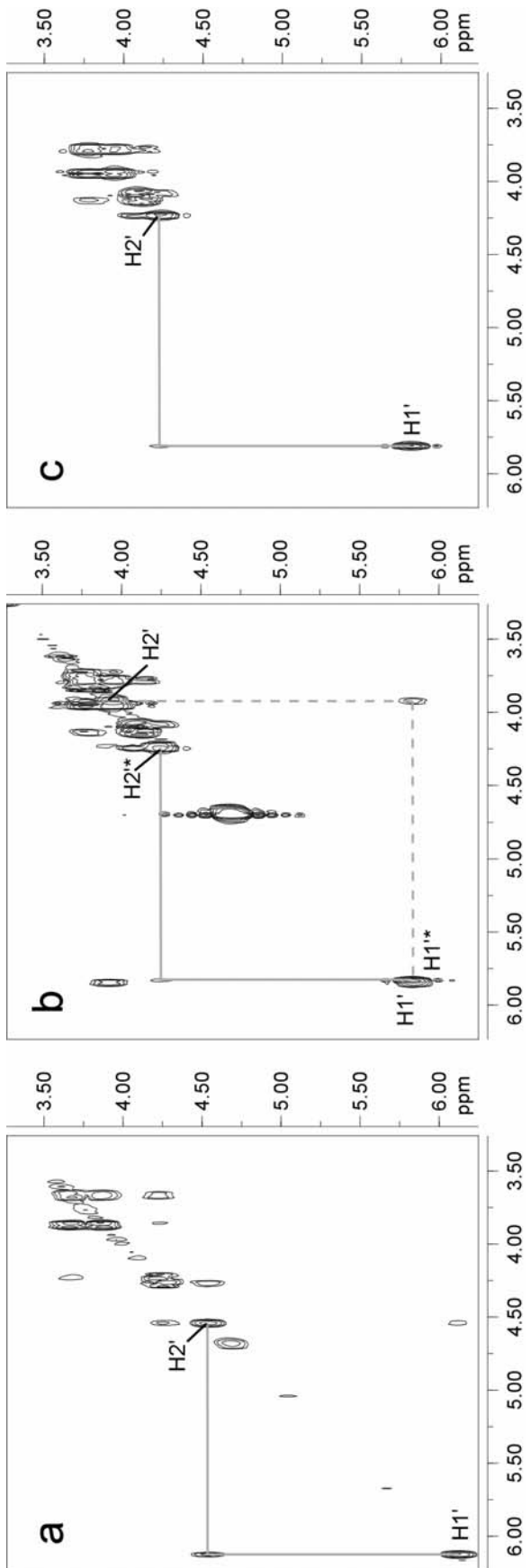
Figure 2.10: A comparison of H1′ and H2′ sugar proton chemical shifts and associated cross peaks from 2D COSY spectra of the 2-pyrimidinone nucleoside products and the zebularine standard. a) HPLC Peak 1. b) HPLC peak 2, which contains two species. The protons of the major species, zebularine, are denoted with asterisks. c) The zebularine standard.
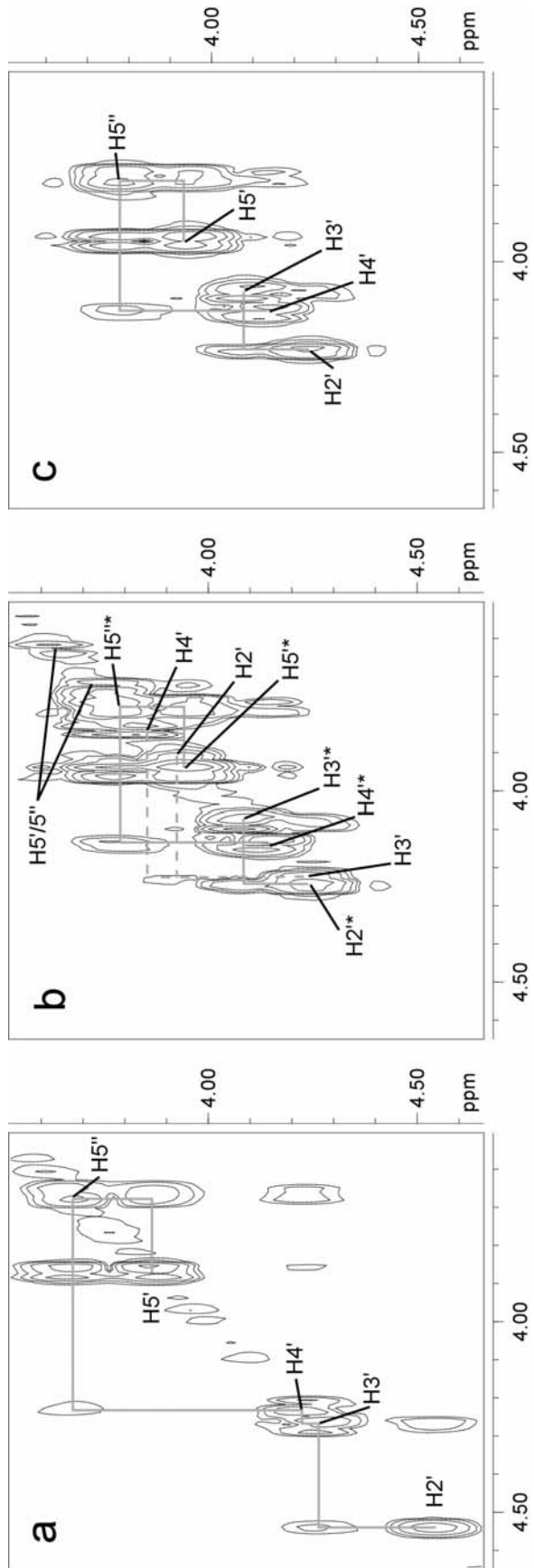
Figure 2.11: A comparison of proton chemical shifts and cross peaks from the H2′/H3′/H4′/H5′/H5″ region of 2D COSY spectra of the 2-pyrimidinone nucleoside products and the zebularine standard. a) HPLC peak 1. b) HPLC peak 2, which contains two species. The protons of the major species, zebularine, are denoted with asterisks. c) The zebularine standard.

38

pyrimidinone nucleoside (*72*). Several characteristics of the NMR spectrum of HPLC peak 1 suggest that it is the α-furanosyl anomer (Figure 2.8a). The downfield shifts of the H1′ and H2′, downfield shift of H3′ relative to H4′, upfield shift of H5′ and H5″, relatively large $\Delta\delta_{H5'/H5''}$ (~0.2 ppm), and the intermediate $J_{1'/2'}$ coupling constant of ~ 6 Hz that are observed for HPLC peak 1 are also reported for other α-ribofuranosyl nucleosides (*30, 73, 74*).

2.3.3 Quantitation of the β-Furanosyl Ribonucleoside Reaction Products

It is of interest to compare the nucleoside yields from 2-pyrimidinone, uracil, adenine, and hypoxanthine. The relative yields of β-furanosyl nucleosides produced in three different reaction conditions were determined by HPLC analysis as described in Section 2.2.1. Although some reactions generated side products at an estimated 15-20%, (particularly for 2-pyrimidinone, adenine, and hypoxanthine under Orgel's conditions, data not shown), only the starting material and β-nucleoside peak integrals were used in the calculation because the extinction coefficients of the side products are not known. The reported yields are averages of three replicates and have a standard deviation that is as high as 45%, which is likely due to the inherent variability of drying-heating reactions.

NMR spectra of HPLC peak 2 from the 2-pyrimidinone nucleoside synthesis reactions revealed that two nucleoside species, the β-furanosyl and β-pyranosyl ribonucleosides, co-elute as one peak. Thus, it was necessary to quantify the relative yields of these two nucleosides in each reaction so that zebularine yields, based upon HPLC peak integration, could be adjusted for the presence of the β-pyranosyl ribonucleoside. The ratio of β-furanosyl to β-pyranosyl ribonucleosides in HPLC peak 2 was determined for a representative set of reactions by first using semi-preparative HPLC

(Section 2.2.3) to isolate all products (i.e., products with longer column retention times than 2-pyrimidinone) from unreacted 2-pyrimidinone. NMR spectroscopy was then used to determine the relative concentration of the two β-nucleosides in the resulting samples. The ratios determined for the representative reactions were then applied to the yields measured by integration of HPLC peak 2 (Equation 2.4, Section 2.2.1) to provide individual β-furanosyl and β-pyranosyl yields.

For the Orgel reaction conditions, the total yield of peak 2 nucleosides (i.e., β-nucleosides) was determined to be 15% by Equation 2.4 (Section 2.2.1), the ratio of β-furanosyl to β-pyranosyl ribonucleoside was 4:1 as measured by NMR, and therefore the β-furanosyl yield was adjusted to 12%. The total yield of peak 2 nucleosides for the pH 2.1 reaction products was determined to be 2% and the measured ratio of β-furanosyl to β-pyranosyl ribonucleoside was 24:1, resulting in an adjusted yield of β-furanosyl ribonucleoside that was still, within experimental error, 2%. These results presented in Table 2.1 demonstrate that 2-pyrimidinone, surprisingly, forms β-nucleosides in greater yields than both adenine and hypoxanthine. Uracil did not form nucleosides in detectable yields under any of the conditions tested. As discussed in Section 2.2.1, the β-furanosyl ribonucleosides of adenine and hypoxanthine were identified solely by HPLC co-injection with adenosine and inosine standards, respectively. The actual yields for adenosine and inosine might, therefore, be less than reported if other reaction products co-elute with these β-furanosyl nucleosides.

In addition, we conducted nucleoside degradation studies using the same reaction conditions that were used for the nucleoside syntheses, presented in Table 2.2 for each set of reaction conditions. These data show that the relative glycosidic bond stabilities of

adenosine, inosine, uridine, and zebularine correlate with nucleoside synthesis yields measured for the same reaction conditions – the glycosidic bond of zebularine is the most labile under all three conditions tested, followed by inosine or adenosine, depending on the reaction conditions. Uridine degradation was not detected under any of these conditions. Transition state theory predicts that, under the same reaction conditions, the most easily formed nucleosides will also be those for which the glycosidic bond is most easily broken, and we observe a positive correlation between the propensity for a particular nucleoside to degrade and its formation yield.

Table 2.1: Comparison of β-furanosyl ribonucleoside yields for three bases under selected reaction conditions

Reaction	% Adenosine	% Inosine	% Zebularine
Orgel rxn	< 1	3	12
pH 2.1	1	1	2
pH 6.3	< 1	< 1	< 1

Uridine was not detected under any of the conditions tested. Reactions conditions are described in Section 2.2.1. Yields are absolute and neglect side products, which we estimate at < 20%. The maximum variability in the measured yields was 45%. Previously reported nucleoside yields were calculated BRSM (*8*).

Table 2.2: Comparison of glycosidic bond cleavage for four nucleosides under the reaction conditions tested for nucleoside synthesis

Reaction	% Adenosine	% Inosine	% Uridine	% Zebularine
Orgel rxn	0.9	6.3	< 0.1	29.0
pH 2.1	24.6	11.3	< 0.1	38.3
pH 6.3	< 0.1	< 0.1	< 0.1	< 0.1

Reaction conditions described in Section 2.2.2.

2.3.4 Modeling β-Pyrimidine Nucleoside Formation Between 2-Pyrimidinone and Ribose

In collaboration with Leszczynski and co-workers, we have used high-level calculations to model the zebularine formation reaction to understand why 2-pyrimidinone readily forms a glycosidic bond with ribose (*7*). Protonated 2-pyrimidinone has a pKa of 2.24 (*75*). Thus, most of the base will be protonated under the acidic reaction condition during the actual reaction, although this does not mean that the base must be the proton donor in an acid catalyzed reaction. In the presence of Mg^{2+} at pH 6.3, less than 0.1% of the base will be protonated. We have modeled glycosylation involving a closed sugar in analogy with other glycosylation and deglycosylation reactions (*76, 77*), although an alternative route may also exist with the open-chain sugar (*78*). The stationary structures obtained at the B3LYP/6-31G(d) level reveal that protonated 2-pyrimidinone forms a complex with ribose that includes a strong H-bond between the N1 proton of 2-pyrimidinone and the O1′ hydroxyl group of ribose (Figure 2.12). This feature of protonated 2-pyrimidinone reduces the activation energy dramatically by 18 kcal/mol over the gas phase pathway (*7*). This proton is also transferred to the O1′

Figure 2.12: Acid catalyzed reaction between ribose and 2-pyrimidinone in the absence of Mg^{2+}. Labeled distances are in Ångstroms. The structures are depicted using standard CPK coloration. See (*7*) for associated energies and for calculation methodology.

hydroxyl group during cleavage of the C1'-O_{OH} bond. These results demonstrate why acidic conditions promote glycosidic bond formation between ribose and 2-pyrimidinone. Although not modeled, proton delivery to the O1' hydroxyl group could also come from the solvent to conceivably produce the same transition state.

The reaction mechanism for zebularine formation in the presence of Mg^{2+} was also investigated at the B3LYP/6-31G(d) level. An initial complex was formed between ribose and protonated 2-pyrimidinone with the assistance of the Mg^{2+} cation (**Mg-COM**, Figure 2.13). In a transition state structure **Mg-TS1**, the breaking of the C1'-O_{OH} bond is facilitated by the Mg^{2+} cation and a bond is formed between the leaving HO$^-$ group and
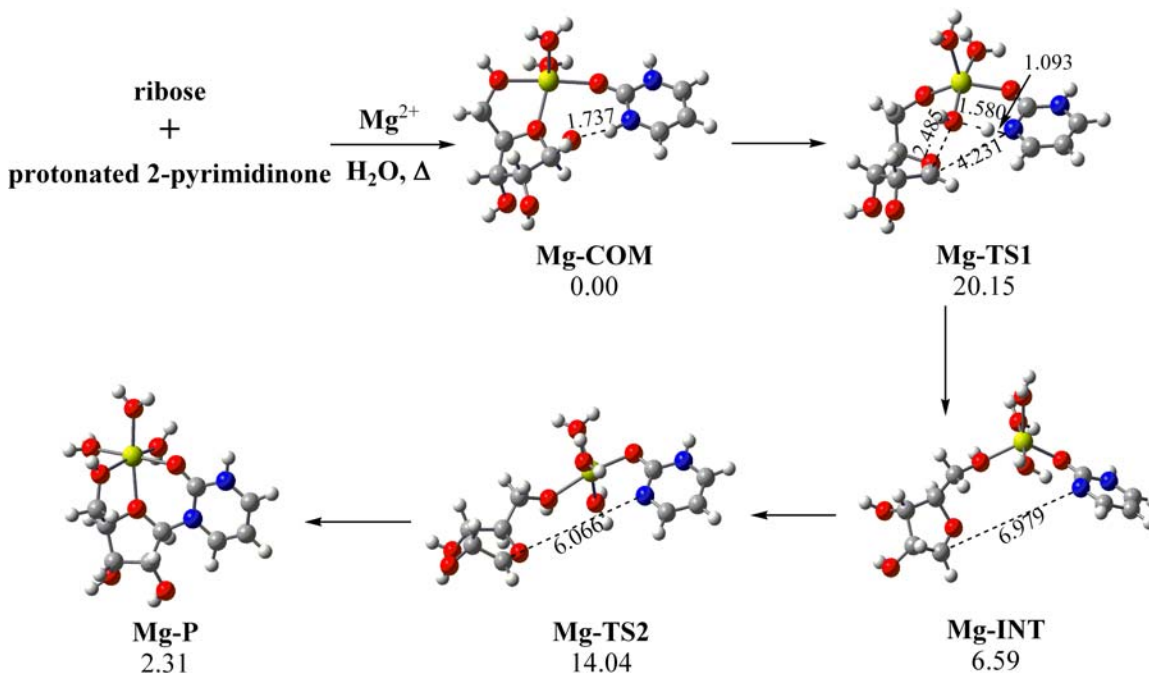
Figure 2.13: The structures and relative energies of reactants, complex, intermediate, transition structures, and product for the reaction between ribose and 2-pyrimidinone in the presence of Mg^{2+}. Labeled distances are in Ångstroms. Magnesium is drawn in yellow, and all other atoms follow standard CPK coloration. See (*7*) for associated energies and for calculation methodology.

the Mg^{2+} cation. The activation energy for this reaction step is calculated to be 20.2 kcal/mol, which is 14 kcal/mol lower than the activation energy for the most favorable reaction mechanism determined in the absence of Mg^{2+} (Table 2.3). Comparison of results obtained from all calculations performed provides the following conclusions: 1) Acid catalyzes glycosidic bond formation by facilitating cleavage of the $C1'-O_{OH}$ bond, and the acidic proton can be delivered by the protonated base; 2) Mg^{2+} facilitates the reaction between ribose and protonated 2-pyrimidinone by both lowering the energy of the transition states and by holding ribose and the protonated base in close proximity and in a relative orientation that is compatible with the transition states and the nucleoside product.

Table 2.3: The B3LYP/6-31G(d) predicted activation energies (in kcal/mol) for glycosidic bond formation between ribose and 2-pyrimidinone or uracil.

Reactants	Activation Energy
Ribose, pyrimidinone	52.7
Ribose, uracil	50.8
Ribose, protonated-pyrimidinone	34.3
Ribose, protonated-pyrimidinone, Mg^{2+}	**20.2**
Ribose, uracil, Mg^{2+}	39.9

See (*7*) for pathway structures and associated energies, and for calculation methodology.

It is instructive to consider that, at neutral pH, the bases that form glycosides under the present conditions, namely adenine, hypoxanthine, and 2-pyrimidinone, all have a reacting system that consists of an amidine moiety (the imidazole portion of the purine) or a vinylogous amidine moiety. This feature allows a ring nitrogen atom that carries an in-plane electron pair to act as a nucleophile while the second nitrogen atom, which carries an in-plane hydrogen atom, can simultaneously or subsequently lose this

hydrogen as a proton (Figure 2.14). Cytosine, in its canonical keto-amino tautomer, also

has such a structure, but the lone pair is on N3, a nitrogen atom that is sterically hindered,

being flanked by an NH_2 and a carbonyl group, apparently preventing reaction.
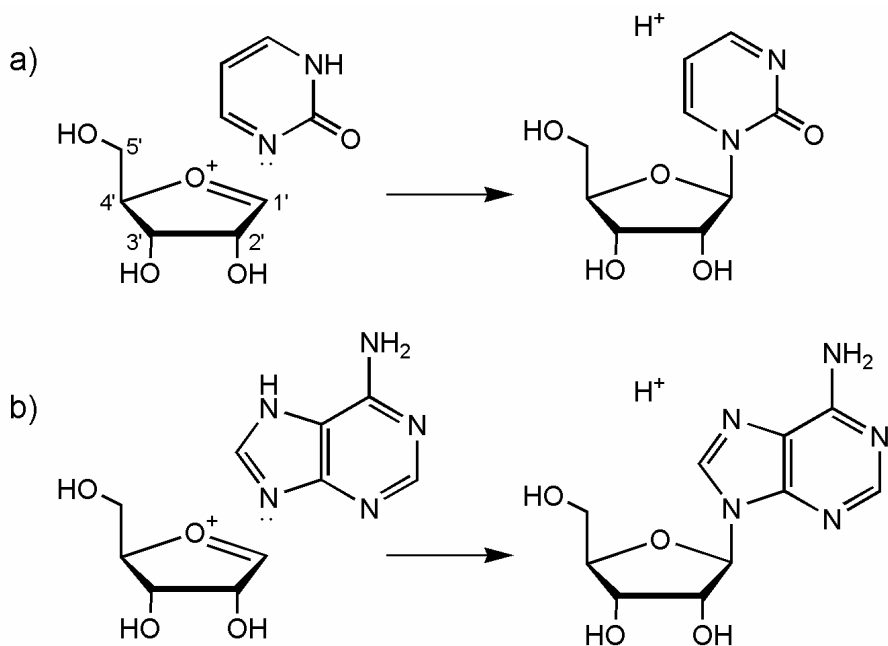
Figure 2.14: Proposed intermediates of nucleoside formation. The lone pair on the vinylogous amidine moiety of 2-pyrimidinone (a) or the amidine moiety of adenine (b) nucleophilically attacks C1′ of the oxonium sugar intermediate to form the glycosidic bond. A second nitrogen atom in the base (N3 of 2-pyrimidinone or N7 of adenine) either simultaneously or subsequently loses its hydrogen as a proton, completing the reaction.

Deprotonation of N1 of uracil and cytosine would provide an in-plane lone pair

that, by our proposed mechanism, would facilitate nucleoside formation. The pKa of N1-

H of uracil is about 9.5, so that there is less than 1% of the anion at neutral pH, and the

alkaline pH that is would be necessary for deprotonation is incompatible with the

oxonium sugar intermediate necessary for glycosidic bond synthesis. However, Kimura *et

al.* have shown that placement of the uracil base near a cationic species can selectively

reduce the pKa of N1-H by more than 2 pH units (*79*). Thus, coordination of uracil within

a supramolecular assembly, such as in a coaxial stack with cationic intercalators (*2*), might sufficiently lower the N1-H pKa to allow glycosylation.

The synthesis of zebularine and its anomers is, to our knowledge, the first demonstration of a prebiotic synthesis for a pyrimidine nucleoside from a free base and an unactivated sugar, demonstrating that pyrimidine nucleosides are prebiotically plausible compounds. We do not propose, however, that zebularine was necessarily a functioning nucleoside in proto-RNA. With the ability to form hydrogen bonds only via the C2 carbonyl and N3 as H-bond acceptors, the incorporation of 2′-deoxyzebularine into a DNA duplex opposite an adenine greatly destabilizes the duplex (ΔTm ~ - 9 °C/ substitution) (*80*). The significance of the work presented in this chapter is the experimental support that is provided for our hypothesis that, given the correct substitutions on the base to provide an in-plane lone pair on an unhindered nitrogen, prebiotic synthesis of pyrimidine nucleosides would have been possible. Other bases have been proposed that meet the chemical requirements of glycosylation, but posses greater hydrogen bonding capabilities (e.g., 5-azauracil, 4-pyrimidinone) (*81*). Indeed, initial experimentation with these bases indicates that ribosylation is occurring under Orgel's reaction conditions (*81*).

If 2-pyrimidinone nucleosides existed in the first RNA-like polymers, they could have been replaced by a post glycosylation base modification or by evolution. Conversion to uridine could have occurred by the addition of water across the N3-C4 bond (*82*), followed by hydride transfer of the C4 hydrogen to an acceptor (e.g., glyoxylate). The latter process would have been driven by the superior stability of uridine against glycosidic bond hydrolysis (Table 2.2) and greater functionality of the uracil

substituents for base pairing (*83*). It is possible that the urea group of the extant pyrimidine bases (i.e., N1, C2 carbonyl, and N3) is an inherited trait from earlier, more easily glycosylated pyrimidine bases. In any case, we suggest that the studies presented here indicate what might have been the essential features of pyrimidine (or pyrimidine-like) bases in early life.

CHAPTER 3

ACETAL NUCLEIC ACIDS AS A REVERSIBLE BACKBONE IN PREBIOTIC NUCLEIC ACIDS[†]

3.1 INTRODUCTION

The persistent lack of a feasible solution to the prebiotic synthesis of RNA has lead most researchers to conclude that RNA was preceded by one or more ancestral RNA-like polymers (*84*). Recently, both peptide nucleic acid (PNA) and threose nucleic acid (TNA) have garnered much attention as possible ancestors to RNA (*85-91*) (Figure 3.1). However, the PNA backbone is so chemically distinct from RNA, including a lack of negative charges, that it is difficult to imagine how PNA could have ever functioned in place of RNA (*50, 51*). TNA, which has a simpler sugar than ribose, still includes
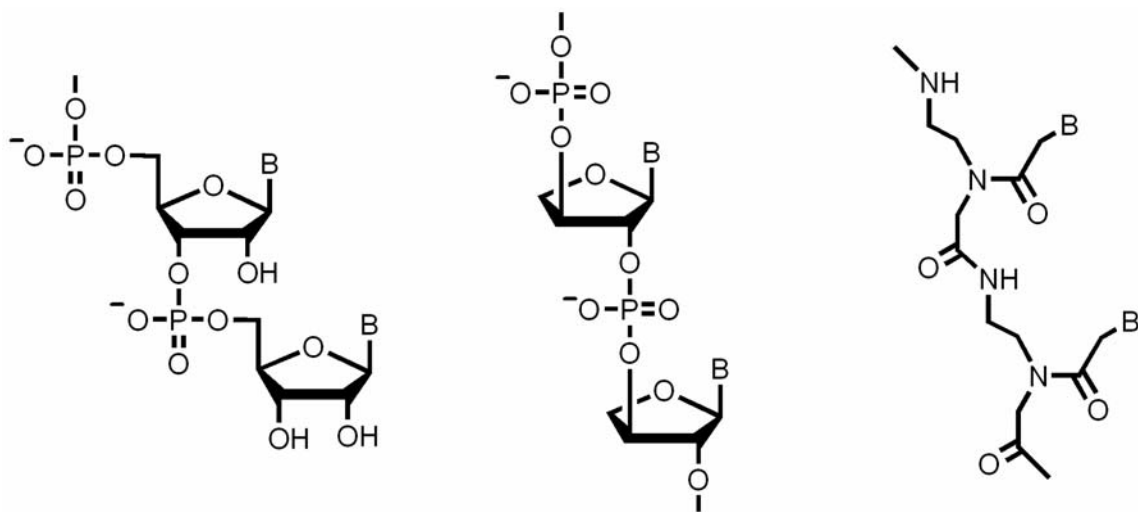
Figure 3.1: Structures of RNA (left), TNA (center) and PNA (right).

phosphodiester bonds, and thus suffers from some of the same difficulties associated with the prebiotic synthesis of RNA. In particular, phosphodiester bond formation requires chemical activation, and suitable phosphate precursors were likely scarce on the early Earth (*52*).

The experiments with the imine/amine backbone by Lynn and co-workers (described in Chapter 1) established thermodynamically-controlled polymerization through initial reversible condensation on a template as a robust approach for achieving high fidelity transfer of sequence information (*66, 68, 69*). Additionally, the ANP backbone is a good isosteric replacement of the phosphodiester backbone as it possesses six bonds from monomer to monomer and a tetrahedral geometry about the carbon of the amine linkage. As the ANP backbone is positively charged, it also satisfies the majority of the physical and chemical characteristics of phosphodiester-linked polymers, such as aqueous solubility, genotypic-phenotypic decoupling, and an extended backbone conformation. These physical and chemical characteristics would allow ANPs to fold into tertiary structures, replicate, and evolve.

That said, depolymerization of the amine backbone linkage of the ANPs would require oxidation, and thus error correction of ANPs would be a redox process. Water remains a necessity for life, and certainly life on Earth exploits condensation in the linkages of the nucleic acid monomers (glycosidic and phosphoester bonds), the amides of polypeptides, the esters of triglycerides, and the acetals of polysaccharides, and utilizes the reverse reaction of hydrolysis for metabolism (e.g., hydrolysis of ATP) and for the recovery of biopolymer building blocks. Therefore we have sought to extend the principles learned in the amine backbones to possible backbone chemistries that allow

equilibrium control of template-directed polymerization through condensation and hydrolysis reactions.

Acetal-linked nucleic acids (*a*NAs) have several characteristics that make them interesting as possible progenitors of RNA via this two-stage approach (Figure 3.2) (*2*).
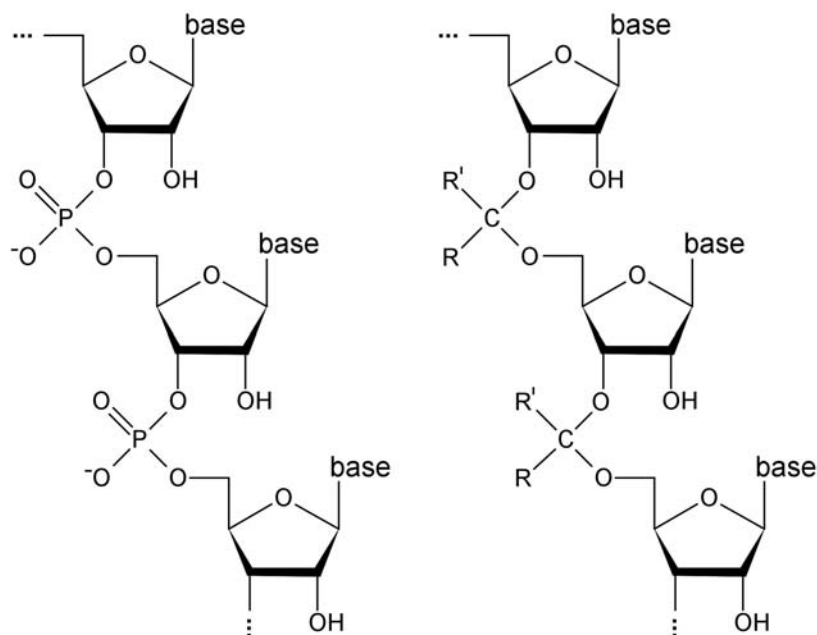
Figure 3.2: A comparison of the RNA (left) and the *a*NA (right) linkages. Both the phosphodiester linkage of RNA and the acetal linkage of *a*NA possess a tetrahedral geometry, resulting in similar helical properties of their duplexes (*6*). However, the acetal linkage is enthalpically favored over its hydrolysis product, and is more kinetically accessible than phosphodiester bonds.

First, acetals readily equilibrate in H_2O, for two-step error correction during replication, and condensation is enthalpically favored over hydrolysis when water is removed (*92*). Second, the *a*NA backbone is structurally similar to the RNA backbone as both the acetal bond and phosphodiester bond have a tetrahedral geometry (Figure 3.2). Moreover, it is already known that as phosphodiester isosteres, *a*NAs can base pair with RNAs (*93, 94*).

Third, acetal linkages can be functionalized to include charged groups that would maintain the chemical and physical properties of the nucleic acids as discussed above.
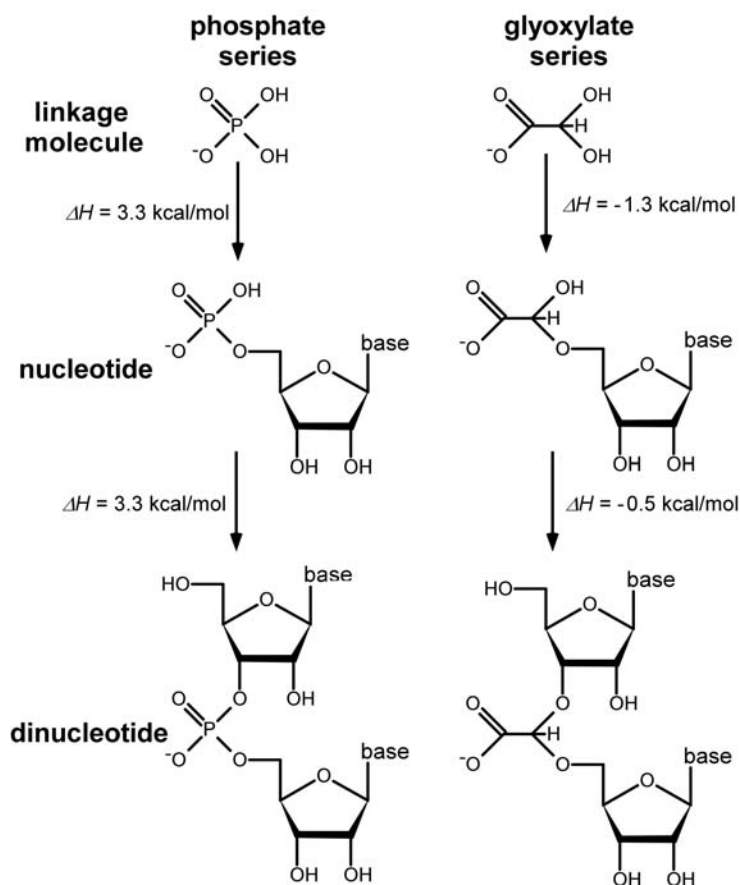
Figure 3.3: Comparison of phosphate and glyoxylate chemical structures, and their associated nucleotides and dinucleotides. ΔH values represent enthalpies of phosphoester bond formation (3.3 kcal/mol) (*3*), glyoxylic acid-methyl hemiacetal bond formation (−1.3 kcal/mol) and glyoxylic acid-dimethyl acetal bond formation (−0.5 kcal/mol). The experiments used to determine glyoxylic acid hemiacetal and acetal bond formation enthalpies are described in Section 3.2.1.

In this chapter we present experimental and computational evidence supporting our hypothesis that glyoxylic acid could have preceded phosphate in proto-RNA backbones (*6*). Glyoxylic acid exists in water at neutral pH in equilibria with its ionized form and also with its hydrate form, but almost entirely as the ionized aldehyde hydrate, a

51

form that for simplicity we will refer to as *glyoxylate*. Glyoxylate is the smallest aldehyde molecule that is negatively charged under neutral conditions, and its prebiotic synthesis from glycolaldehyde has been demonstrated (*4*). The structural similarity of glyoxylate and dihydrogen phosphate is clear (Figure 3.3). Close structural analogs of modern nucleotides would be created by the formation of hemiacetal bonds between glyoxylate and nucleoside sugars, which we term *ga*-nucleotides (Figure 3.3). Furthermore, nucleosides connected through glyoxylate by acetal bonds would likewise represent close structural analogs of RNA and DNA, termed collectively as glyoxylate-acetal nucleic acids, or *ga*NA (Figure 3.3). The method described here for creating *ga*-nucleotide linkages represents, to the best of our knowledge, the first demonstration that an RNA-like linkage can be created with plausible prebiotic chemistry.

3.2 METHODS

3.2.1 Determination of Enthalpies of Hemiacetal and Acetal Bond Formation

The enthalpy of hemiacetal formation by glyoxylate (Figure 3.3) was determined by measuring the equilibrium constant (based upon peak integration in NMR spectra) of glyoxylic acid hydrate and the glyoxylic acid hemiacetal (*R* and *S* forms) formed with methanol in a 100 mM glyoxylic acid solution, 20% w/v methyl-d_3 alcohol-d/80% w/v D$_2$O (pD 1) at 65 °C (K = 9.4; $\Delta G_{65\,°C}$ = −1.5 kcal/mol) and 85 °C (K = 8.4; $\Delta G_{85\,°C}$ = −1.5 kcal/mol). The free energy of acetal formation by glyoxylic acid (from its hemiacetal form) was determined by measuring the equilibrium constant of the glyoxylic acid-methyl hemiacetal (*R* and *S* forms) and the glyoxylic acid-dimethyl acetal in a 100 mM glyoxylic acid solution, 20% w/v methyl-d_3 alcohol-d/80% w/v D$_2$O (pD 1) at 65 °C (K = 1.4; $\Delta G_{65\,°C}$ = −0.7 kcal/mol). The enthalpy of glyoxylic acid acetal formation was

derived from the free energy measurement using the entropy measured for the hydrate/hemiacetal reaction in the same sample, which also coincides with the addition of one methanol and the liberation of one water molecule. For both free energy derivations reported, peak integration in NMR spectra was taken over the glyoxylic acid and the methyl-ester forms of glyoxylic acid that exist in equilibrium exchange at pD 1.

3.2.2 Synthesis of *ga*-Dinucleotides with Metal Ion Hydrates

ga-Dinucleotide yields presented below were measured from reaction samples that contained 20 μmol dT, 10 μmol glyoxylate (pH 6.4), and 100 μmol of the chloride salt for each metal ion. In the case of $MgCl_2$, *ga*-dinucleotide yield was determined to be highest at a relative concentration of salt to glyoxylate of approximately 10:1 (data not shown). This same ratio was used for all metal ion species tested. The samples were dried at 22 °C under vacuum or at 65 or 85 °C at atmospheric pressure, and then resuspended in a 0.25 N NH_4OH solution.

3.2.3 Mass Spectrometry

Mass spectra of *ga*-nucleotides and *ga*-dinucleotides were collected in negative ion mode on a Micromass Quattro LC mass spectrometer and ionized under electrospray conditions (capillary voltage, 2.5 kV; cone voltage, 40 V). Accurate mass analyses, used to confirm *ga*-nucleotide and *ga*-dinucleotide formation, were accomplished using the negative ion mode on an Applied Biosystems QSTAR XL hybrid Quadrupole-TOF mass spectrometer under electrospray conditions (capillary voltage, −4.2 kV; cone voltage, −70 V).

3.2.4 HPLC and NMR Spectroscopy

HPLC separation of dT/glyoxylate reaction products was achieved on a Phenomenex ODS analytical column (4.6 × 250 mm, 5 μ) with mobile phase conditions of 10–30% B (0–5 min), 30–50% B (5–30 min), flow rate = 1 ml/min. Mobile phase A = 0.1 M triethyl ammonium acetate (TEAA), B = 25% (v/v) acetonitrile in 0.1 M TEAA. dT-*ga*-dT *ga*-dinucleotides for the NMR study shown in Figure 3.7B were prepared in bulk by drying under vacuum at room temperature ten solutions of 20 μmol dT nucleoside, 10 μmol glyoxylic acid, and 15 μmol HCl in 200 μl dH_2O. Reaction products were resuspended in 0.25 N NH_4OH, pooled and purified by the same HPLC gradient with a Phenomenex ODS semipreparative column (10 × 250 mm, 5 μ). The process of *ga*-dinucleotide isolation and purification resulted in less than 15% degradation into monomer dT, as determined by NMR. Products of bulk dT-*ga*-dT *ga*-dinucleotide preparations were correlated with products of $MgCl_2$-catalyzed reactions by simultaneous injection onto an analytical HPLC column. NMR spectra of *ga*-dinucleotides were recorded on a Bruker DRX500 spectrometer in D_2O at 25 °C.

3.2.5 Molecular Modeling

Partial electrostatic charges for the glyoxylate atoms and ribose atoms of *ga*-oligonucleotides were determined by fitting the electrostatic surface potential obtained from a high-level quantum mechanical calculation of ribose-3′-glyoxylate-5′-ribose as described in the Cornell force field paper (*95*). The Amber suite of programs (*96*) was used for generation, energy minimization, and molecular dynamics simulations of a *ga*RNA duplex. The model *ga*RNA duplex with the self-complementary sequence A_6U_6 was constructed by replacing the phosphate groups of an RNA duplex $[A_6U_6]_2$ in a

canonical A-form helix with glyoxylate. *R* and *S* isomers of the glyoxylate-acetal linkages between nucleotides were varied along the two strands as: 5′-A*Rga*A-*Sga*A-*Sga*A-*Rga*A-*Rga*A-*Sga*U-*Rga*U-*Sga*U-*Rga*U-*Rga*U-*Sga*U-3′ and 5′-A-*Sga*A-*Rga*A-*Sga*A-*Rga*A-*Sga*A-*Rga*U-*Rga*U-*Rga*U-*Sga*U-*Rga*U-*Sga*U-3′, respectively. The energy minimization and molecular dynamics simulation protocol used was adapted from that described previously for nucleic acids simulations (*97*). Briefly, prior to energy minimization, 22 Na$^+$ counterions and 4142 solvent water molecules were added around the *ga*RNA duplex in a truncated octahedral solvent box that extended at least 8 Å from all solute atoms. Solvent water molecules and cations were energy minimized by 500 steps of steepest descent and 500 steps of conjugate gradient, while holding the *ga*RNA duplex atoms fixed. A second minimization was then carried out with no constraints on the duplex or solvent. The model duplex and solvent molecules were heated to 300 K over the course of 20 ps, and then subjected to 1 ns of molecular dynamics simulation.

3.3 RESULTS

3.3.1 Equilibria of Glyoxylate Acetals

Glyoxylic acid exists in neutral aqueous solution predominantly as its ionized, hydrated form, glyoxylate hydrate (Figure 3.4). Through the removal of a water molecule and the addition of an alcohol, such as a nucleoside, the hemiacetal is formed (Figure 3.4). This reaction can take place in acidic, basic, or neutral solutions. The formation of the acetal from the hemiacetal occurs under acidic conditions via the loss of a second water molecule followed by the addition of a second alcohol (Figure 3.4B). When the alcohols are nucleosides, hemiacetal formation yields *ga*-nucleotides and acetal formation yields *ga*-dinucleotides (Figure 3.3). Therefore, the overall reaction for the synthesis of

ga-dinucleotides requires the addition of two alcohols and the removal of two waters, given by Equation 3.2.

A
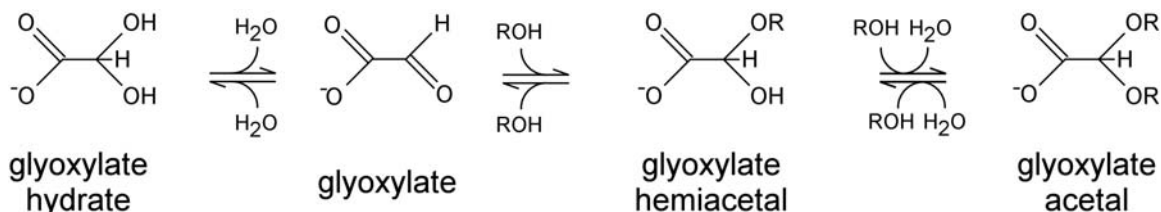
B
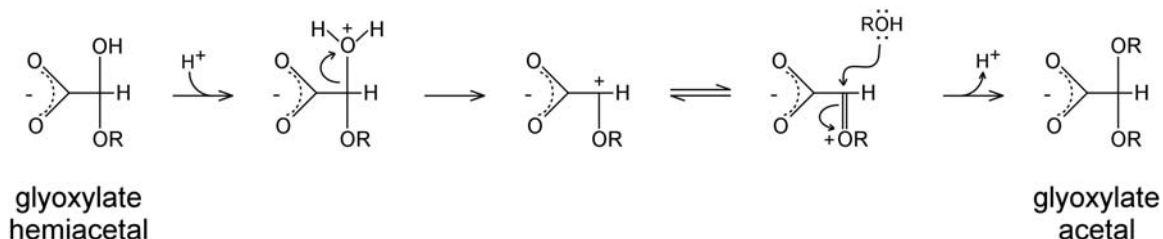
Figure 3.4: Equilibria and mechanisms of glyoxylate acetal formation. A) Glyoxylic acid exists in neutral aqueous solutions as glyoxylate hydrate. The hemiacetal is formed via the loss of a water molecule, followed by the addition of an alcohol, which is a nucleoside in the case of *ga*NA synthesis. Via the loss of a second water molecule and the addition of a second alcohol, the acetal is formed. B) The conversion from the hydrate to the acetal is an acid catalyzed reaction that proceeds through a carbocation intermediate.

$$\text{hydrate} + 2\text{ROH} \rightleftharpoons \text{acetal} + 2\text{H}_2\text{O} \qquad\qquad \text{Equation 3.1}$$
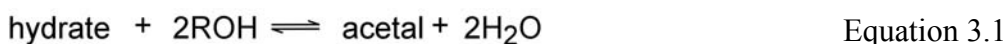
$$K_{sum} = \frac{K_{hemiacetal} * K_{acetal}}{K_{hydrate}} = \frac{[\text{acetal}][\text{H}_2\text{O}]^2}{[\text{hydrate}][\text{ROH}]^2} \qquad\qquad \text{Equation 3.2}$$

3.3.2 Enthalpies of Glyoxylate Hemiacetals and Acetals

Based upon NMR analysis of glyoxylic acid dissolved in a water-methanol mixture, we have determined that a glyoxylic acid-methyl hemiacetal is enthalpically favored over its hydrate species by 1.3 kcal/mol, and a glyoxylic acid-dimethyl acetal is

enthalpically favored over its hemiacetal by an additional 0.5 kcal/mol. These enthalpies of hemiacetal and acetal bond formation between glyoxylic acid and methanol are used as estimates for the bond formation enthalpies of glyoxylate-nucleoside hemiacetal and acetal species (Figure 3.3). It is expected that these enthalpy values are quite reliable, as the enthalpies of hemiacetal and acetal formation have previously been shown to only change by a few percent with the identity of the alcohol (*98*). These enthalpies of bond formation are in contrast to the 3.3 kcal/mol of enthalpy that must be provided to phosphorylate a nucleoside, and the 3.3 kcal/mol required for the subsequent formation of a phosphodiester linkage (*3*). The negative enthalpies of formation of hemiacetals and acetals suggest that *ga*-nucleotides, *ga*-dinucleotides, and *ga*NA polymers will form spontaneously under appropriate conditions.

3.3.3 Formation of Glyoxylate Hemiacetals with Nucleosides

The formation of a hemiacetal by an alcohol (e.g., the OH of ribose) and an aldehyde hydrate (e.g., glyoxylate) is a dehydration reaction that releases one molecule of water and takes place in acidic, basic, or neutral pH conditions. The existence of *ga*-nucleotides in neutral aqueous solutions containing a mixture of glyoxylate and each of the RNA and DNA nucleosides is supported by mass spectrometry. Samples containing 100 mM glyoxylate and 100 mM nucleoside (pH 6.9) were prepared at room temperature and injected directly into the mass spectrometer. In each case a prominent peak in the mass spectrum confirmed formation of the *ga*-nucleotide (Table 3.1). *ga*-rU and *ga*-dT nucleotides were further characterized by high-resolution mass spectrometry, which confirmed that the observed masses for these nucleotides match their theoretical masses within 3 ppm error (Table 3.3).

Table 3.1: Mass spectrometry of hemiacetal peaks

ga-nucleotide[a]	observed m/z[b]
ga-rU	317
ga-rC	316
ga-rA	340
ga-rG	356
ga-dT	315
ga-dC	300
ga-dA	324
ga-dG	340

[a] The observed mass is consistent with a ga-nucleotide. Molecular structure has not been determined.
[b] Data collected in negative ion mode as described in Methods

Table 3.2: Mass spectrometry of acetal peaks

ga-dinucleotide[a]	observed m/z[b]
rU-ga-rU	543
rC-ga-rC	541
rA-ga-rA	589
rG-ga-rG	nd[c]
dT-ga-dT	539
dC-ga-dC	509
dA-ga-dA	nd[c]
dG-ga-dG	nd[c]

[a] The observed mass is consistent with a ga-dinucleotide. Molecular structure has not been determined.
[b] Data collected in negative ion mode as described in Methods
[c] Not detected

Table 3.3: High resolution mass spectrometry data for selected hemiacetal and acetal species

species	molec. formula[a]	observed m/z[b]	calculated mass	ppm error
ga-rU	$C_{11}H_{13}N_2O_9$	317.0616	317.0621	1.6
ga-dT	$C_{12}H_{15}N_2O_8$	315.0837	315.0828	2.7
rU-ga-rU	$C_{20}H_{23}N_4O_{14}$	543.1177	543.1211	6.2
dT-ga-dT	$C_{22}H_{27}N_4O_{12}$	539.1593	539.1626	6.0

[a] Formula of the negatively ionized species
[b] Data collected in negative ion mode as described in Methods

Although all eight RNA and DNA *ga*-nucleotides were observed by mass spectrometry, the electrospray ion source of the mass spectrometer desolvates samples prior to ion separation and detection. This desolvation process within the mass spectrometer creates dehydrating conditions that promote hemiacetal (e.g., *ga*-nucleotide) formation. To confirm that *ga*-nucleotides are also formed in the solution state, a sample containing 1 M glyoxylate (pH 6) and 1 M 1-O-Methyl-2-deoxy-D-ribose (dOM), a model of a nucleoside sugar, was analyzed by solution state NMR spectroscopy (Figure 3.5). New chemical species were observed to form upon mixing glyoxylate and dOM, which are consistent with the formation of *ga*-dOM nucleotide analogs in solution. The use of dOM instead of any single nucleoside was done for solubility reasons. The equilibrium constant of the glyoxylate hydrate-hemiacetal system in H_2O-ROH is approximately 10 (Section 3.2.1, Figure 3.4). Thus, in a solution with D_2O present at a nominal concentration of 55 M, confirmation of *ga*-nucleotide formation by NMR spectroscopy in the presence of free dOM and the glyoxylate hydrate requires a dOM concentration that is beyond the solubility limit of the natural nucleosides (i.e., approximately 1 M) and is therefore difficult to study in the solution state.

3.3.4 Formation of Glyoxylate-Linked Dinucleotides

The formation of an acetal bond in the reaction of a hemiacetal and an alcohol, such as that required for the formation of a *ga*-dinucleotide, also represents a dehydration reaction (loss of a water molecule), but requires acid catalysis. It does not take place under neutral or basic conditions, in contrast to the reaction for hemiacetal formation. Acetal bond formation would typically be carried out in a synthetic laboratory using a
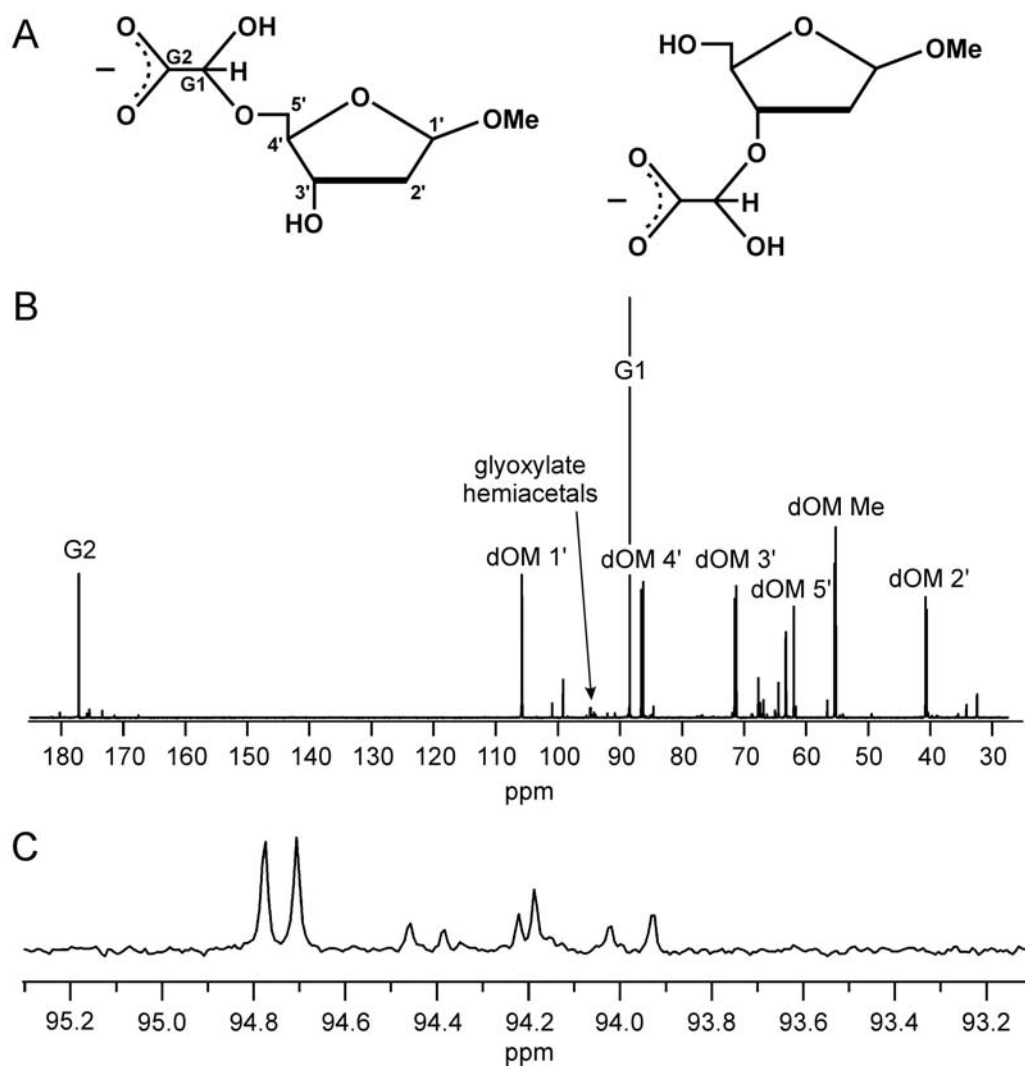
Figure 3.5: Verification of *ga*-nucleotide formation in solution by glyoxylate and 1-O-Methyl-2-deoxy-D-ribose (dOM). **(A)** Structures of the 3′ and 5′ hemiacetals. The dOM exists in both α and β forms, and the glyoxylate moiety is prochiral, giving rise to eight possible hemiacetal structures. **(B)** Natural abundance ^{13}C NMR spectrum of 1 M glyoxylate and 1 M dOM in D_2O (pH 6). At this approximate ratio of 55:1 D_2O:dOM, the dominant signals in the NMR spectrum correspond to free dOM and free glyoxylate. The smaller peaks indicated by arrow correspond to the glyoxylate CH (G1) resonances of 3′ and 5′ *ga*-dOM hemiacetals. Resonance assignments are based upon correlations observed in 2D HMQC and HMBC spectra (not shown). **(C)** Expanded region of the NMR spectrum (B) that contains the G1 resonances of the *ga*-dOM hemiacetals.

strong acid catalyst and an anhydrous medium. However, life likely emerged from an aqueous environment within the pH range of 6 - 9 (*99*). Given these restraints, we sought to catalyze acetal bond formation using cationic metal hydrates as general acids (Table 3.4) and evaporation of water by drying as a means to reduce the water activity to a very low level. Similar conditions have previously proven fruitful in the prebiotic synthesis of nucleosides from the free bases and ribose (*8*).

Table 3.4: pKas of several of the metal hydrates tested for dT-*ga*-dT synthesis. Adapted from (*1*).

Metal	Oxidation State	Hydration number	pKa of aqua ion
Group 1A			
Li	1+	4	13.8
Na	1+	6	14.48
Group 2A			
Mg	2+	6	11.42
Ca	2+	8	12.7
Sr	2+	6	13.18
1st row transition metals			
Fe	2+	hs 6	9.5
		ls 6	
Fe	3+	hs 6	0.70
		ls 6	2.19
Ni	2+	T_d4	9.86
		sq 4	

Samples containing 50 mM glyoxylate, 100 mM nucleoside, and 500 mM $MgCl_2$ (pH 6.4) were dried at 85 °C in an open-air test tube and then resuspended in dH_2O at room temperature. Using mass spectrometry we have verified that nucleosides dT, dC, rU, rC, and rA will couple under these conditions to form homo *ga*-dinucleotides (Table 3.2). The *ga*-dinucleotides dT-*ga*-dT and rU-*ga*-rU were further characterized by high-resolution mass spectrometry (Table 3.3). The remaining nucleosides, dA, dG, and rG,

did not produce *ga*-dinucleotides in detectable yields, and no dinucleotide formation was observed without sample drying. A correlation exists between nucleoside solubility and *ga*-dinucleotide production, strongly suggesting that nucleosides dA, dG, and rG crystallize from solution before acetal bond formation can occur (see Chapter 4 for more details).

We have also investigated the effects of temperature and cationic metal species on *ga*-dinucleotide formation during sample drying. For this study, samples of dT and glyoxylate were prepared with eight different metal chlorides and dried at 22, 65, or 85 °C (Figure 3.6). At 22 °C only the divalent metal ions Fe^{2+}, Ni^{2+}, and Zn^{2+} produced detectable yields of dT-*ga*-dT *ga*-dinucleotides. At 65 and 85 °C all metal ions tested produced appreciable amounts of *ga*-dinucleotides. The highest yield was observed for the solution containing Mg^{2+} that was dried at 85 °C, and corresponded to the conversion of approximately 1% of starting dT nucleoside into *ga*-dinucleotides. All other values shown in Figure 3.6 are normalized to this reaction, and lie roughly within an order of magnitude for all metal ions tested at 65 and 85 °C. Thus, acetal bond formation by glyoxylate and nucleosides requires a metal ion catalyst, but does not strongly depend on the elemental species of the metal ion. From the perspective of prebiotic chemistry, Mg^{2+} was likely to have been widely distributed in liquid water on the prebiotic Earth, as it is today. Furthermore, temperatures around 65 °C would have likely been a common occurrence on the early Earth (*99*). Hence, nucleosides on the prebiotic Earth could have been coupled through glyoxylate-acetal linkages without the need for any additional catalysts, energy sources, or chemical activation, provided that nucleoside organization occurred before reactant solubility became limiting (Section 3.4.1).
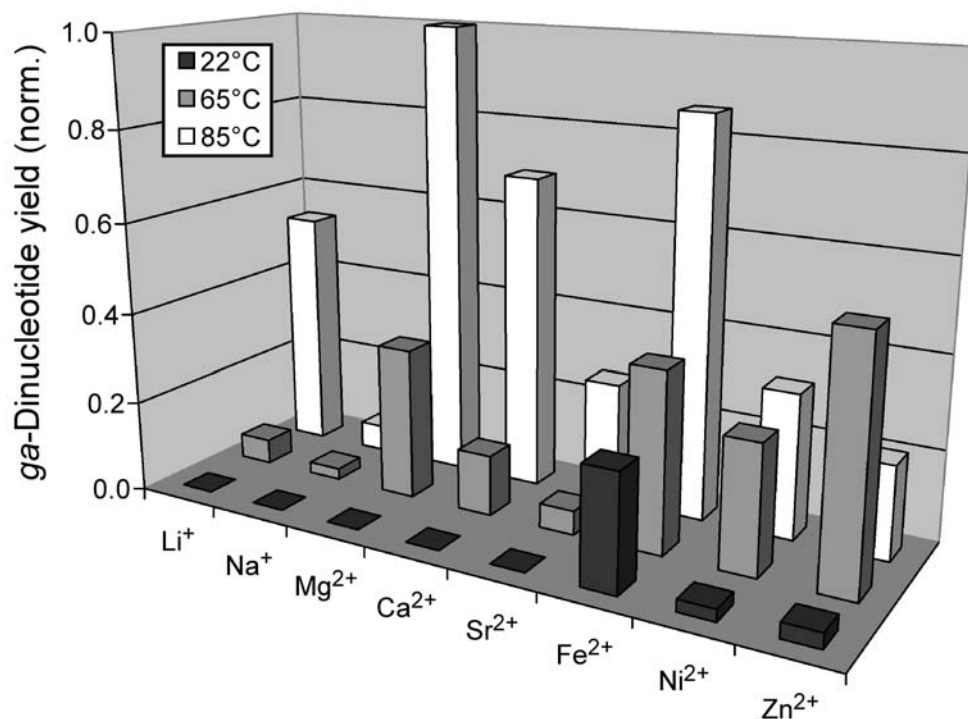
Figure 3.6: Effects of metal ion species and temperature on the formation of dT-*ga*-dT *ga*-dinucleotides. Each sample contained 20 μmol dT, 10 μmol glyoxylate (pH 6.4), and 100 μmol of the chloride salt for each metal ion. The 65 and 85 °C reaction mixtures were dried at atmospheric pressure; the room temperature samples were dried under vacuum. Five identical samples (initially 200 μl) were dried and analyzed by HPLC for each salt/temperature combination. Yields are based upon the average integrated HPLC peak intensities for the five samples of each reaction condition. All yields are normalized with respect to the yield of the $MgCl_2$ samples dried at 85 °C.

3.3.5 Characterization of dT-*ga*-dT Dinucleotide Linkages

Deoxyribose nucleosides and glyoxylate have the potential to form acetal-linked dinucleotides involving all combinations of the 3′-OH and 5′-OH groups of deoxyribose (i.e., 3′-*ga*-5′, 5′-*ga*-5′ and 3′-*ga*-3′ linkages), and ribonucleosides can also form linkages involving the 2′-OH group of ribose. Additionally, the CH carbon of glyoxylate is prochiral and can give rise to *R* and *S* isomers when the two acetal groups are different from one another. The possible structural implications of coexisting *R* and *S* isomers are

63

addressed below (Section 3.3.7). In the case of *ga*-dinucleotide formation, acetal bonds involving the 5′-OH, a primary alcohol, would be expected to be more favored than bonds involving one or more of the secondary alcohols (i.e., 2′-OH and 3′-OH) (*98, 100*). To discern which of the possible *ga*-dinucleotides are actually synthesized, we conducted a more in-depth analysis of the products formed when dT is dried from a neutral solution containing glyoxylate and $MgCl_2$. HPLC analysis of reaction products exhibited several peaks with relative mobilities close to that of the natural dinucleotide dT-p-dT (Figure 3.7A). Mass spectrometry confirmed that several of these elution peaks have a mass corresponding to a dT-*ga*-dT dinucleotide. HPLC peaks labeled A, B, and C were also present in a bulk acid-catalyzed preparation, allowing the compounds in these peaks to be obtained in sufficient quantity for characterization by NMR.

In Figure 3.7B proton NMR spectra for the *ga*-dinucleotide HPLC peaks A, B, and C are shown along with an NMR spectrum of the natural dT-p-dT dinucleotide. The proton resonance in the 4.8 − 5.0 ppm region of the *ga*-dinucleotide spectra belongs to the single CH proton of the glyoxylate moiety. 2D correlation ^{1}H-^{13}C spectroscopy (HMBC) revealed peak C to be largely the 5′-*ga*-5′ dinucleotide, and peaks A and B to be largely 3′-*ga*-5′ linked dinucleotides (Tables 3.5 − 3.10). Analysis of additional HPLC elution peaks gave no indication of dinucleotides with 3′-*ga*-3′ linkages, which is consistent with the reaction of two secondary alcohols with glyoxylate being the least favored product of the four possible linkages. We note that the observation of more than three peaks in the dinucleotide region of the HPLC trace implies the creation of more products than the 5′-*ga*-5′ and 3′-*ga*-5′ dinucleotides. Cyclic mononucleotides have been identified by mass
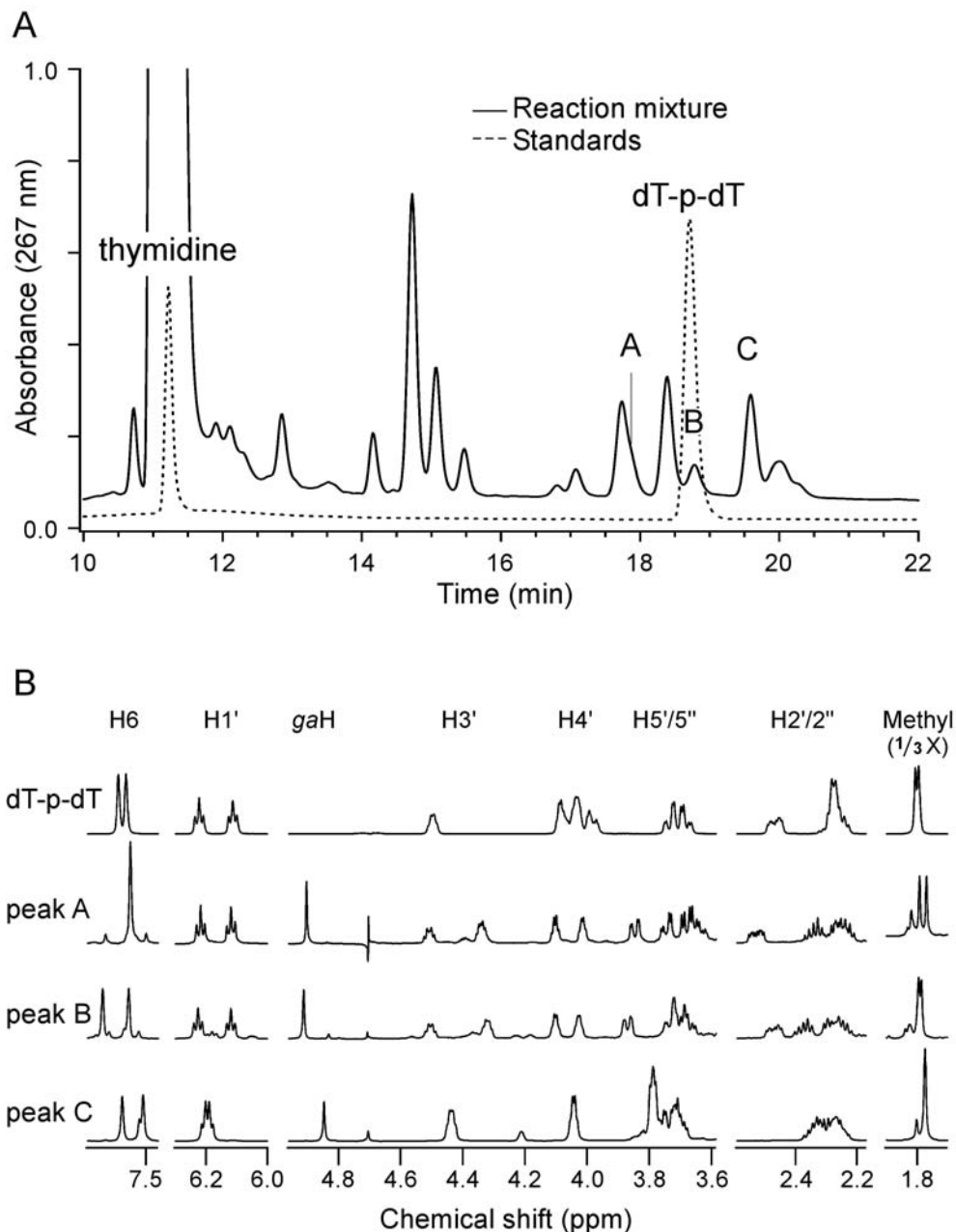
Figure 3.7: Chromatographic and spectroscopic analyses of *ga*-dinucleotide products of deoxythymidine reacted with glyoxylate under drying conditions. **(A)** Overlay of an HPLC chromatograph of a glyoxylate/thymidine reaction mixture (solid line) with a chromatograph of thymidine and dT-p-dT standards (dashed line). Seven peaks with relative mobilities similar to dT-p-dT, including those marked A, B, and C, were confirmed to be *ga*-dinucleotides by mass spectrometry. **(B)** Comparison of the ^{1}H spectrum of the dT-p-dT dinucleotide with spectra of the *ga*-dinucleotide products designated as HPLC peaks A, B, and C. The H3′ resonance of the 5′ nucleoside of dT-p-dT lies beneath the solvent peak (4.7 ppm), which was suppressed by a presaturation pulse.

Table 3.5: NMR data for peak A in D_2O

group[e]	position	δ_C, ppm	δ_H, ppm[a]
dT1	2	151.8	
	6	137.7	7.53 (s)
	1'	85.9	6.10 (t)
	2'[c]	37.2	2.23 (m)
			2.51 (m)
	3'	77.4	4.32 (m)
	4'	85.1	4.08 (m)
	5'[d]	61.4	3.65 (d,d)
			3.73 (d,d)
glyoxylate	G1 (CH)	100.4	4.89 (s)
	G2 (COO⁻)	173.9	
dT2	2	152.0	
	6	137.9	7.53 (s)
	1'	84.8	6.20 (t)
	2'[c]	38.7	2.23 (m)
			2.31 (m)
	3'	70.5	4.49 (m)
	4'	85.0	4.00 (m)
	5'[d]	64.7	3.64 (d,d)
			3.83 (d,d)
unassigned[b]	4	166.4	
	4	166.6	
	5	111.6	
	5	111.8	
	M	11.9	1.77 (s)
	M	11.9	1.74 (s)

[a] Multiplicities are indicated in parentheses
[b] Unable to assign these peaks to a particular nucleoside
[c] H2' is highfield, H2" is lowfield (Wood, *et al.*, 1974)
[d] H5' is lowfield, H5" is highfield (Wood, *et al.*, 1974)
[e] Group and position numbers refer to structure in Table 3.6

Table 3.6: HMBC correlations for peak A

group	proton	carbon
dT1	6	2, 4, M
	1'	2, 6
	2'	1', 3'
	2"	3'
	4'	3'
	5'	3'
	5"	3'
glyoxylate	G1	G2, dT2 5', dT1 3'
dT2	6	2, 4, M
	1'	2, 6
	2'	1', 3'
	2"	3'
	4'	3'
unassigned [a]	M	4, 5, 6
	M	4, 5, 6

[a] Unable to assign these peaks to a particular nucleoside
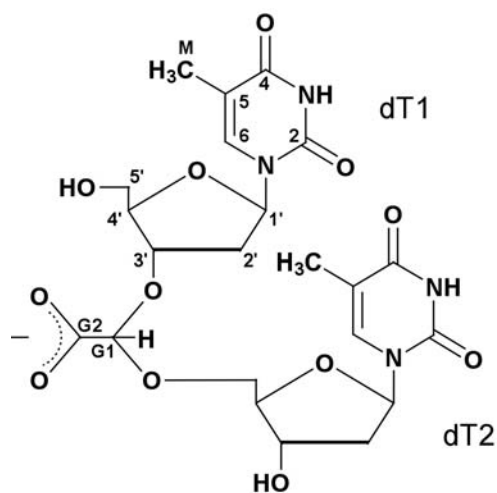
Table 3.7: NMR data for peak B in D$_2$O

group[e]	position	δ_C, ppm	δ_H, ppm[a]
dT1	2	152.2	
	4	167.4	
	5	112.4	
	6	138.4	7.53 (s)
	1'	86.6	6.09 (t)
	2'[c]	38.3	2.20 (m)
			2.44 (m)
	3'	76.4	4.29 (m)
	4'	85.9	4.08 (m)
	5'[d]	62.1	3.65 (m)
			3.72 (m)
glyoxylate	G1 (CH)	100.4	4.89 (s)
	G2 (COO$^-$)	174.3	
dT2	2	152.4	
	4	167.0	
	5	112.4	
	6	138.8	7.62 (s)
	1'	85.9	6.20 (t)
	2'[c]	39.1	2.25 (m)
			2.34 (m)
	3'	71.4	4.48 (m)
	4'	85.6	4.00 (m)
	5'[d]	66.7	3.69 (m)
			3.84 (d,d)
unassigned[b]	M	11.9	1.77 (s)
	M	11.9	1.78 (s)

[a] Multiplicities are indicated in parentheses
[b] Unable to assign these peaks to a particular nucleoside
[c] H2' is highfield, H2" is lowfield (Wood, *et al.*, 1974)
[d] H5' is lowfield, H5" is highfield (Wood, *et al.*, 1974)
[e] Group and position numbers refer to structure in Table 3.8

Table 3.8: HMBC correlations for peak B

group	proton	carbon
dT1	6	2, 4, M, 1'
	1'	2, 6
	2'	1', 3'
	2"	3'
	3'	4', G1
	4'	1', 3'
	5'	3', 4'
	5"	3', 4'
glyoxylate	G1	G2, dT2 5', dT1 3'
dT2	6	2, 4, M, 1'
	1'	2, 6
	2'	3'
	2"	1', 3'
	3'	4', 5'
	4'	3'
	5'	3'
	5"	3'
unassigned[a]	M	4, 5, 6
	M	4, 5, 6

[a] Unable to assign these peaks to a particular nucleoside
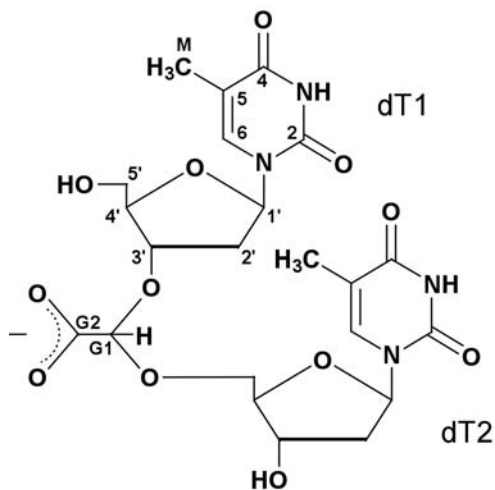
Table 3.9: NMR data for peak C in D$_2$O

group[e]	position	δ_C, ppm	δ_H, ppm[a]
glyoxylate	G1 (CH)	100.1	4.83 (s)
	G2 (COO$^-$)	173.6	
unassigned[b]	2	151.7	
	2	151.7	
	4	166.7	
	4	166.7	
	5	111.7	
	5	111.7	
	6	137.8	7.49 (s)
	6	138.0	7.56 (s)
	M	12.1	1.76 (s)
	M	12.1	1.76 (s)
	1'	85.5	6.17 (t)
	1'	85.5	6.18 (t)
	2'[c]	38.8	2.23 (m)
			2.32 (m)
	2'[c]	38.8	2.25 (m)
			2.30 (m)
	3'	71.5	4.42 (m)
	3'	71.5	4.42 (m)
	4'	85.3	4.03 (m)
	4'	85.3	4.03 (m)
	5'[d]	66.1	3.71 (m)
			3.77 (m)
	5'[d]	66.4	3.72 (m)
			3.74 (m)

[a] Multiplicities are indicated in parentheses
[b] Unable to assign these peaks to a particular nucleoside
[c] H2' is highfield, H2" is lowfield (Wood, *et al.*, 1974)
[d] H5' is lowfield, H5" is highfield (Wood, *et al.*, 1974)
[e] Group and position numbers refer to structure in Table 3.10

Table 3.10: HMBC correlations for peak C

group	proton	carbon
glyoxylate	G1	G2, pro-*R* 5', pro-*S* 5'
unassigned[a]	6	2, 4, 5, M, 1'
	M	4, 5, 6
	1'	2, 6, 3'
	2'	1', 3'
	2"	1', 3'
	3'	1', 3'
	4'	1', 3'
	5'	G1, 3', 4'
	5"	G1, 3', 4'

[a] Unable to assign these peaks to a particular nucleoside

pro-*R*   pro-*S*
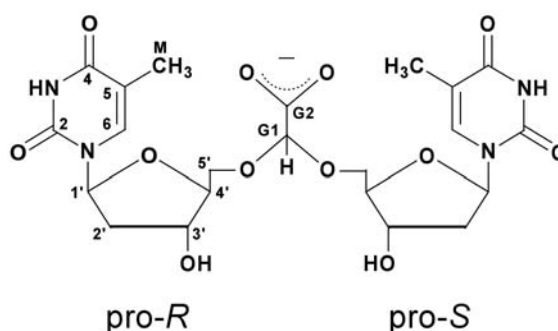
spectrometry (m/z = 297 in negative ion mode), and it is possible that some of these products represent ester-linked nucleosides, cyclic dinucleotides, the 3'-*ga*-3' dinucleotide, or *ga*-dinucleotides linked through one or both bases by iminal-type bonds.

3.3.6 Stability of dT-*ga*-dT Dinucleotide Linkages

Acetal bonds are known to be stable in neutral and basic pH conditions, but generally considered to be readily labile in an aqueous solution of acidic pH. Nevertheless, we have found that the acetal bonds formed by glyoxylate and nucleosides are relatively stable in aqueous solution, even at pH 1. When dissolved in water acidified

by HCl to pH 1, no appreciable degradation of dT-*ga*-dT dinucleotides was observed

after a two-week incubation at 22 °C (Figure 3.8). Under the same acidic conditions at 60

°C, considerable degradation of dT-*ga*-dT into its components was observed. The

addition of 1 M MgCl$_2$ to the acidic *ga*-dinucleotide solution was found to accelerate

degradation at 60 °C, but had no visible effect at 22 °C. Under neutral conditions (pH 7)
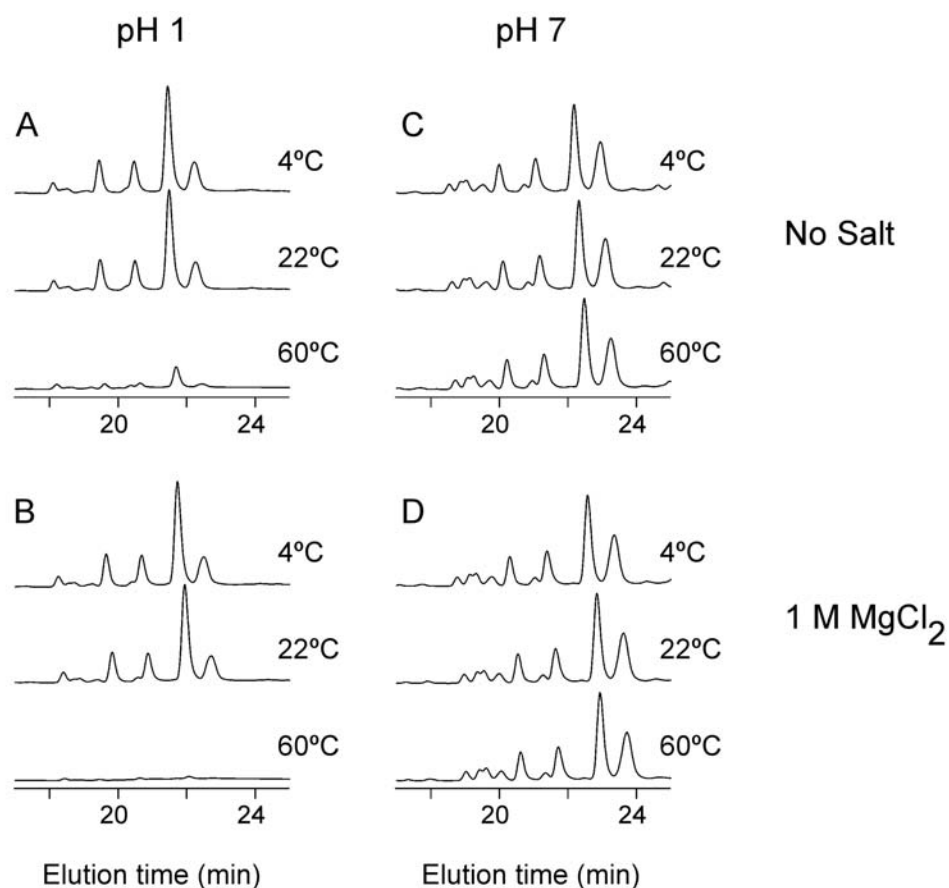
Figure 3.8: HPLC chromatographs illustrating the relative stability of dT-*ga*-dT *ga*-dinucleotides under various solution conditions. Dinucleotides were prepared by drying under vacuum at room temperature solutions of 20 µmol dT, 10 µmol glyoxylic acid, and 15 µmol HCl in 200 µl dH$_2$O. After drying, dT-*ga*-dT reaction solutions were resuspended at pH 1 or pH 7, in the presence or absence of 1 M MgCl$_2$, and then incubated for two weeks at 4, 22, or 60 °C. At day 14, samples were diluted five-fold by a 0.25 N NH$_4$OH solution prior to HPLC analysis. **(A)** Samples resuspended at pH 1 without MgCl$_2$. **(B)** Samples resuspended at pH 1 with 1 M MgCl$_2$. **(C)** Samples resuspended at pH 7 without MgCl$_2$. **(D)** Samples resuspended at pH 7 with 1 M MgCl$_2$.

neither heating to 60 °C nor incubation with 1 M MgCl$_2$ for two weeks resulted in appreciable *ga*-dinucleotide degradation (Figure 3.8).

3.3.7 Modeling Studies of *ga*RNA Electrostatics and Duplex Structure

As mentioned above, the CH carbon of glyoxylate is prochiral, and therefore 3′-*ga*-5′ linked *ga*-dinucleotides can form as either the *R* or *S* isomers (Figure 3.9). In the absence of a mechanism that specifically enhances the formation of one isomer over the other, one would expect little selectivity in a prebiotic world. Formation of *ga*-dinucleotides with non-identical components raises the possibility that mixing glyoxylate-acetal *R* and *S* isomers *external* to a *ga*-oligonucleotide backbone could destabilize duplex formation, but this is very different from mixing ribose stereoisomers *internal* to the backbone, where dramatic effects are to be expected. For example, Watson-Crick base pair formation is not possible for a single L-ribose nucleotide incorporated within an otherwise D-ribose nucleotide duplex, as the Watson-Crick face of the base attached to the L-ribose is rotated approximately 100° away from the duplex helical axis (*101*). In contrast, switching between the *R* and *S* isomers of glyoxylate is not expected to directly affect bond angles between the backbone and the nucleoside bases.

Molecular modeling was used to investigate the potential for mixed *R* and *S* isomers of glyoxylate-acetal linkages to interfere with *ga*NA duplex formation. Although determining the specific structural details of a *ga*-oligonucleotide duplex is beyond the present capabilities of computer-based molecular modeling, molecular dynamics simulations have successfully reproduced experimentally observed changes in the helical structures of nucleic acids, such as the DNA B-form to A-form transition (*102*). Thus, we have performed modeling studies with the Amber force field (*95*) as an initial means to

73

test if a *ga*RNA duplex containing mixed *R* and *S* isomers of the glyoxylate-acetal linkage would be expected to maintain a double helical structure resembling that of A-form RNA.

Partial charges for the atoms of the glyoxylate-acetal linkage were required to perform energy minimization and molecular dynamics simulations with the Amber force field (*95*). These partial atom charges were determined by an *ab initio* quantum mechanical calculation (Section 3.2.5). The partial charges of the four oxygen atoms of a glyoxylate-acetal linkage are predicted to be very close to those of the corresponding atoms of a phosphodiester linkage (Figure 3.9A). Additionally, the two carbon atoms of glyoxylate and the CH hydrogen in this linkage are predicted to have partial positive charges with a sum total that is approximately equal to that of the phosphorus atom of a phosphodiester linkage (Figure 3.9A).

Figure 3.9: Comparison of computer generated models of RNA and *ga*RNA. **(A)** Ball and stick models of dinucleotides connected by a phosphate (RNA), an *R*-glyoxylate-acetal, and an *S*-glyoxylate-acetal linkage, respectively. For clarity, no hydrogen atoms are shown, except the CH hydrogen of glyoxylate. Partial electrostatic atomic charges, based upon quantum mechanical calculation (Section 3.2.5), are shown for the atoms of the phosphate linkage and the *S*-glyoxylate-acetal linkage. Partial electrostatic atomic charges assigned to the *R*-glyoxylate-acetal linkage were the same as those shown for the *S*-glyoxylate-acetal linkage. **(B)** Space-filling representation of the RNA duplex $[A_6U_6]_2$ in a canonical A-form helix. **(C)** Space-filling representation of an energy minimized model (Section 3.2.5) of the *ga*RNA duplex with the self-complementary sequence A_6U_6, with a mixture of the *R*- and *S*-glyoxylate-acetal linkages along each strand (Section 3.2.5). **(D and E)** Vector model of backbone and base pair geometries generated by the Curves program (*9*) for the models shown in B and C. For the *ga*RNA model (E), backbone linkages of the *R*-glyoxylate-acetal nucleotides are shown in red, backbone linkages of the *S*-glyoxylate-acetal nucleotides are shown in blue. Note that the helix of *ga*RNA (models C and E) is slightly underwound from the canonical A-form helix (models B and D). Also note that the backbone and base pair vectors of the *ga*RNA duplex indicate that the helical form of *ga*RNA does not vary appreciably between *R*- and *S*-glyoxylate-acetal linkages. Models in D and E are rotated 90° from those shown in B and C.

rA-p-rU

rA-ga-rU
(R isomer)

rA-ga-rU
(S isomer)

B

C

D

E
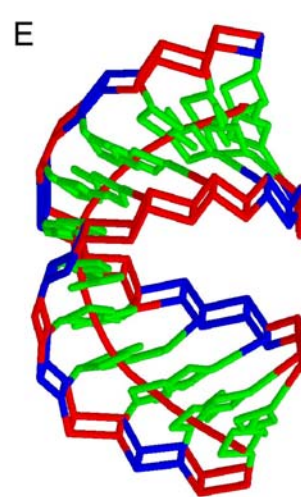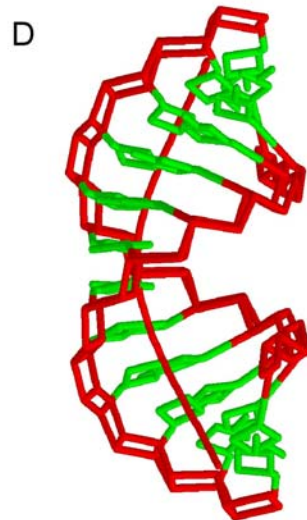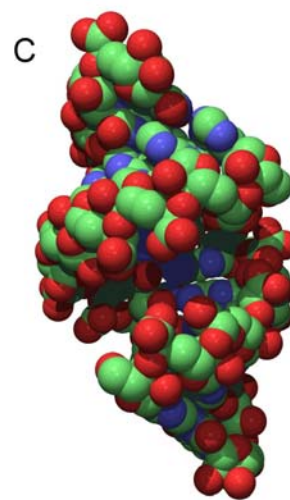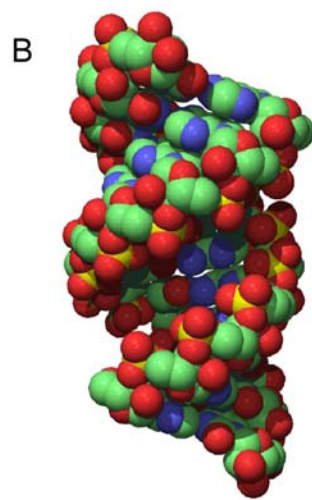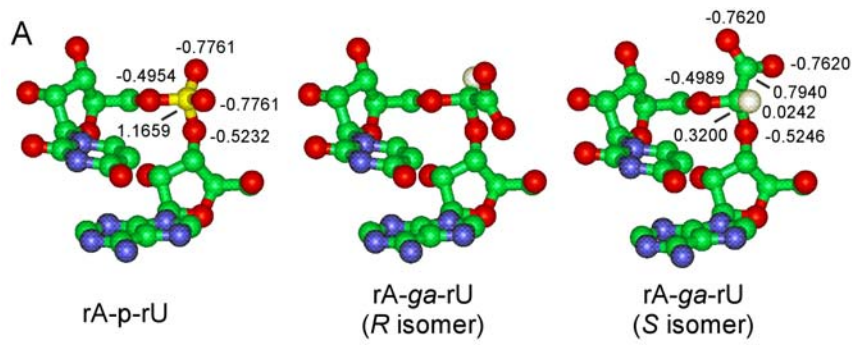
An initial model of a *ga*RNA duplex was constructed by changing the phosphate groups to glyoxylate in a model RNA double helix in the canonical A-form. The duplex was twelve base pairs in length, with the self-complementary sequence A_6U_6. *R* and *S* isomers of the glyoxylate-acetal linkages between nucleotides were varied along the two strands (Section 3.2.5). In a periodic solvent box with explicit water molecules and Na^+ counterions, the *ga*RNA duplex model was energy minimized and subjected to 1 ns of molecular dynamics simulation (Section 3.2.5). The model retained a double helix structure similar to that of RNA, except for a partial unwinding of the helix (Figures 3.9B and C). This modeling study demonstrated that helix structure and base pair geometry within a *ga*-oligonucleotide duplex does not depend significantly on the particular isomer of a glyoxylate-acetal linkage, and is not affected by the coexistence of *R* and *S* linkages within the same backbone (Figure 3.9E). The similarities between a *ga*RNA duplex and a natural RNA duplex also suggests that it should be possible for *ga*RNA and RNA strands to form hybrid duplexes, which would be essential for the continuity of genetic transfer from one nucleic acid system to another during the evolutionary process.

3.4 DISCUSSION

The RNA world hypothesis presents an attractive scenario for a very early stage of life (*1*). However, wide acceptance of the RNA world hypothesis will ultimately require the discovery of a process by which RNA polymers, or RNA-like polymers, form spontaneously from plausible prebiotic molecules. The formation of the nucleotide bases and ribose from simple precursors (*55*), and the coupling of purine and pyrimidine bases and ribose upon heating in the presence of divalent cations (*8*), give reason to believe that the nucleosides are plausible prebiotic compounds. In contrast, the abiotic coupling of

nucleotides into RNA polymers has only been achieved with chemically activated mononucleotides, and the activation chemistries employed are hard to imagine in the prebiotic world (*84*). If RNA was preceded by another informational polymer, it stands to reason that this polymer must have been easier to form under prebiotic conditions. Furthermore, the structure of this polymer would have been sufficiently close to RNA that molecular evolution would have allowed it to eventually and continuously evolve into RNA. By replacing phosphate with glyoxylate we have introduced a thermodynamically favorable and plausibly prebiotic ancestor of the phosphodiester linkage of RNA. The glyoxylate group is also remarkably close in structure and electrostatic charge distribution to that of a phosphate group. Here we discuss the additional reasons why *ga*NA is attractive as a prebiotic ancestor of RNA.

3.4.1 Prospects for *ga*NA Polymer Formation

We have shown that drying nucleosides and glyoxylate from an aqueous solution containing divalent metal cations yields approximately 1% *ga*-dinucleotides. While these experiments demonstrate a plausible prebiotic method for the formation of an RNA-like linkage, we have not yet realized the spontaneous formation of *ga*-oligonucleotides. Experiments in which samples containing nucleosides and glyoxylate were subjected to repeated drying-rehydration cycles did not result in increased *ga*-dinucleotide yield or the growth of *ga*-oligonucleotides (data not shown). It appears that, under the conditions of the experiments presented, the observed 1% yield represents the equilibrium amount of *ga*-dinucleotide formed shortly before the point at which one component of the reaction mixture precipitates or crystallizes from solution.

The thermodynamic parameters we have determined for glyoxylate-acetal formation indicate that *ga*-oligonucleotide formation should proceed spontaneously if *ga*-nucleotides are preorganized prior to the point of sample drying when acetal bond formation is catalyzed by cationic metal hydrates. Solution state experiments with glyoxylic acid in H_2O/methanol mixtures revealed the equilibrium constant between a hemiacetal and an acetal to be 1.5, corresponding to a free energy of -0.5 kcal/mol at 65 °C (Section 3.2.1). Thus, for free *ga*-nucleotides in solution, formation of acetals between two nucleotides only becomes favored over the *ga*-nucleotide hemiacetal state when nucleotide concentration becomes equal to or greater than that of water. However, if a surface or a macromolecular assembly ordered *ga*-nucleotides during the course of sample drying, then the effective local concentration of ribose OH groups to form acetal linkages would be increased dramatically. Preorganization of *ga*-nucleotides, for example by base stacking on an intercalating molecule (*2*), would eliminate the entropic penalty of bringing together two *ga*-nucleotides in solution prior to acetal formation while the entropic gain of releasing a water molecule during acetal formation at each *ga*-dinucleotide step is retained. Additionally, under drying conditions, any water molecules formed upon acetal bond formation could be lost to evaporation, essentially blocking the reverse reaction of acetal hydrolysis and driving the equilibrium towards acetal formation. The sudden addition of water after the acetal has been formed, particularly if the addition takes place at a lower temperature, will quench the reaction. The acetal will not hydrolyze because the hydrolysis rate is then too low, and equilibrium conditions are no longer operative, unless inordinately long times are considered. At 22 °C these acetal bonds are quite stable, with no observable degradation after two weeks at room

temperature in neutral aqueous solution (Figure 3.8). However, at elevated temperatures the glyoxylate linkages become more labile, providing a temperature regime under which the bonds are reversible and equilibrium controlled replication could take place. Therefore, glyoxylate-linked backbone satisfies our vision of a proto-RNA backbone that is structurally similar to RNA, but much more easily assembled and perhaps replicated.

Ferris and co-workers have demonstrated that drying the chemically activated nucleoside 5′-phosphorimidazolide-adenosine (ImpA) onto the surface of a montmorillonite clay leads to the elongation of a pre-adsorbed decanucleotide by up to 30 additional nucleotides, whereas the same decanucleotide is not elongated by ImpA in solution or when dried in the absence of the clay surface (*60*). A similar catalytic inorganic surface may exist that facilitates the polymerization of *ga*-nucleotides. We have previously proposed that during prebiotic times planar, organic molecules, similar to those known to intercalate the bases of DNA and RNA, may have organized the nucleoside bases into assemblies that facilitated the formation of RNA-like polymers (*2, 64*). The chemistry we have described here for *ga*NA synthesis would also be compatible with this suggested mode of polymer assembly and replication by "molecular midwives" (*2*). Additionally, preorganization of the bases or nucleosides into columnar stacks of base pairs, triples, or tetrads prior to backbone synthesis would protect the bases from iminal bond formation with glyoxylic acid.

3.4.2 The Advantage of Reversible Backbone Linkages in Early Evolution

Template-directed synthesis using activated nucleoside monophosphates (e.g., 5′-phosphorimidazolides) is highly prone to nucleotide misincorporation, the production of truncation products, and strand cyclization (*55-57*). Such processes with high rates of

mutation would have been catastrophic to early life (*55*). In contemporary life, error correction occurs through the reversible association of the incoming nucleotide with the growing product strand (Figure 1.6) and through the proofreading mechanism of polymerase enzymes that excise misincorporated nucleotides (i.e., errors) from a product strand that do not form Watson-Crick base pairs with the template strand, providing the polymerase with another opportunity to attach the correct nucleotide. Since proofreading protein polymerases likely appeared after the era of the RNA world, it would have been advantageous for early self-replicating systems to have had some other mechanism by which replication errors were corrected. As discussed in the Chapter 1, Lynn and co-workers have shown that backbone linkages with reversible covalent bonds can provide a form of thermodynamically-controlled proofreading during template-directed synthesis (*70*). Correct base pairing between a template strand and substrates (e.g., mononucleotides, dinucleotides) is ultimately selected from incorrect pairings based upon the lower energy of a template paired with a product strand containing all Watson-Crick base pairs. With this approach, Lynn and co-workers have achieved both sequence-specific and chain-length-specific reading of DNA templates into product strands using DNA-like mononucleosides with backbone coupling chemistries that form reversible imine bonds, prior to conversion to stable amine linkages upon reduction (*70*).

The reversible formation of acetal bonds in an aqueous environment could have likewise provided a means for proofreading during the template-directed synthesis of *ga*-oligonucleotides (*2*). Based upon the *ga*-dinucleotide formation and degradation data presented in the current study, glyoxylate-acetal bond formation and cleavage at neutral or basic pH in a drying sample is apparently only catalyzed by divalent cations close to

80

the point at which liquid water is completely removed. Therefore, an existing *ga*NA strand could be long-lived in an aqueous solution containing metal hydrates. As a solution of *ga*NA strands, nucleosides, and glyoxylate progressed towards dryness, intercalating molecular midwives could have promoted base pairing between free nucleotides and the parent *ga*NA polymer before acetal bond cleavage/formation commenced. If this multimolecular assembly remained intact as liquid water was completely removed, then acetal bond formation could have been catalyzed along the backbone of the daughter strand nucleotides without loss of the parent *ga*NA polymer.

3.4.3 *ga*NA and RNA Evolution

A molecule as functionally versatile as RNA has likely undergone more than one stage of molecular evolution. The proposed role of glyoxylate as a precursor to phosphate in modern nucleic acids would, as illustrated above, alleviate some serious difficulties associated with the prebiotic synthesis of RNA. Nevertheless, the origin of nucleosides, including problems associated with ribose synthesis and glycosidic bond formation (Chapter 2), remain separate and formidable questions regarding the origin of RNA. Eschenmoser and co-workers have suggested that alternative sugars may have preceded ribose in an ancestor of RNA (*36, 37*). Orgel and co-workers have even proposed that RNA may have evolved from a polymer containing glycerol-derived acyclonucleosides, a suggestion aimed at alleviating perceived problems associated with RNA synthesis from a racemic mixture of ribonucleotides (*103*). Our proposal of glyoxylate as an ancestor of phosphate is compatible with these and other proposals for the molecular evolution of RNA, as long as the proposed nucleosides contain at least two alcohol groups for the formation of acetal linkages.

Even if glyoxylate-acetal linkages are ancestors to the phosphodiester linkage, there may have been even earlier linkage molecules than glyoxylate. Glycolaldehyde, a widely accepted prebiotic molecule, would be a possible ancestor to glyoxylate. Just as we have shown that glyoxylate can link nucleosides in a plausible prebiotic reaction, glycolaldehyde and other aldehydes would be expected to form similar linkages under similar reaction conditions (see Chapter 4 for supporting data). It is even possible that the earliest RNA-like polymers had backbones that contained a mixture of aldehyde linkage groups and a variety of nucleoside sugars (*2*) before Darwinian evolution gradually resulted in the appearance of polymers with backbones of uniform structure and chemical composition.

It is important to note the difference between the kinetic stabilities of *ga*-nucleotides and nucleotide monophosphates, and the possible role that this could have played in evolution. The phosphoester bond of a monophosphate nucleotide is kinetically stable to hydrolysis, with a half-life in the absence of catalysis on the order of a thousand years at room temperature (*104, 105*). In contrast, the hemiacetal bond of a *ga*-nucleotide is freely labile in the presence of water (at neutral, acidic, or alkaline pH), and a *ga*-nucleotide in aqueous solution will exist in equilibrium with its nucleoside and glyoxylate. Thus, a *ga*-nucleotide could not function in the many roles that are played by nucleotide monophosphates (or polyphosphates) in modern life. For example, the kinetic stabilities of phosphoester and polyphosphate linkages are of paramount importance for the transfer and utilization of energy within all living cells. Thus, the positive attributes of nucleotide kinetic stability would have been a strong selective pressure that favored replacement of glyoxylate by phosphate.

CHAPTER 4

OBSERVATIONS ON CONDENSATION REACTIONS

4.1 INTRODUCTION

The focus of much of the work discussed in previous chapters has been achieving nucleoside and dinucleotide synthesis under prebiotically plausible conditions, with tolerance for low yields, as even a small amount of product is an achievement in this field of exploration. Through experimentation with condensation reactions data have been obtained that illuminate the limitations and potentials of prebiotic proto-RNA syntheses. The correlations between product yield and thermodynamics are discussed, with supporting evidence from aNA syntheses using ketones as linking molecules. Reactant solubility plays a very large role in the yields of heating-drying reactions, which will be demonstrated through comparisons of gaNA yields of rU versus dT. Also, a strong dependence of the zebularine and aNA product yields on reaction kinetics has been observed for the non-equilibrium reaction conditions used. Finally, evidence for the selective influence of salts on the distribution of 2-pyrimidinone nucleoside anomers is presented.

4.2 MATERIALS AND METHODS

4.2.1 gaNA Reactions

Homo dinucleotides of dT, dC, dA, dG, rU, rC, rA, and rG were synthesized by drying under vacuum at room temperature solutions of 20 μmol nucleoside, 10 μmol glyoxylic acid, and 15 μmol HCl in 200 μl dH$_2$O. Mass spectra of the reaction products

were collected in negative ion mode on a Micromass Quattro LC mass spectrometer and ionized under electrospray conditions (capillary voltage, 2.5 kV; cone voltage, 40 V).

4.2.2 Pyruvate Acetal Equilibria

NMR spectra were collected for samples containing 100 mM methyl pyruvate (CH_3-CO-$COOCH_3$) in 0, 25, 50, and 75% (v/v) MeOD in D_2O. The 100% D_2O sample had a pD = 3.2. The spectra were recorded on a Bruker DRX500 spectrometer at 25 °C. The percent of each species was determined by dividing the NMR peak height by the sum of the peak heights for the ketone, hydrate, and hemiacetal species.

4.2.3 Pyruvic Acid-Linked *a*NAs

Pyruvate-linked acetal nucleic acids (*pa*NAs) were synthesized and analyzed as described for gaNA homodinucleotides in Section 4.2.1 using dT and rU nucleosides.

4.2.4 Kinetics of Acetal Formation

The kinetics of glyoxylic acid and glycolaldehyde acetal formation were determined by NMR experimentation (Figure 4.9). The percentage of aldehyde as acetal was calculated by dividing the integral for the acetal species by the sum of the integrals for hydrate, hemiacetal, and acetal species. For glyoxylic acid, both the acid and the methyl ester were observed for each species (Figure 4.9a), and both sets of integrals were included in the calculation. The percentage of aldehyde as acetal was plotted versus time and the results were fit with an exponential to obtain the time constant.

The necessity for accurate integrals required obtaining relaxation data through inversion recovery experiments. In the absence of a relaxation agent, T_1 values for the aldehydic proton of the glyoxylic acid exceeded 10 s at 80 °C. Therefore, $CuSO_4$ was

used in order to shorten the cycle time of each NMR experiment. Table 4.1 contains the results of the T_1 experiment and the NMR experimental parameters that were used for obtaining rate constants for acetal formation. The sample conditions for obtaining the T_1 data were 400 mM glycolaldehyde or glyoxylic acid, 25% (v/v) MeOD, 75% D_2O, 2% DCl, and 100 μM $CuSO_4$. For the kinetics experiments, the final sample concentrations were 200 mM in aldehyde, 25% (v/v) MeOD, 75% D_2O, pD = 1.6, and 100 μM $CuSO_4$. Kinetics data were collected at four temperatures for glycolaldehyde and three for glyoxylate (Table 4.1). The samples were prepared without MeOD and allowed to equilibrate to the NMR probe temperature for a minimum of five minutes. The NMR was tuned, locked, and shimmed and the first spectrum was collected without MeOD. The sample was then removed from the NMR, MeOD was added and the series of NMR experiments was started with a preset delay to allow for reinserting the sample tube into the NMR.

Table 4.1: NMR experimental parameters for measuring acetal kinetic as determined by inversion recovery experiments.

Aldehyde	Temp. (K)	Longest T_1 (s)	$5*T_1$ (s)	kinetics experiment parameters		
				d1 (s)	aq (s)	cycle time (d1 + aq) (s)
Glycolaldehyde	287.8	1.34	6.7	2.0	5.45	7.5
	298.1	1.83	9.2	4.0	5.45	9.5
	307.9	2.752	13.8	8.5	5.45	14.0
	318.0	4.059	20.3	15.0	5.45	20.5
Glyoxylate	333.0	6.631	33.2	27.7	5.45	33.2
	343.0	8.481	42.4	37.0	5.45	42.5
	353.0	7.169	35.8	30.4	5.45	35.9

4.3 THERMODYNAMICS, EQUILIBRIA, AND SOLUBILITY

4.3.1 Ketones as Linking Molecules in *a*NAs

As discussed in Chapter 3, one of the reasons that acetals are attractive as proto-RNA linking molecules is that hemiacetal (or nucleotide) formation is enthalpically favored over the hydrate by an average of 4 – 5 kcal/mol, and acetal (or dinucleotide) formation is favored over the hemiacetal by another 4 – 5 kcal/mol (*92, 106*). This is a general trend for all acetals (*92, 106*). However, Wiberg and coworkers report that the equilibrium amount of hydrate, hemiacetal, and acetal are highly sensitive to the substituent group on the aldehyde (*92*). Translated to *a*NA synthesis, the yield of dinucleotide will be dependent upon the aldehyde linker used.
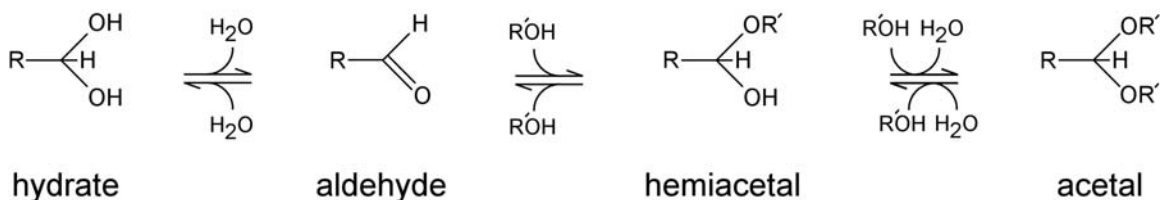
Figure 4.1: Equilibria of aldehydes with their hydrate, hemiacetal, and acetal species.

The amount of unreacted aldehyde correlates with steric effects such as strain and bond eclipsing and the electron donating or withdrawing character of the aldehyde R group (Figure 4.1) (*92, 106*). There is a positive correlation between the energy of the nucleophilic additions of water or an alcohol in the formation of aldehyde hydrates or hemiacetals, respectively, with the electropositivity of the carbonyl carbon. That is, the more electropositive the carbon, the greater the enthalpy of aldehyde hydrate or hemiacetal formation (*92*). The neighboring partial electronegative charge on the OH of

hydrates or OR′ group of hemiacetals stabilizes the partial positive charge on the carbonyl (*106*). Alkyl substituents on the aldehyde stabilize the carbonyl by donating electron density, and this stabilizing effect is greatly reduced for hydrates (*106*). Therefore, hydrate formation is less favored with an increase in alkyl substitutions (acetaldehyde: $K_{hydrate}$ = 1.25; acetone: $K_{hydrate}$ = 1.75 x 10^{-3}, Equation 4.1, 25 °C) (*92, 106*). This reasoning can be extended to the reduced equilibrium constants for hemiacetal formation with increased alkyl substitution (acetaldehyde: $K_{hemiacetal}$ = 18.0; acetone: $K_{hemiacetal}$ = 7.32 x 10^{-3}, Equation 4.2, 25 °C) (*92*).

Because alkyl substitutions reduce the equilibrium amount of hemiacetal formed, acetal production is also reduced, and therefore ketones are generally poor acetal-forming molecules (acetaldehyde: acetal completely formed; acetone: K_{acetal} = 7.68 x 10^{-3}, 25 °C) Exceptions to the rule of poor acetal formation by ketones can be found. Substituting one of the R groups of a ketone with a strongly electron-withdrawing group such as CF_3 or CCl_3 will lead to complete hydration (*106*), and as hemiacetal and acetal formation are favored over hydration, acetal formation is also complete in acidic alcoholic solutions. The positively charged carbons of these R groups destabilize the ketone carbonyl, which favors nucleophilic addition by water or alcohol (*106*). The same effect is seen upon the substitution of electron-withdrawing groups on aldehydes, as observed for glyoxylic acid, which is almost completely hydrated in water whereas acetaldehyde is not. Therefore, we hypothesized that the ketone, pyruvic acid (CH_3-CO-COOH), may be a good candidate for *a*NA bond formation.

*ga*NA nucleic acids were calculated to be able to form a duplex with helical properties that are very similar to phosphodiester-linked duplexes, but with a slight under

winding (Figure 3.9). The under winding could be caused by the reduced bulk about the aldehydic carbon in comparison to the phosphate (Chapter 3, Figure 3.3). In this light, pyruvic acid could be a better isosteric replacement for phosphate than glyoxylic acid (Figure 4.2). It has an additional bulky group (CH_3) on the acetal carbon, which is a closer size mimic than the proton of glyoxylate for the oxygen atom of phosphate. With the enhanced steric bulk about the acetal carbon, the energetically-favored conformational space of a pyruvate-linked dinucleotide would be more restricted than that of a *ga*-dinucleotide, and could more closely match the favored conformations of phosphodiester-linked nucleic acids.
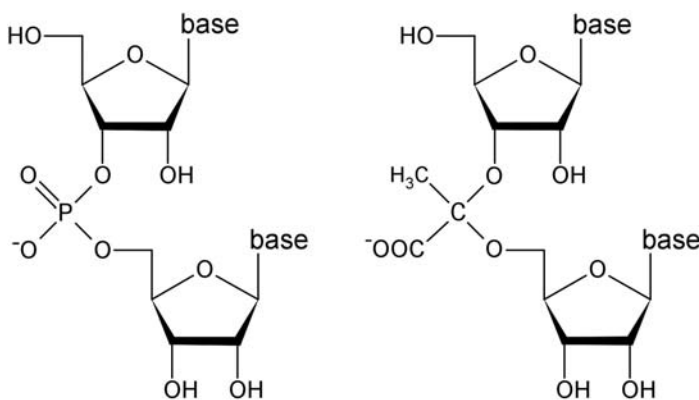
Figure 4.2: Comparison of the phosphodiester-linked RNA backbone (left) and the pyruvate-acetal backbone (right).

The ability of pyruvic acid to form hydrates and hemiacetals was established by NMR, as described in Section 4.2.2 (Figure 4.3). Although pyruvic acid does not undergo these nucleophilic additions as readily as glyoxylic acid, the equilibrium constant for hemiacetal formation was calculated to be 0.3 M^{-1}, which is approximately thirty times greater than for acetone (*92*). Next, acetal bond formation with rU and dT
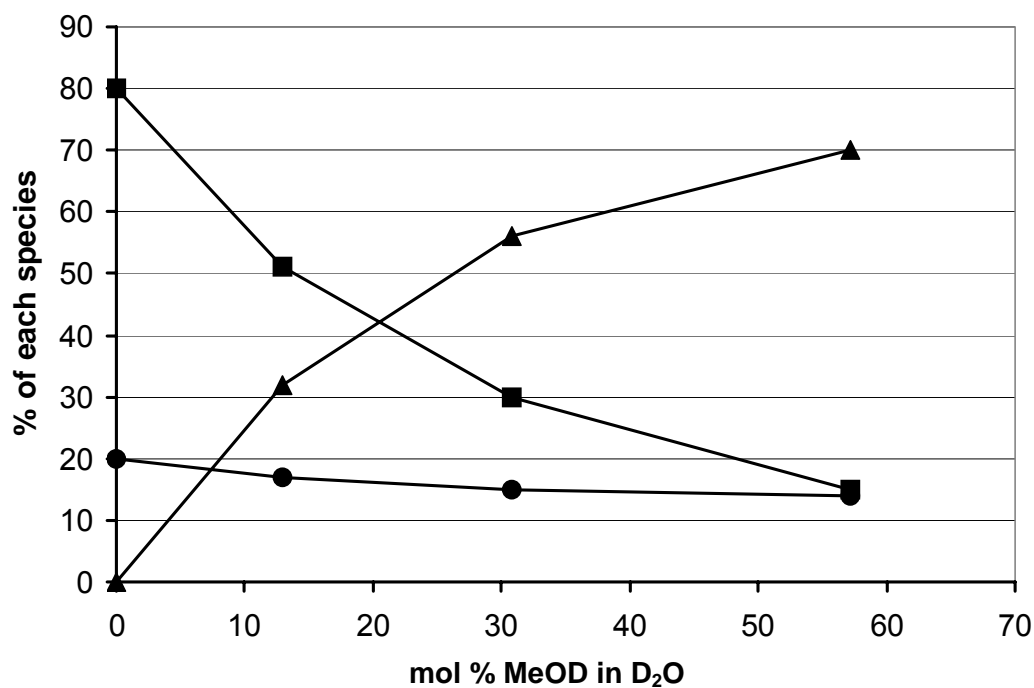
Figure 4.3: Equilibria of methyl pyruvate in water-methanol mixtures. In water methyl pyruvate exists primarily as the hydrate (squares), but 20% ketone remains (circles). As the mole percent of MeOD is increased, the hemiacetal concentration increases (triangles), and both ketone and hydrate concentrations decrease.

nucleosides was confirmed in heating-drying experiments as described for *ga*-dinucleotide formation in Section 3.2.4. Mass spectrometry analysis proved the existence of species with the same mass as cyclic *pa*-dT and *pa*-rU nucleotides (m/z = 311, 313 Da, respectively), as well as dT-*pa*-dT and rU-*pa*-rU dinucleotides (m/z = 553, 557 Da, respectively). However, the yield of the pyruvate-linked dinucleotides is approximately an order of magnitude less than the yields observed for the *ga*-dinucleotides, as determined by HPLC-UV (Figure 4.4). This result correlates with the expected trend for the substitution of a methyl group to the glyoxylic acid aldehydic carbon. Although the low yield of pyruvate-linked dinucleotides under the acid-catalyzed heating-drying

reactions suggests that pyruvate is not a plausible proto-RNA linker, modulating reaction conditions that affect nucleoside solubility (Section 4.3.2) and acetal reaction kinetics (Section 4.4.1) could boost *pa*-dinucleotide yields.
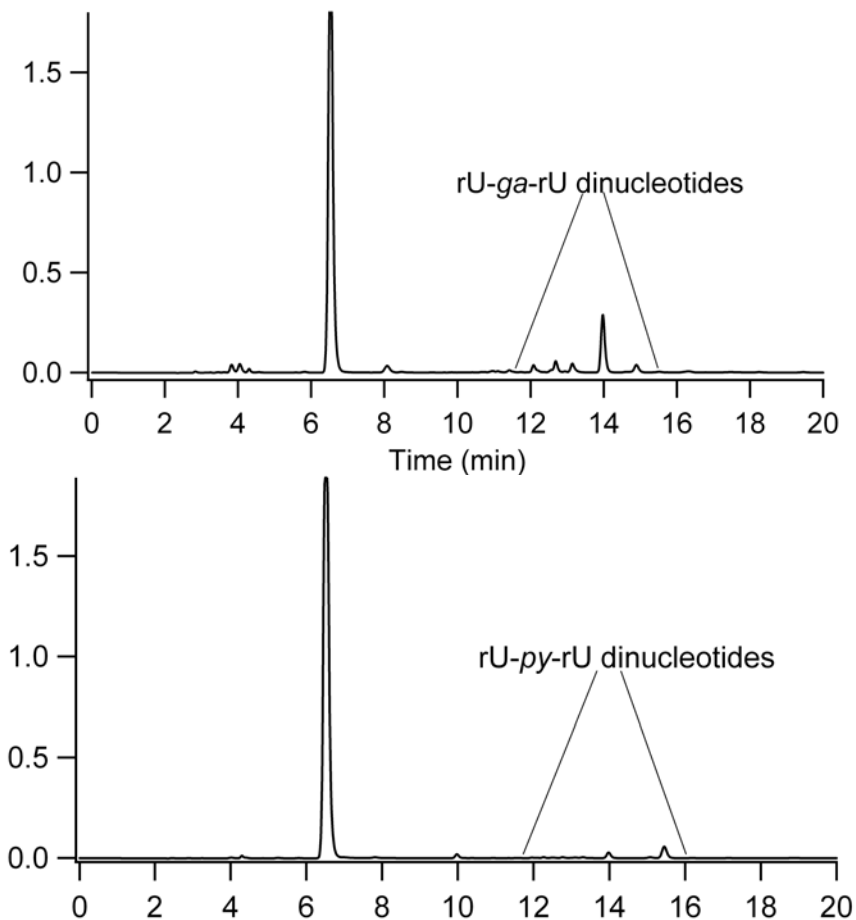
Figure 4.4: Comparison of glyoxylate and pyruvate *a*NA syntheses. HPLC-UV chromatographic analysis shows that, under the same reaction conditions, rU-*ga*-rU yields are approximately ten-fold higher (top) than rU-*pa*-rU (bottom).

4.3.2 Solubility and Condensation Product Yield

Evaporating an aqueous reaction mixture to dryness was the means by which the *ga*NA dinucleotides and the 2-pyrimidinone nucleosides were synthesized, but with limited success (approximately a 3% yield for all of the dT-*ga*-dT dinucleotides in an

$$[\text{nucleoside}] = K \frac{[\text{2-pyrimidinone}] \, [\text{ribose}]}{[H_2O]} \qquad \text{Equation 4.1}$$

$$[\text{dinucleotide}] = K \frac{[\text{nucleoside}]^2 \, [\text{glyoxylate hydrate}]}{[H_2O]^2} \qquad \text{Equation 4.2}$$

acid-catalyzed reaction and a 17% yield of the nucleosides). The yield of a condensation reaction is directly related to the ratio of reactant to water, as demonstrated by the equations for 2-pyrimidinone nucleoside formation (Equation 4.1) and *ga*NA dinucleotide formation (Equation 4.2). In order to drive the reaction toward nucleoside or dinucleotide formation, the water is removed, concentrating the reactants. However, in a drying reaction from an aqueous solution, the ratio of reactant to water reaches an upper limit, which is the aqueous solubility of the reactant; therefore the yield of the condensation product is governed by reactant solubility. This is born out in the syntheses of the homodinucleotides of dT, dC, dA, dG, rU, rC, rA, and rG, nucleosides that range in aqueous solubility by three orders of magnitude, with rU being the most soluble at 1.8 M, and dG being the least soluble at less than 2 mM at room temperature. When drying nucleosides and glyoxylate from an aqueous solution containing a Lewis acid catalyst, only the five most soluble nucleosides, rU, dT, rC, dC, and rA, produced dinucleotides in observable yields as determined by ESI- mass spectrometry (Chapter 3, Table 3.2). Presumably for the three least soluble nucleosides, water evaporated from the reaction mixture until the solubility limit of the nucleoside was reached, at which point the ratio of nucleoside to water remained fixed, and the nucleoside to water ratio was too low for detectable dinucleotide formation. The correlation between dinucleotide yield and
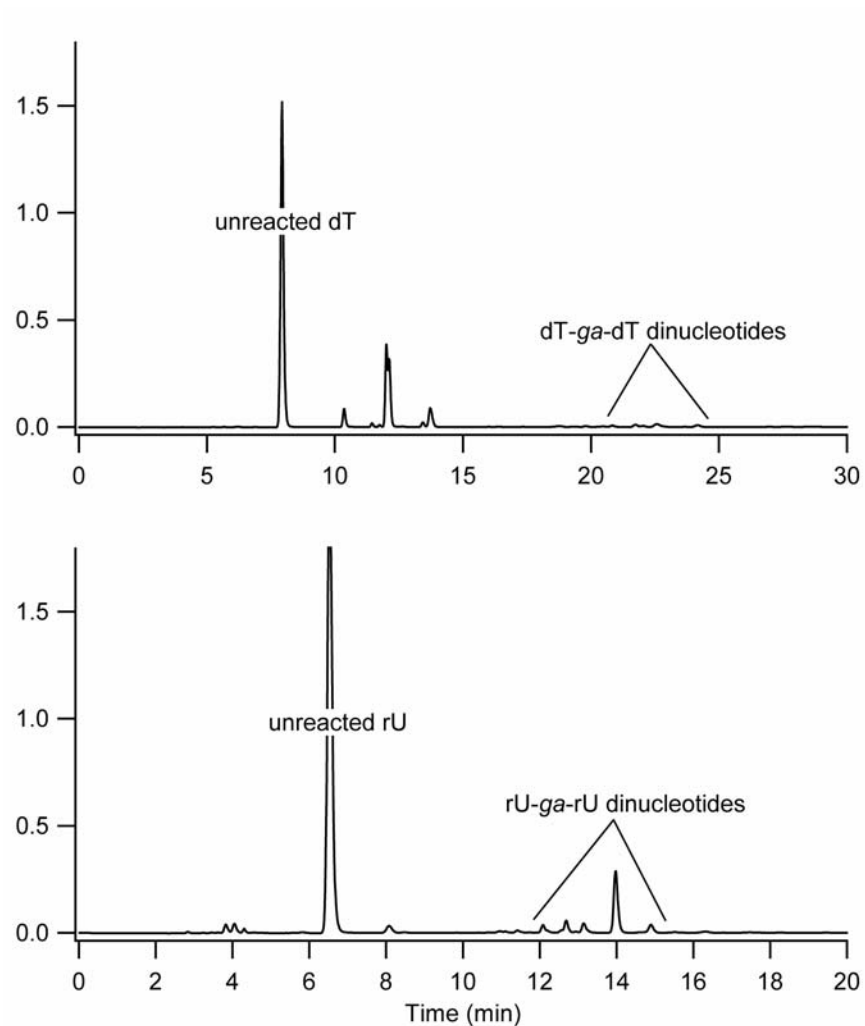
Figure 4.5: HPLC chromatographs showing the relative yields of dT-*ga*-dT (top) and rU-*ga*-rU (bottom) dinucleotides at approximately 1.5 and 15%, respectively.

nucleoside solubility is further supported by the fact that the yield of rU dinucleotides is roughly ten-fold higher than that of dT dinucleotides (Figure 4.5), which directly corresponds to the relative aqueous solubilities of the two nucleosides.

The limited solubility of thymidine in the nucleotide reactions also limited the length of the product *ga*NA oligonucleotide. With a 3% yield for a single coupling of *ga*-dT acetals, it can be expected that the yield of a *ga*-dT trinucleotide would be on the

order of 0.1%. As a result, a *ga*-dT trinucleotide was never identified by HPLC or by

mass spectrometry. As with dinucleotide formation, the yield of *ga*-rU trinucleotides is

expected to be higher (approximately 2%) due to the ten-fold greater aqueous solubility

of rU versus dT. Indeed, mass spectrometry analysis of the crude reaction mixture of a

1:1 ratio of rU and glyoxylate indicates that trinucleotides were produced in a drying-

heating reaction (Figure 4.6).
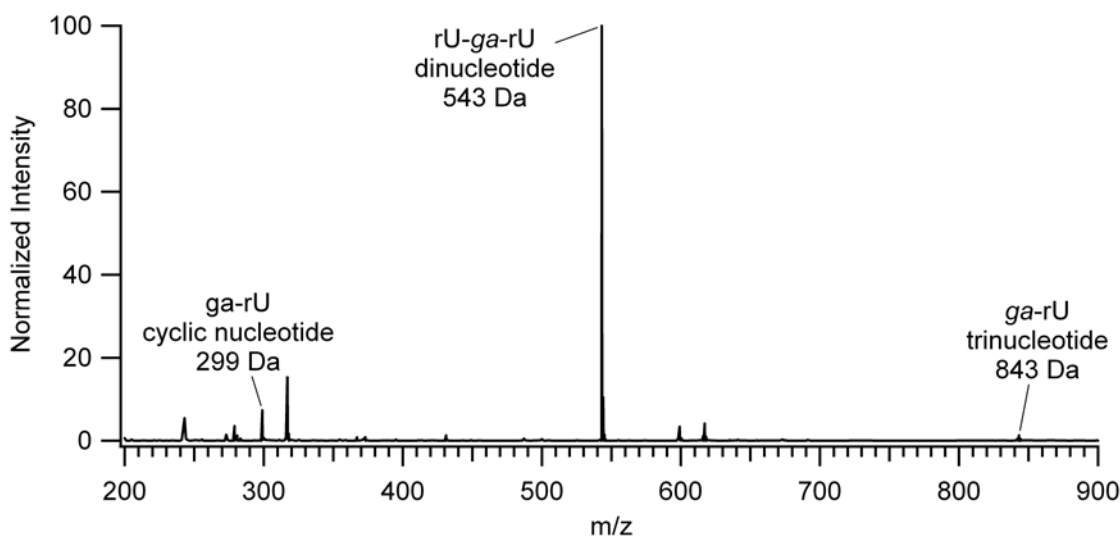
Figure 4.6: ESI- mass spectrum of a gloxylate-rU reaction mixture showing the presence of *ga*-rU cyclic nucleotides (m/z = 299 Da), rU-*ga*-rU dinucleotides (m/z = 543 Da) and *ga*-rU trinucleotides (m/z = 843 Da).

The observed correlation between condensation product yield and reactant

solubility suggests the need for solution conditions where water can be removed, but

where reactants remain in solution, able to diffuse and rearrange. Typically, chemists

perform acetal reactions in anhydrous organic solvents, but these solvents do not adhere

to the model of a prebiotic earth environment. Rather, the ideal solvent would be one in

which nucleosides and aldehydes are freely soluble, that is miscible with water, yet

readily dehydrates. One category of solvents that can meet these requirements is ionic liquids.

Ionic liquids (ILs) are generally defined as molten salts (i.e. liquids that almost exclusively contain ions) with melting temperatures below 100 °C. They come in a myriad of ion pair combinations, each with a unique suite of properties, such as melting temperatures, liquid ranges, thermal and aqueous stabilities, solvating and catalytic properties. Notably, 1-methylimidazolium tetrafluoroborate (HmimBF$_4$) has been reported to specifically catalyze acetal formation. In this work the authors report that this Brønsted acidic ionic liquid specifically catalyses acetal formation in high yields from diols and a variety of carbonyls. Additionally, HmimBF$_4$ is heat and water stable, water miscible, yet dehydrates readily under vacuum and/or heating, making this IL a good candidate for *ga*NA dinucleotide reactions.

As HmimBF$_4$ is a known catalyst for acetal formation, it can be used with the *ga*NA synthesis as a proof of principle to establish the potential role of ionic liquids as catalysts and solvents in boosting the yields of prebiotic condensation reactions. Subsequent investigations could involve catalysis of other condensation reactions, such as peptide bond formation, nucleoside formation, etc. Additionally, there are other plausibly prebiotic compounds such as urea, choline, formate, and glyoxylate, that have been demonstrated to function as an ionic liquid or a deep eutectic solvent (ionic solvents that are mixtures forming a eutectic with melting points that are much lower than either component) under various reaction conditions. The influence of these ionic liquids on condensation and other prebiotic reactions should be explored.

4.4 KINETICS

4.4.1 Kinetics of Acetal Formation

While investigating the thermodynamics of glyoxylate dimethyl acetals (Section 3.2.1), the slow kinetics of acetal formation were noted. If the acidic water-methanol-glyoxylate reaction mixture was kept at room temperature, very little acetal was observed to form, even days after mixing. However, if the mixture was heated to 85 °C for 5 – 10 min then cooled to room temperature, acetal peaks were observed. The slow kinetics of acetal formation were accounted for in the experiments described in Section 3.2.1 by allowing an equilibration period after reaching the target temperature in the NMR probe. The equilibrium amounts of hydrate, hemiacetal, and acetal species were only measured when no further change in their ratios was observed in the NMR spectrum.

As heating-drying reactions are non-equilibrium environments with rapidly changing solution conditions, it is possible that the slow kinetics of glyoxylate acetal formation could be limiting *ga*NA dinucleotide yield. We hypothesized that the kinetics of acetal formation could be dictated by aldehyde substituents effects, much in the same way as the equilibria of acetal formation are governed by substituents factors such as sterics, bond eclipsing, and electron donation or withdrawal. It is possible that the partial positive charge on the carboxylate group of glyoxylate not only destabilizes the carbonyl to promote hydrate and hemiacetal formation, but also destabilizes the carbocation transition state in acetal formation (Figure 4.7), increasing the activation energy and slowing the kinetics of the reaction. If this proposed mechanism is affecting the reaction kinetics, then *a*NA yields may increase if an aldehyde linker was used that possessed a less destabilizing R group, yet still promotes hemiacetal formation. To test this
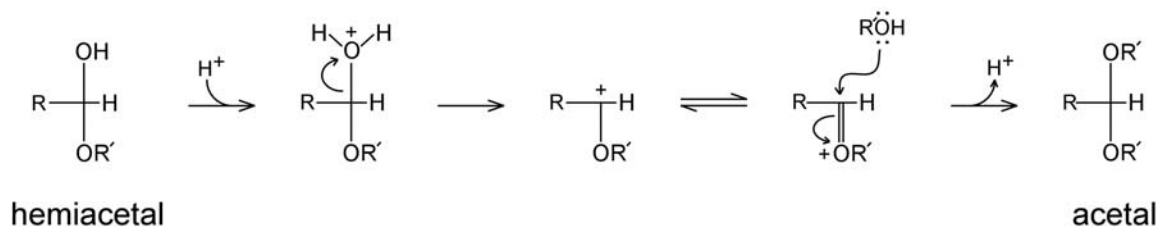
hemiacetal   acetal

Figure 4.7: Mechanism of acetal formation. The nucleophilic substitution reaction of converting the hemiacetal to an acetal goes through a carbocation intermediate.

glyoxalate   glyoxal   glycolaldehyde
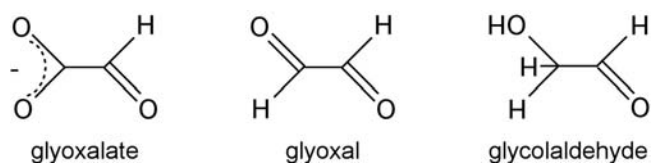
Figure 4.8: The glyoxylate family of aldehydes. Glyoxal and glycolaldehyde are related to glyoxylate through successive reductions of the carboxylic acid. The prebiotic synthesis of glyoxylate from glycolaldehyde has been demonstrated (*4*).

hypothesis we aimed to study glyoxal and glycolaldehyde (Figure 4.8), which possess aldehyde and alcohol substituents, respectively, and represent a proposed intermediate and starting material for prebiotic glyoxylate synthesis (*4*). However, glyoxal has a high self-association constant and forms polymers even in dilute aqueous solution, similar to formaldehyde, complicating data interpretation on acetal formation. Therefore, we focused our kinetic studies on glyoxylate and glycolaldehyde.

The kinetics of glyoxylate and glycolaldehyde acetal formation were measured by NMR as described in Section 4.2.4. Figure 4.9a shows the acetal proton region of two representative NMR spectra for glyoxylate acetal formation. For glyoxylate, the hydrate, hemiacetal, and acetal peaks were integrated over both the acid and the methyl ester species that formed in the acidic methanol-water mixture. The amount of acetal formed was calculated as a percentage of all aldehyde species and plotted as a function of time

for several temperatures (Table 4.1). Figure 4.9b displays the results for the reaction mixture at 80 °C. Even at this elevated temperature, the time constant for acetal formation is 1.1 h, and the solution has not reached equilibrium even at 3 h. The slow kinetics are reflected in the markedly different dT-*ga*-dT yields obtained at 85 °C versus 65 °C versus 22 °C (Figure 3.6), and NMR data suggest that a further increase in temperature could increase *ga*NA yields, although higher temperatures become less prebiotically plausible.

The kinetics of glycolaldehyde dimethyl acetal formation were also measured at 80 °C (data not shown). At this temperature the mixture reaches equilibrium within minutes, in the time it takes to tune, match, lock, and shim the NMR for spectrum collection. However, at 35 °C the reaction requires hours to reach equilibrium, with a time constant of 1.9 h (Figure 4.9c), which is on the same order as the glyoxylate dimethyl acetal reaction at a temperature that is 45 °C higher. These data support the hypothesis that more electron-withdrawing substituents increase the activation energy of the hemiacetal to acetal transition state, possibly by destabilization of the carbocation intermediate.

Destabilization of the carbocation transition state can also contribute to the stability of dT-*ga*-dT dinucleotides, as the nucleophilic substitution reaction to the hemiacetal would require traversal of the same chemical pathway. Electron-withdrawing substituents on aldehydes have also been noted for stabilizing the glycosidic bond, as the unsubstituted 2′-deoxyribonucleosides are more labile than ribonucleosides, which are less stable than 2′-fluororibonucleosides. Taken together, these lines of evidence suggest that although glyoxylate is a good prebiotic replacement for phosphate as its acetals are thermodynamically favored and quite stable, other aldehydes, such as glycolaldehyde,
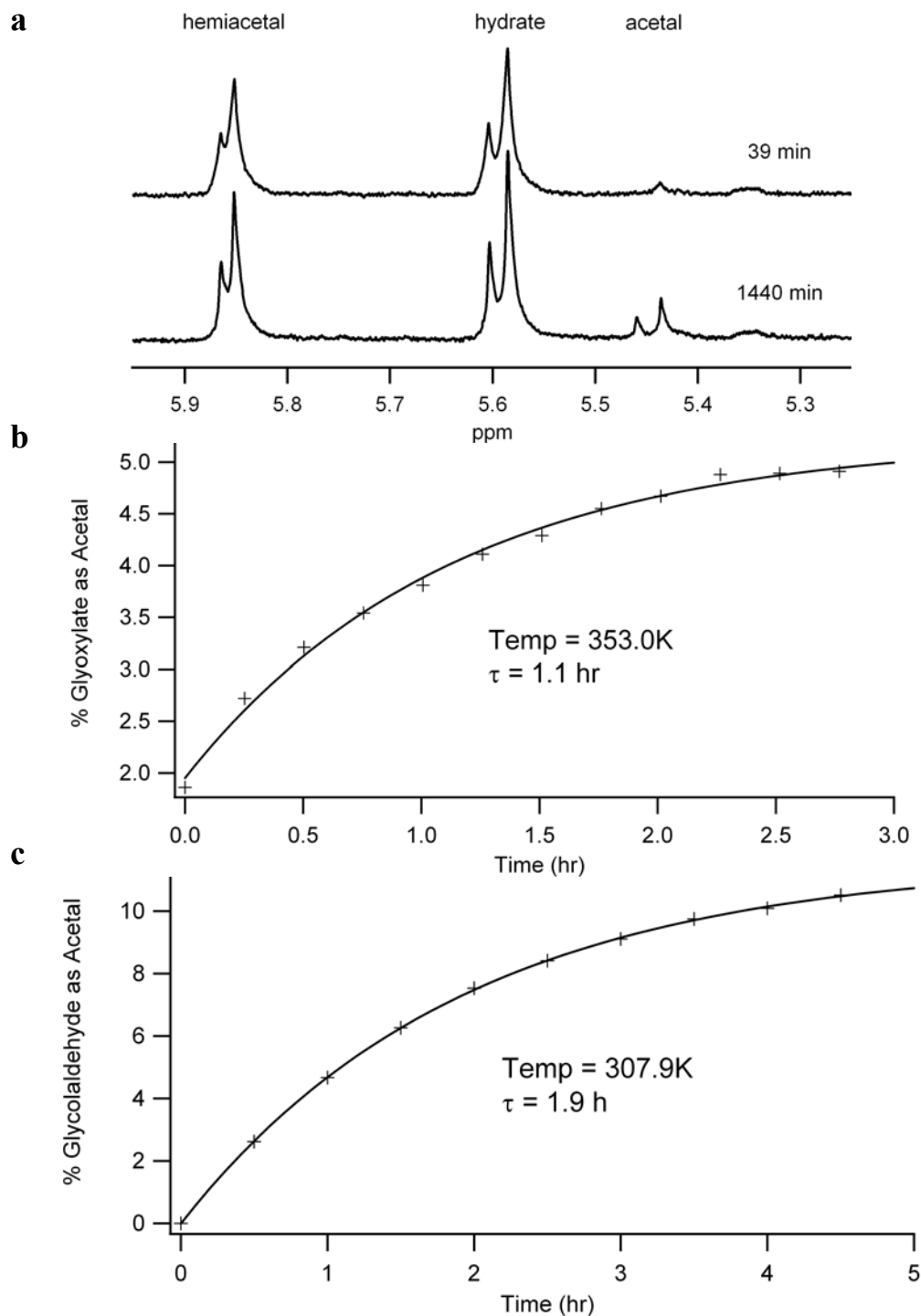
Figure 4.9: Kinetics of dimethyl acetal formation. a) NMR was used to monitor the kinetics of dimethyl acetal formation. These spectra were collected for glyoxylate at 80°C. See Section 4.2.4 for method details. b) Glyoxylate dimethyl acetal formation as a function of time. c) Glycolaldehyde dimethyl acetal formation as a function of time. Both reactions were 100 mM aldehyde, 25% (v/v) MeOD in D_2O, pD = 1.6.

could form *a*NAs in higher yield in the same heating-drying reactions. Indeed, the synthesis of glycolaldehyde-linked dT dinucleotides was confirmed by LC-UV (data not shown) and LC-MS (m/z = 525 Da, ESI-), and acetal yield was sufficient to synthesize trinucleotides (m/z = 809 Da).

The kinetics of glyoxylate dimethyl acetal formation catalyzed by $MgCl_2$ were also determined. A sample that consisted of 100 mM glyoxylate (neutralized), 25% (v/v) MeOD, 75% D_2O, and 100 mM $MgCl_2$ was flame sealed in an NMR tube and incubated at 85 °C for two months, with periodic removal to room temperature for NMR analysis. The data are shown in Figure 4.10. The time constant for the glyoxylic acid acetal reaction catalyzed by $MgCl_2$ was calculated to be greater than 12 d, or approximately 300 h, meaning that the reaction is two orders of magnitude slower than the acid catalyzed reaction. In the absence of $MgCl_2$, no acetal formation was observed (data not shown). Again, the slow kinetics of the magnesium catalyzed reaction could explain the lower dT-*ga*-dT yields than for the HCl catalyzed reaction. An interesting side note is that although $MgCl_2$ can promote the acid-catalyzed formation of acetals, it is not able to catalyze the formation of methyl esters, as determined by NMR (Figure 4.10a). Therefore, although magnesium promotes slower acetal formation than HCl, it may be a more specific catalyst with the potential to reduce non-specific synthesis of side products.

The drying-heating reactions for *ga*NA synthesis maintain the reactants in the solution phase for only a few hours, which is a serious limitation when the reaction kinetics are on the same order as the drying time (e.g., acid-catalyzed glyoxylate acetal reaction at 85 °C), and especially when the kinetics are orders of magnitude slower (e.g., magnesium-catalyzed glyoxylate acetal reaction at 85 °C). We envision two possible
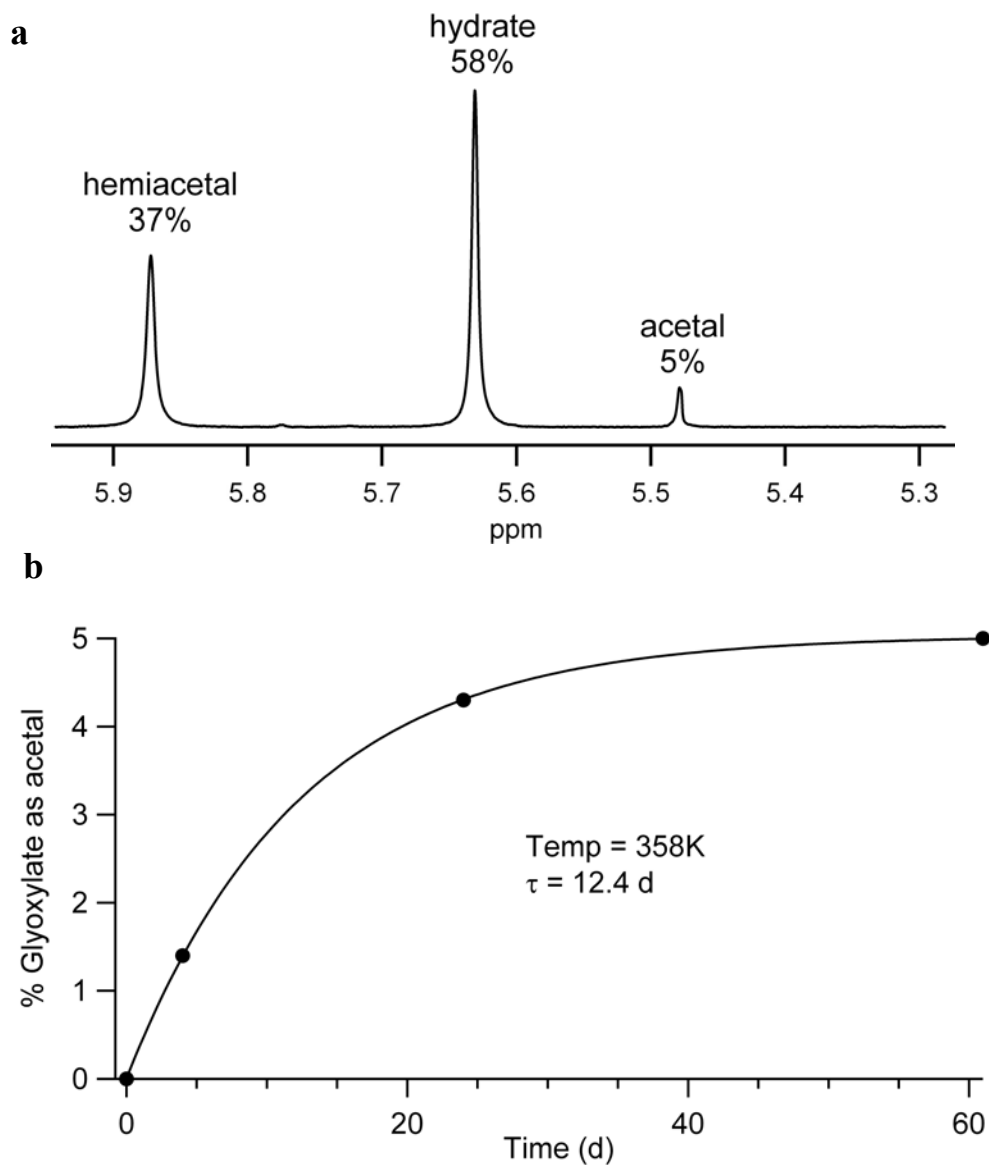
Figure 4.10: Kinetics of glyoxylate dimethyl acetal formation catalyzed by magnesium. a) The aldehydic proton region of the NMR spectrum for the magnesium-catalyzed sample at 61 d. b) Glyoxylate dimethyl acetal formation was monitored by NMR as a function of time. The data were fit with an exponential to obtain the time constant. The sample, 100 mM glyoxylate (neutralized), 25% (v/v) MeOD in D_2O, 100 mM $MgCl_2$, was incubated at 85°C. The NMR spectra were taken at room temperature.

scenarios that could circumvent these problems. First, one can imagine that in a mixture of aldehydes, the molecules with the lower energy barrier to acetal formation will form nucleotides in higher concentrations, particularly on short time scales, forming backbone linkages that are more kinetically accessible. Glycolaldehyde, for instance, would form these backbone linkages more readily at lower temperatures than would glyoxylate. Then, once formed, this backbone could be modified to make more kinetically stable linkages. For instance, glycolaldehyde could be oxidized, perhaps by a metal ion such as Fe^{3+}, to glyoxylate (Figure 4.11). This model also fits with the two step thermodynamic-kinetic backbone synthesis as described for the ANP backbone in Chapter 1. The second scenario would again be facilitated by ionic liquids, which are perhaps the ideal solvent for prebiotic condensation reactions. Since water removal from an ionic liquid will not require that the reaction solution goes to dryness, slow kinetics will not be as much of a hindrance, allowing the reactants to continuously diffuse and rearrange in the absence of water, even over the course of days.
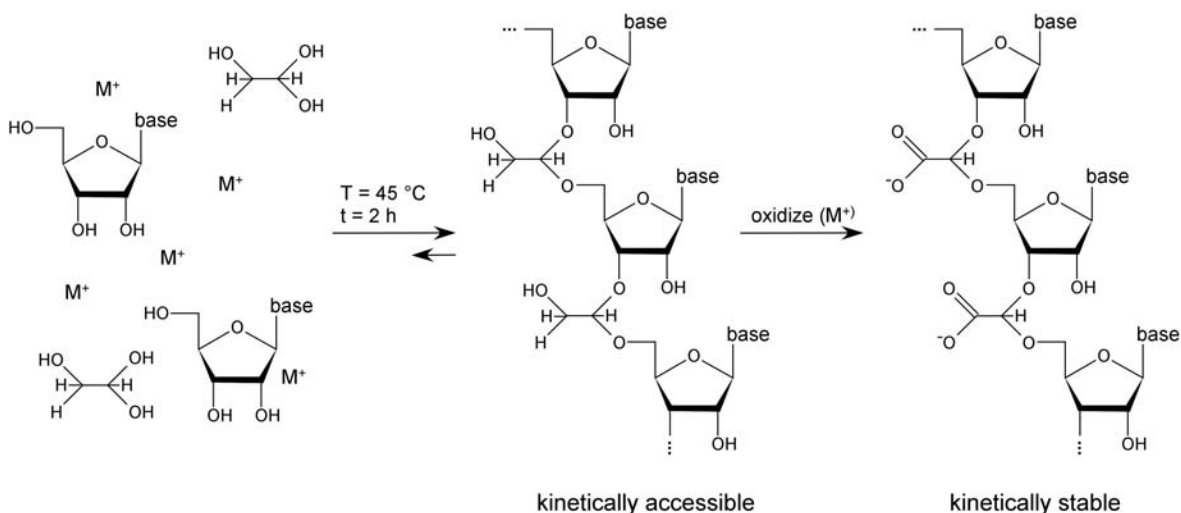
FIGURE 4.11: Two-step mechanism for *a*NA synthesis. First, the more kinetically-accessible glycolaldehyde linkage is formed, followed by an oxidative step to generate a more kinetically-stable backbone.

4.4.2 Kinetics in Nucleoside Formation

In Chapter 2 we proposed that the inability of uracil and cytosine to form nucleosides is of kinetic origin, attributed to the lack of an in-plane lone pair of electrons to nucleophilically attack the C1′ of the sugar to form a glycoside bond. This limitation would not be a universal characteristic of all pyrimidine bases, and therefore it was hypothesized that other bases would be able to form nucleosides, given the appropriate substituents on the pyrimidine ring to generate a nucleophilic nitrogen. Support for this hypothesis was provided with the successful synthesis of 2-pyrimidinone nucleosides, although no direct kinetic data was obtained.

An NMR experiment devised to screen for the nucleoside-forming ability of modified nucleic bases was found to provide insights into the role of kinetics of nucleoside formation. In these experiments, acetaldehyde, a model for an open chain aldose, is combined with a base in DMSO-d6. The mixture is analyzed by NMR for the addition of the acetaldehyde to the base to form the hemiaminal, observed through the formation of new carbon and proton resonances. The experiment is carried out in DMSO-d6 to obtain base concentrations in the $0.1 - 1$ M range, and to reduce competition with water to form the acetaldehyde hydrate. The bases tested were uracil, 6-azauracil, 5-azauracil, 2-pyrimidinone, 4-pyrimidinone, cyanuric acid, maleic hydrazide, 5-hydroxy-6-azauracil, 5-bromo-6-azauracil, and purine. Uracil and 6-azauracil were tested as negative controls as the glycolysis of these bases have been attempted many times in many different labs with no observed products. The 2-pyrimidinone was used as a positive control as glycosylation had already been demonstrated, as reported in Chapter 2.
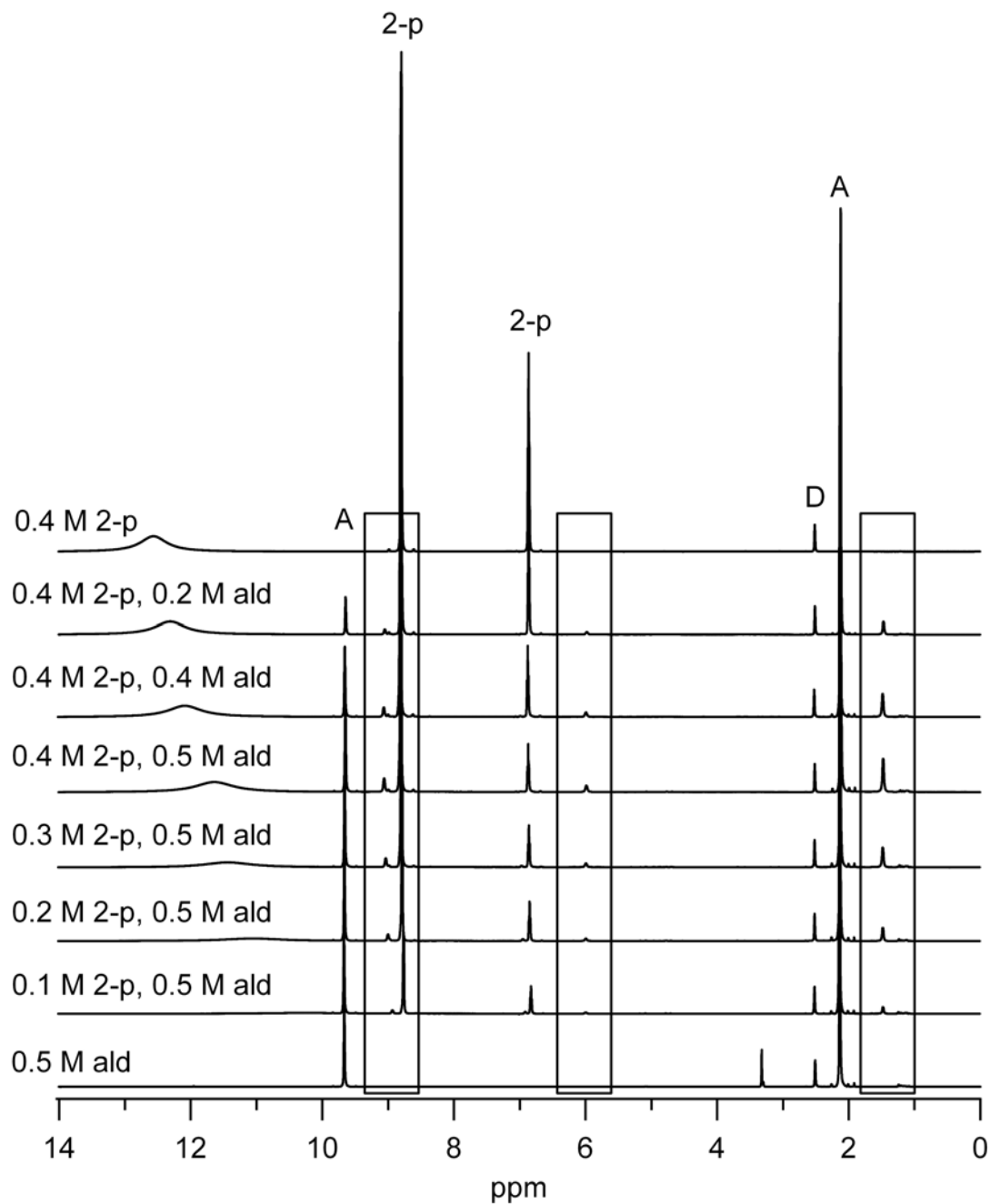
Figure 4.12: NMR of varying ratios of 2-pyrimidinone and acetaldehyde in DMSO-d6. In mixtures of 2-pyrimidinone (2-p) and acetaldehyde (ald), new peaks can be observed (boxed) which are consistent with hemiaminal formation. Peaks labeled with an "A" are acetaldehyde protons, "2-p" are 2-pyrimidinone protons, and "D" are DMSO protons. The peak at 3.3 ppm in the 0.5 M acetaldehyde spectrum is HOD.

The results of the screen were mixed. The positive control for 2-pyrimidinone worked and the intensity of the new peaks that were observed were dependent upon both the concentration of the base and of the acetaldehyde, as determined through NMR titration experiments (Figure 4.12). New proton peaks were also observed for all of the rest of the bases of interest, suggesting glycosylation, although the data that were obtained did not discriminate between hemiaminal bond formation between the acetaldehyde and a base nitrogen, versus the glycosylation of the base through an exocyclic group. Unexpectedly, the formation of new peaks was also observed for both uracil (Figure 4.13) and 6-azauracil. However, the rates of these glycosylation reactions differed drastically from base to base, with uracil, 6-azauracil, 6-hydroxy-6-azauracil, and
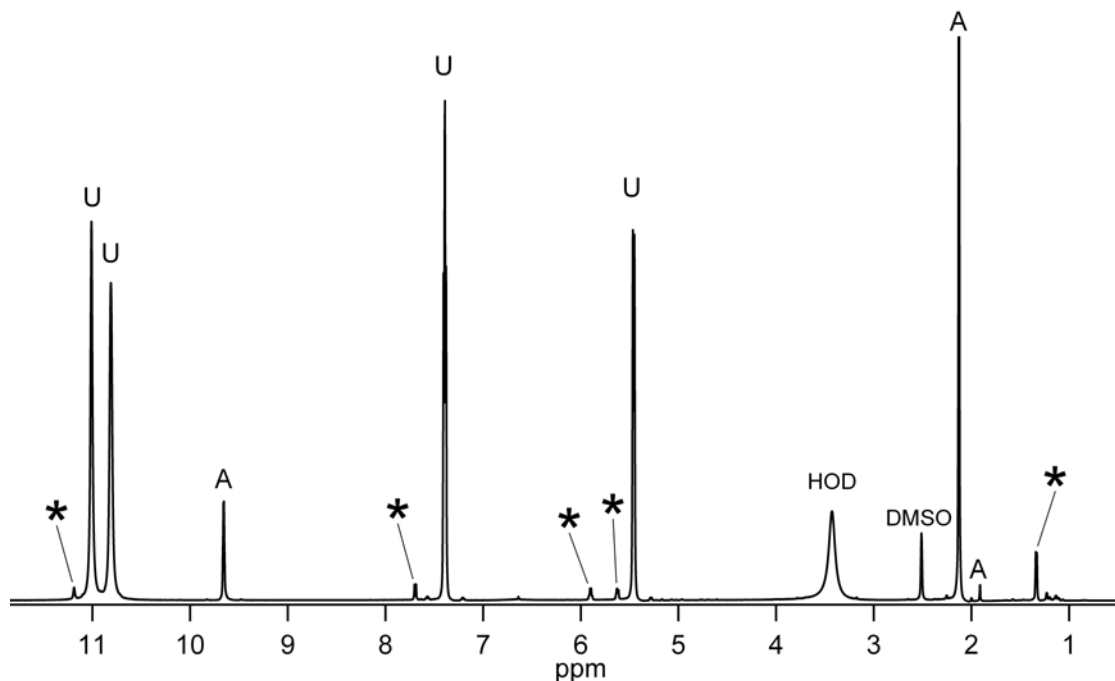
Figure 4.13: NMR spectrum of a mixture of acetaldehyde and uracil in DMSO-d6 after eight days at room temperature. The peaks marked with a "U" are assigned to unreacted uracil, "A" to unreacted acetaldehyde, HOD and DMSO to solvent protons. New peaks that were observed to grow as a function of time are marked with an asterisk. Uracil and acetaldehyde were each 100 mM.

5-bromo-6-azauracil addition to acetaldehyde being the slowest reactions (i.e., the NMR spectra of these samples were still changing after eight days at room temperature). Although further studies need to be conducted to confirm that the new NMR peaks in the uracil sample are glycosides of N1, these data suggest that uridine synthesis is possible, but the reaction rate is very slow.

Kinetics not only plays a role in the ability for a base to be glycosylated, but also in the distribution of nucleoside anomers in the product mixture. In the synthesis of urazole ribonucleosides, Dworkin noted that on short time scales, a mixture of nucleoside anomers is produced, with distributions at time zero extrapolated to be 20% each of β-pyranose, α-pyranose, and α-furanose, with β-furanose comprising the other 40% (*31*). However, once the system is allowed to come to equilibrium, the β-pyranosyl anomer represents 78% of the product, the β-furanosyl 9%, α-pyranosyl 8%, and α-furanosyl the least at 5% (*31*). For all four of the aldopentose and all eight of the aldohexose urazole nucleosides that were synthesized, the pyranosyl anomers were dominant at equilibrium, which mirrors the distributions of the free sugars in solution. The abundance of the furanosyl anomers of the 2-pyrimidinone nucleoside synthesis suggests that kinetics rather than thermodynamics may dictate the distribution of anomers in the product mixture (Table 4.2). However, the influence of the magnesium on the product distribution can not be discounted (Section 4.5).

4.5 The Role of Salts in the Selective Synthesis of Prebiotic Compounds

Prebiotic molecular selection was certainly not restricted to multi-molecular complexes, such as the midwife stacks, but likely active at the molecular and metal ion levels. Sugars, with their tandem hydroxyl groups, are well suited for chelating cations

Table 4.2: Comparison of 2-pyrimidinone nucleoside anomer distribution to equilibrium distributions for ribose

	Distribution of ribose anomers (%)	
	Furanose	Pyranose
Nucleoside, MgCl$_2$	92	18
Equilibrium, D$_2$O, 31°C	20	80
Equilibrium, D$_2$O, 80°C	30	70

However, due to their variance in ionic radii and coordination geometries, different metal ions are best chelated by different sugars (*107*). Additionally, a given salt will not be chelated equally by all anomers of a given sugar (e.g., pyranosyl versus furanosyl ribose) (*107*). Because the anomers of a sugar are in dynamic equilibrium, the relative population of anomers can be shifted in the presence of a cation, generating selection. For instance, the methylation of glucose results in 92% methyl pyranosides and 8% methyl furanosides under catalysis by HCl, which is close to the equilibrium distribution in water of 100% pyranosyl anomers (*33, 108*). However, methylation in the presence of SrCl$_2$ shifts the product distribution to 35% methyl furanosides as the strontium ion is preferentially chelated by the furanosyl anomer of glucose (*108*). This selective power may have also influenced the 2-pyrimidinone nucleoside formation reactions reported in Chapter 2. The ribosylation of 2-pyrimidinone under catalysis with HCl results in a roughly equal distribution of α- and β-furanosyl nucleosides (Figure 2.4). However, in the presence of magnesium, the β-furanosyl nucleoside is produced in 6:1 ratio to the α-furanosyl anomer (Figure 2.4), perhaps because the furanosyl-magnesium chelate occludes nucleophilic attack by the pyrimidine from the α-face of the sugar.

Magnesium was chosen as the salt for investigating 2-pyrimidinone nucleoside synthesis so that comparisons could be made with the yields of purine nucleosides

reported by Orgel and coworkers (*8*). Although Orgel chose his mixture of magnesium chloride and sulfate salts to mimic the composition of extant seawater (*8*), it was also established that $MgCl_2$ yields the greatest concentration of 2-pyrimidinone nucleosides in comparison to lithium, sodium, calcium, iron (II), strontium, nickel, and zinc chlorides (*81*). As mentioned above, not only would the yield of nucleosides be affected by the salts used during synthesis, but so should the anomer distribution. Since ribose chelates magnesium more weakly than it chelates calcium, a greater shift in the anomer distribution of 2-pyrimidinone nucleosides may be observed if calcium were present in the reaction medium. However, due to the low yields reported for many of the other chloride salts tested, the dependence of anomer distribution on the presence and concentration of the other cations could not be established. In order to study the effects of salt chelation on the 2-pyrimidinone nucleoside synthesis, it may be necessary to use the salts in combination with magnesium or HCl in order to boost nucleoside yield. As the prebiotic environment certainly contained a mixture of cations that each influenced reaction mechanisms, a combinatorial study of salt effects could prove informative.

Given the experimental observations regarding the effects of metal ions on glycosylation reactions, it appears possible that metal ion salts also affected other prebiotic synthesis reactions. Decker and co-workers observed differences in the relative distribution of sugar products of the formose reaction dependent upon the calcium concentrations in the reaction mixture, with low or no calcium increasing the production of branched sugar products (*27*). Therefore, whether through the association of intercalators with nucleic acid monomers and polymers, the association of salts with sugars during synthesis or nucleoside formation, or other small molecule associations,

non-covalent interactions conceivably played a significant role in the selective synthesis of nucleic acid monomers and polymers.

CHAPTER 5

CONCLUDING REMARKS

5.1 BUILDING A PROTO-RNA

The RNA world hypothesis was proposed decades ago as a solution to a fundamental quandary of the origin of life (*10-12*). This hypothesis states that RNA molecules, which are capable of storing genetic information and performing catalysis, could have preceded both DNA and proteins, which perform these functions in extant biology. However, biological synthesis of RNA is a highly evolved process that requires protein enzymes, leading many to conclude that RNA evolved from simpler, more easily assembled, and perhaps more heterogeneous informational polymers, termed proto-RNAs (*55*). The prebiotic synthesis of a proto-RNA is not trivial, first requiring the selection of appropriate molecules (i.e., base pairs, sugars, and linking moieties) from the myriad of chemically similar, but less functional molecules in the prebiotic chemical inventory. Once selected, the molecules must be coupled in the correct way to form nucleosides, nucleotides, and oligonucleotides that are homogeneous enough to form duplexes, a requirement for heritability.

The use of a combination of molecular midwives and reversible backbones were proposed by Hud and Anet as a means of overcoming many of the challenges of selection and coupling in the synthesis of the earliest proto-RNAs (*2*). The midwives would have acted as nanometer scale templates upon which base pairs of a uniform width would have been assembled into a columnar stack with a spacing of 6.8 Å (i.e. the thickness of two stacked ring systems). Once the base pairs were assembled, linking molecules that were long enough to span the intervening midwife molecule could stitch together sequential

bases into a proto-RNA polymer. The midwife molecules, which are not covalently attached to the proto-RNA assembly, could be dissociated from the newly-formed helix by changes in solution conditions. Through assembly with midwives, the synthesis of proto-RNAs with uniform helix width and backbone length is achieved. Additionally, experimental evidence proves that midwives can assemble very short sequences at concentrations that are orders of magnitude lower than ever demonstrated for non-intercalated non-enzymatic template-directed synthesis (*64, 5*).

Reversible backbone linkage chemistry plays a significant role in the midwife model in that it allows the selection of the thermodynamically most favored product (e.g., correct base pairs, backbones of the optimal conformation and length). This feature is essential for high-fidelity sequence information transfer in non-enzymatic systems, which lack the proof-reading functions of polymerase enzymes. Reversible backbones promote the synthesis of full-length proto-RNAs with the correct complementary sequence and significantly suppress the production of side products, as demonstrated by Lynn and coworkers for amine-nucleoside polymers (ANPs) (*67-69*).

In this thesis, acetals were explored as a reversible backbone linker for proto-RNAs. Chapter 3 focused on glyoxylate as the aldehyde for forming acetal-linked nucleic acids (*a*NAs) as it is the smallest aldehyde molecule that is negatively charged at neutral pH, a chemical characteristic that is integral to the function and stability of nucleic acid polymers (*50, 51*). Acetal formation is generally enthalpically favored over the hemiacetal and hydrate, which was confirmed for glyoxylate in this work, meaning that chemical activation of the nucleoside is not required for backbone formation. The prebiotic synthesis of glyoxylate-acetal nucleic acids (*ga*NAs) through simple heating-

drying reaction from neutral aqueous solutions was demonstrated, and these linkages were shown to be hydrolytically stable under a considerable range of temperature, pH, and salt concentrations. The kinetic stability of *ga*NAs may also contribute to the relatively low yields during synthesis, which was addressed in Chapter 4. The glyoxylate linkage was calculated to be a good electronic and isosteric replacement for phosphate and computational models of *ga*NA duplex have structural properties that closely match a phosphate-linked RNA helix, suggesting the possibility for cross-pairing between *ga*NAs and RNAs, allowing for sequence transfer and genetic continuity through the evolution from proto-RNAs to RNA.

Another, more fundamental problem for proto-RNA synthesis, the lack of a prebiotic synthesis for pyrimidine nucleosides, has persisted over the past four decades. In 1972 Orgel and coworkers demonstrated that the purine nucleosides (adenosine and inosine) could be synthesized from the free bases (adenine and hypoxanthine) and ribose in a magnesium-catalyzed heating-drying reaction from a neutral aqueous solution (*8*). These reaction conditions do not yield the pyrimidine nucleosides uridine or cytidine, and despite many attempts by several labs in the intervening years, a plausibly prebiotic pyrimidine nucleoside synthesis remained elusive. The lack of a synthesis for pyrimidine nucleosides have lead many researchers to conclude that pyrimidines may have been absent in the earliest proto-RNA molecules (*11, 12, 42-44*). The discovery presented in Chapter 2 that a different pyrimidine base, 2-pyrimidinone, is capable of forming nucleosides by Orgel's reaction mechanism demonstrates that pyrimidine nucleosides could have been available in the prebiotic chemical inventory, but that uridine and cytidine were likely not abundant.

A general model for nucleoside synthesis was proposed that suggests that an in-plane lone pair of electrons on an unhindered nitrogen of the base is necessary for the nucleophilic attack at the C1′ of the sugar for glycosidic bond formation. As uracil and cytosine do not possess this chemical functionality, their glycosylation will be kinetically hindered. However, with the appropriate arrangement of function groups on the base, 2-pyrimidinone and other pyrimidine bases do have a nitrogen nucleophilic enough to obtain high yields of nucleoside on short time scales. NMR data presented in Chapter 4 supports the hypothesis that uracil can be glycosylated, but that the kinetics of the reaction are slow. Therefore, the earliest proto-RNAs may not have contained uracil or cytosine, but rather other pyrimidines that were more readily synthesized in the prebiotic environment. However, the superior functionality and stability of the extant pyrimidines would have provided the selective pressure for their incorporation in RNA.

A great deal of information that is applicable to other prebiotic condensation reactions has been gained through the exploration of pyrimidine nucleoside and *ga*NA nucleotide synthesis. A positive correlation between high yields for aqueous heating-drying reactions and high reactant solubility has been established (Chapter 4). A positive correlation also exists between yields and fast reaction kinetics. Furthermore, the selective influence of salts on prebiotic reactions is an area of rich potential that should be explored. These data in Chapter 4 are pertinent to the prebiotic synthesis of many other biomolecules, including peptides and polysaccharides.

5.2 EVOLUTION AND THE EMERGENCE OF THE RNA WORLD

Selective chemistries, such as supramolecular assemblies formed by bases and molecular midwives, and backbone formation through reversible coupling chemistries

could have provided a route from the monomer world of the prebiotic chemical inventory to a proto-RNA world that was comprised of easily-assembled, thermodynamically favored polymers (Figure 5.1). Although molecular midwives would have promoted the synthesis of proto-RNA polymers with a particular backbone length and helix width, the chemical composition of these polymers may not have been homogeneous. We envision that at the dawn of the proto-RNA world, before the rise of catalytic RNAs, the oligonucleotide backbones would have been composed of a mixture of sugars and low-energy linking molecules and perhaps bases that resembled those of current RNA (i.e., able to Watson-Crick base pair), but not of identical chemical structure. However, as
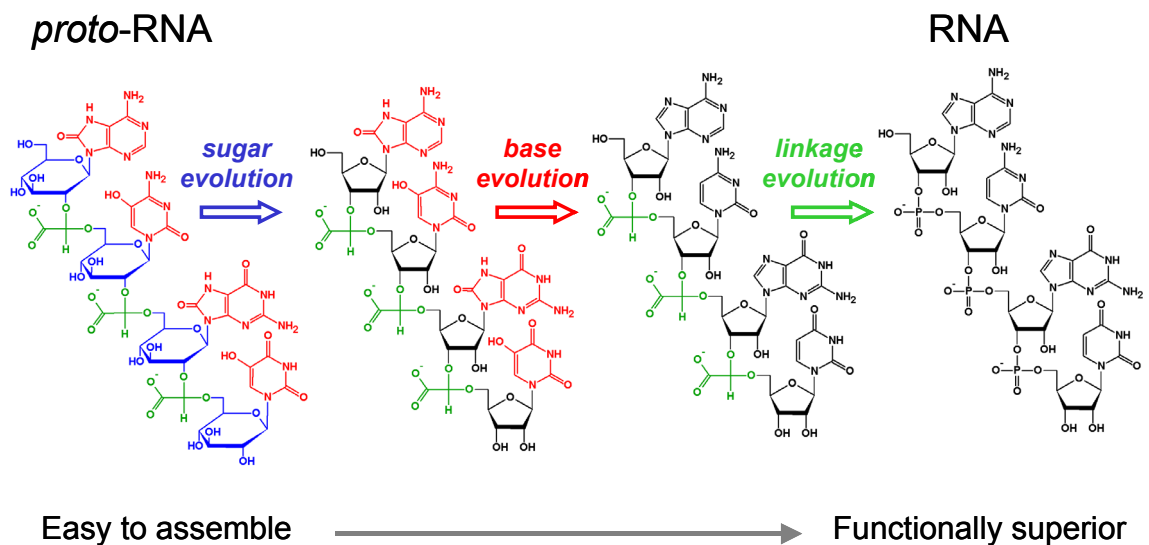
Figure 5.1: Proposed evolutionary pathway from the proto-RNA world to the RNA world. The earliest proto-RNA polymers would have been composed of monomer units that were easy to assemble. Although the proto-RNA depicted here has a backbone comprised of a single type of sugar (blue) and a single linking molecule (green), in the absence of RNA or protein enzymes the earliest proto-RNAs were likely more heterogeneous. As proto-RNA sequences that promoted RNA synthesis were spontaneously generated and selected through Darwinian evolutionary processes, the populations of proto-RNAs would have become more homogeneous, stable, and functional, eventually leading to the catalyzed synthesis of RNA.

113

catalytic proto-RNAs were selected through Darwinian evolution, proto-ribozymes that facilitated RNA synthesis, e.g., ribose synthesis, base synthesis and glycosylation, and phosphodiester bond formation, etc., the proto-ribozymes could have gradually promoted the emergence of RNA-like polymers that were more uniform, stable, functional, and more like RNA (Figure 5.1).

REFERENCES

1. Gesteland, R. and Atkins, J. F., eds. *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World.* 1999, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 709 pp.

2. Hud, N. V. and Anet, F. A. L., *Intercalation-Mediated Synthesis and Replication: A New Approach to the Origin of Life.* J. Theor. Biol. **2000,** *205,* 543-562.

3. Ould-Moulaye, C. B., Dussap, C. G. and Gros, J. B., *A Consistent Set of Formation Properties of Nucleic Acid Compounds. Nucleosides, Nucleotides and Nucleotide-Phosphates in Aqueous Solution.* Thermochim. Acta **2002,** *387,* 1-15.

4. Weber, A. L., *The Sugar Model: Catalysis by Amines and Amino Acid Products.* Origins Life Evol. Biosphere **2001,** *31,* 71-86.

5. Hud, N. V., Jain, S. S., Li, X. H. and Lynn, D. G., *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.* Chem. Biodiversity **2007,** *4,* 768-783.

6. Bean, H. D., Anet, F. A. L., Gould, I. R. and Hud, N. V., *Glyoxylate as a Backbone Linkage for a Prebiotic Ancestor of RNA.* Origins Life Evol. Biospheres **2006,** *36,* 39-63.

7. Bean, H. D., Sheng, Y. H., Collins, J. P., Anet, F. A. L., Leszczynski, J. and Hud, N. V., *Formation of a Beta-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a Plausible Prebiotic Reaction.* J. Am. Chem. Soc. **2007,** *129,* 9556-9557.

8. Fuller, W. D., Sanchez, R. A. and Orgel, L. E., *Studies in Prebiotic Synthesis. VII Solid-State Synthesis of Purine Nucleosides.* J. Mol. Evol. **1972,** *1,* 249-257.

9. Lavery, R. and Sklenar, H., *Defining the Structure of Irregular Nucleic Acids - Conventions and Principles.* J. Biomol. Struct. Dyn. **1989,** *6,* 655-667.

10. Woese, C., in *The Genetic Code. The Molecular Basis for Genetic Expression.* 1967, Harper & Row: New York. p. 179-195.

11. Crick, F. H. C., *The Origin of the Genetic Code.* J. Mol. Biol. **1968,** *38,* 367-379.

12. Orgel, L. E., *Evolution of the Genetic Apparatus.* J. Mol. Biol. **1968,** *38,* 381-393.

13. Schwartz, A. W. and Chang, S., *From Big Bang to Primordial Planet. Setting the Stage for the Origin of Life*, in *Life's Origin: The Beginnings of Biological Evolution*, Schopf, J. W., Editor. 2002, University of California Press: Berkeley, CA. p. 46 - 77.

14. Hollis, J. M., Jewell, P. R., Lovas, F. J. and Remijan, A., *Green Bank Telescope Obserations of Interstellar Glycolaldehyde: Low-Temperature Sugar.* Astrophys. J. **2004,** *613,* L45-L48.

15. Pizzarello, S., *The Chemistry of Life's Origin: A Carbonaceous Meteorite Perspective.* Acc. Chem. Res. **2006,** *39,* 231-237.

16. [cited 24 February 2008]; Available from: www.astrochemistry.net.

17. Kasting, J. F., *Earth's Early Atmosphere.* Science **1993,** *259,* 920-926.

18. Rye, R., Kuo, P. H. and Holland, H. D., *Atmospheric Carbon Dioxide Concentrations before 2.2 Billion Years Ago* Nature **1995,** *378,* 603-605.

19. Miller, S. L., *A Production of Amino Acids under Possible Primitive Earth Conditions.* Science **1953,** *117,* 528-529.

20. Oro, J., *Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions.* Nature **1961,** *191,* 1193-1194.

21. Reid, C. and Orgel, L. E., *Synthesis of Sugars in Potentially Prebiotic Conditions.* Nature **1967,** *216,* 455-455.

22. Reid, C. and Orgel, L. E., *Nucleoside Synthesis under Potentially Prebiotic Conditions.* Nature **1967,** *216,* 936-936.

23. Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H. J. and Miller, S. L., *Prebiotic Synthesis from CO Atmospheres: Implications for the Origins of Life.* PNAS **2002,** *99,* 14628-14631.

24. Plankensteiner, K., Reiner, H., Schranz, B. and Rode, B. M., *Prebiotic Formation of Amino Acids in a Neutral Atmosphere by Electric Discharge.* Angew. Chem. **2004,** *43,* 1886-1888.

25. Saladino, R., Crestini, C., Costanzo, G. and DiMauro, E., *Advances in the Prebiotic Synthesis of Nucleic Acids Bases: Implications for the Origin of Life.* Curr. Org. Chem. **2004,** *8,* 1425-1443.

26. Cleaves, H. J., Nelson, K. E. and Miller, S. L., *The Prebiotic Synthesis of Pyrimidines in Frozen Solution.* Naturwissenschaften **2006,** *93,* 228-231.

27. Decker, P., Schweer, H. and Pohlmann, R., *Bioids. X. Identification of Formose Sugars, Presumable Prebiotic Metabolites, Using Capillary Gas Chromatography/Gas Chromatography-Mass Spectrometry of N-Butoxime Trifluoroacetates on Ov-225.* J. Chromatogr. **1982,** *244,* 281-291.

28. Larralde, R., Robertson, M. P. and Miller, S. L., *Rates of Decomposition of Ribose and Other Sugars: Implications for Chemical Evolution.* Proc. Natl. Acad. Sci. U. S. A. **1995,** *92,* 8158-8160.

29. Fuller, W. D., Sanchez, R. A. and Orgel, L. E., *Studies in Prebiotic Synthesis: VI. Synthesis of Purine Nucleosides.* J. Mol. Biol. **1972,** *67,* 25-33.

30. Kolb, V. M., Dworkin, J. P. and Miller, S. L., *Alternative Bases in the RNA World: The Prebiotic Synthesis of Urazole and Its Ribosides.* J. Mol. Evol. **1994,** *38,* 549-557.

31. Dworkin, J. P. Ph. D. thesis, *Alternatives to Uracil in the Pre-RNA World,* University of California, San Diego, CA, 1997.

32. Angyal, S. J., *Conformatinal Analysis in Carbohydrate Chemistry. I. Conformational Free Energies. The Conformations and Alpha:Beta Ratios of Aldopyranoses in Aqueous Solution.* Aust. J. Chem. **1968,** *21,* 2737-2746.

33. Angyal, S. J. and Pickles, V. A., *Equilibria between Pyranoses and Furanoses. II. Aldoses.* Aust. J. Chem. **1972,** *25,* 1695-1710.

34. Angyal, S. J. and Pickles, V. A., *Equilibria between Pyranoses and Furanoses. III. Deoxyaldoses. The Stability of Furanoses.* Aust. J. Chem. **1972,** *25,* 1711-1718.

35. Bolli, M., Micura, R. and Eschenmoser, A., *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).* Chem. Biol. **1997,** *4,* 309-320.

36. Eschenmoser, A., *Chemical Etiology of Nucleic Acid Structure.* Science **1999,** *284,* 2118-2124.

37. Schöning, K. U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A., *Chemical Etiology of Nucleic Acid Structure: The Alpha - Threofuranosyl-(3'→2') Oligonucleotide System.* Science **2000,** *290,* 1347-1351.

38. Eschenmoser, A., *The TNA-Family of Nucleic Acid Systems: Properties and Prospects.* Origins Life Evol. Biosphere **2004,** *34,* 277-306.

39. Egli, M., Pallan, P. S., Pattanayek, R., Wilds, C. J., Lubini, P., Minasov, G., Dobler, M., Leumann, C. J. and Eschenmoser, A., *Crystal Structure of Homo-DNA and Nature's Choice of Pentose over Hexose in the Genetic System.* J. Am. Chem. Soc. **2006,** *128,* 10847-10856.

40. Sanchez, R. A. and Orgel, L. E., *Studies in Prebiotic Synthesis V. Synthesis and Photoanomerization of Pyrimidine Nucleosides.* J. Mol. Biol. **1970,** *47,* 531-543.

41. Ingar, A.-A., Luke, R. W. A., Hayter, B. R. and Sutherland, J. D., *Synthesis of Cytidine Ribonucleotides by Stepwise Assembly of the Heterocycle on a Sugar Phosphate.* ChemBioChem **2003,** *4,* 504-507.

42. Wächtershäuser, G., *An All-Purine Precursor of Nucleic-Acids.* Proc. Natl. Acad. Sci. U. S. A. **1988,** *85,* 1134-1135.

43. Zubay, G. and Mui, T., *Prebiotic Synthesis of Nucleotides.* Origins Life Evol. Bioshperes **2001,** *31,* 87-102.

44. Battersby, T. R., Albalos, M. and Friesenhahn, M. J., *An Unusual Mode of DNA Duplex Association: Watson-Crick Interaction of All-Purine Deoxyribonucleic Acids.* Chem. Biol. **2007,** *14,* 525-531.

45. Ts'o, P. O. P., Melvin, I. S. and Olson, A. C., *Interaction and Association of Bases and Nucleosides in Aqueous Solutions.* J. Am. Chem. Soc. **1963,** *85,* 1289-1296.

46. Yang, Z. Y., Hutter, D., Sheng, P. P., Sismour, A. M. and Benner, S. A., *Artificially Expanded Genetic Information System: A New Base Pair with an Alternative Hydrogen Bonding Pattern.* Nucleic Acids Res. **2006,** *34,* 6095-6101.

47. Krueger, A. T., Lu, H. G., Lee, A. H. F. and Kool, E. T., *Synthesis and Properties of Size-Expanded Dnas: Toward Designed, Functional Genetic Systems.* Acc. Chem. Res. **2007,** *40,* 141-150.

48. Mittapalli, G. K., Osornio, Y. M., Guerrero, M. A., Reddy, K. R., Krishnamurthy, R. and Eschenmoser, A., *Mapping the Landscape of Potentially Primordial Informational Oligomers: Oligodipeptides Tagged with 2,4-Disubstituted 5-Aminopyrimidines as Recognition Elements.* Angew. Chem., Int. Ed. **2007,** *46,* 2478-2484.

49. Mittapalli, G. K., Reddy, K. R., Xiong, H., Munoz, O., Han, B., De Riccardis, F., Krishnamurthy, R. and Eschenmoser, A., *Mapping the Landscape of Potentially Primordial Informational Oligomers: Oligodipeptides and Oligodipeptoids Tagged with Triazines as Recognition Elements.* Angew. Chem., Int. Ed. **2007,** *46,* 2470-2477.

50. Westheimer, F. H., *Why Nature Chose Phosphates.* Science **1987,** *235,* 1173-1178.

51. Benner, S. A., *Understanding Nucleic Acids Using Synthetic Chemistry.* Acc. Chem. Res. **2004,** *37,* 784-797.

52. Keefe, A. D. and Miller, S. L., *Are Polyphosphates or Phosphate Esters Prebiotic Reagents?* J. Mol. Evol. **1995,** *41,* 693-702.

53. Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F. and Di Mauro, E., *Nucleoside Phosphorylation by Phosphate Minerals.* J. Biol. Chem. **2007,** *282,* 16729-16735.

54. Pasek, M. A., *Rethinking Early Earth Phosphorus Geochemistry.* Proc. Natl. Acad. Sci. U. S. A. **2008,** *105,* 853-858.

55. Joyce, G. F. and Orgel, L. E., *Prospects for Understanding the Origin of the RNA World*, in *The RNA World*, Gesteland, R. F., Cech, T. R. and Atkins, J. F., Editors. 1999, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 49-77.

56. Kawamura, K. and Okamoto, F., *Cyclization and Dimerization of Hexanucleotides Containing Guanine and Cytosine with Water-Soluble Carbodiimide.* Viva Origino **2001,** *29,* 162-167.

57. Miyakawa, S., Joshi, P. C., Gaffey, M. J., Gonzalez-Toril, E., Hyland, C., Ross, T., Rybij, K. and Ferris, J. P., *Studies in the Mineral and Salt-Catalyzed Formation of RNA Oligomers.* Origins Life Evol. Biosphere **2006,** *36,* 343-361.

58. Orgel, L. E., *The Origin of Biological Information*, in *Life's Origin: The Beginnings of Biological Evolution*, Schopf, J. W., Editor. 2002, University of California Press: Berkeley, CA. p. 140-157.

59. Bernal, J. D., *The Physical Basis of Life.* P. Phys. Soc. Lond. A **1949,** *62,* 537-558.

60. Ferris, J. P., Aubrey R. Hill, J., Liu, R. and Orgel, L. E., *Synthesis of Long Prebiotic Oligomers on Mineral Surfaces.* Nature **1996,** *381,* 59-61.

61. Goodwin, J. T. and Lynn, D. G., *Template-Directed Synthesis: Use of a Reversible Reaction.* J. Am. Chem. Soc. **1992,** *114,* 9197-9198.

62. Ihmels, H. and Otto, D., *Intercalation of Organic Dye Molecules into Double-Stranded DNA - General Principles and Recent Developments*, in *Supermolecular Dye Chemistry*. 2005. p. 161-204.

63. Chaires, J. B., *Energetics of Drug-DNA Interactions.* Biopolymers **1997,** *44,* 201-215.

64. Jain, S. S., Anet, F. A. L., Stahle, C. J. and Hud, N. V., *Enzymatic Behavior by Intercalating Molecules in a Template-Directed Ligation Reaction.* Angew. Chem., Int. Ed. Engl. **2004,** *43,* 2004-2008.

65. Horowitz, E. D. and Hud, N. V., *Ethidium and Proflavine Binding to a 2',5'-Linked RNA Duplex.* J. Am. Chem. Soc. **2006,** *128,* 15380-15381.

66. Zhan, Z.-Y. J. and Lynn, D. G., *Chemical Amplification through Template-Directed Synthesis.* J. Am. Chem. Soc. **1997,** *119,* 12420-12421.

67. Gat, Y. and Lynn, D. G., *Reading DNA Differently.* Biopolymers **1998,** *48,* 19-28.

68. Luo, P. Z., Leitzel, J. C., Zhan, Z. Y. J. and Lynn, D. G., *Analysis of the Structure and Stability of a Backbone-Modified Oligonucleotide: Implications for Avoiding Product Inhibition in Catalytic Template-Directed Synthesis.* J. Am. Chem. Soc. **1998,** *120,* 3019-3031.

69. Ye, J., Gat, Y. and Lynn, D. G., *Catalyst for DNA Ligation: Towards a Two-Stage Replication Cycle.* Angew. Chem. Int. Ed. **2000,** *39,* 3641-3643.

70. Li, X., Zhan, Z.-Y. J., Knipe, R. and Lynn, D. G., *DNA-Catalyzed Polymerization.* J. Am. Chem. Soc. **2002,** *124,* 746-747.

71. Benner, S., Burgstaller, P., Battersby, T. and Jurczyk, S., *Did the RNA World Exploit an Expanded Genetic Alphabet?*, in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, Gesteland, R. F. and Atkins, J. F., Editors. 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. p. 163-181.

72. Kelley, J. A., Driscoll, J. S., McCormack, J. J., Roth, J. S. and Marquez, V. E., *Furanose-Pyranose Isomerization of Reduced Pyrimidine and Cyclic Urea Ribosides.* J. Med. Chem. **1986,** *29,* 2351-2358.

73. Kett, W. C., Batley, M. and Redmond, J. W., *Heterocyclic Derivatives of Sugars: An NMR Study of the Formation of 1-Glycosyl-3,5-Dimethyl-1 H-Pyrazoles from Hydrazones.* Carbohydr. Res. **1997,** *299,* 129-141.

74. Saewan, N., Crowe, M. A., Helliwell, M., Raftery, J., Chantrapromma, K. and Sutherland, J. D., *Exploratory Studies to Investigate a Linked Prebiotic Origin of RNA and Coded Peptides. 4th Communication: Further Obserations Concerning Pyrimidine Nucleoside Synthesis by Stepwise Nucleobase Assembly.* Chem. Biodiversity **2005,** *2,* 2005.

75. Brown, D. J., *2-Hydroxypyrimidine.* Nature **1950,** *165,* 1010-1010.

76. Berti, P. J. and McCann, J. A. B., *Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer.* Chem. Rev. **2006,** *106,* 506-555.

77. Horenstein, N. A., *Mechanisms for Nucleophilic Aliphatic Substitution at Glycosides*, in *Advances in Physical Organic Chemistry, Vol 41*. 2006. p. 275-314.

78. Dworkin, J. P. and Miller, S. L., *A Kinetic Estimate of the Free Aldehyde Content of Aldoses.* Carbohydr. Res. **2000,** *329,* 359-365.

79. Kimura, E., Kitamura, H., Koike, T. and Shiro, M., *Facile and Selective Electrostatic Stabilization of Uracil N(1)(-) Anion by a Proximate Protonated Amine: A Chemical Implication for Why Uracil N(1) Is Chosen for Glycosylation Site.* J. Am. Chem. Soc. **1997,** *119,* 10909-10919.

80. Iocono, J. A., Gildea, B. and McLaughlin, L. W., *Mild Acid Hydrolysis of 2-Pyrimidinone-Containing DNA Fragments Generates Apurinic/Apyrimidinic Sites.* Tetrahedron Lett. **1990,** *31,* 175-178.

81. Collins, J. P. Ph. D. thesis, *Prebiotic Synthesis of Pyrimidine Nucleosides,* Georgia Institute of Technology, Atlanta, GA, 2006.

82. Frick, L., Yang, C., Marquez, V. E. and Wolfenden, R., *Binding of Pyrimidin-2-One Ribonucleoside by Cytidine Deaminase as the Transition-State Analog 3,4-Dihydrouridine and the Contribution of the 4-Hydroxyl Group to Its Binding-Affinity.* Biochemistry **1989,** *28,* 9423-9430.

83. Gildea, B. and McLaughlin, L. W., *The Synthesis of 2-Pyrimidinone Nucleosides and Their Incorporation into Oligodeoxynucleotides.* Nucl. Acids Res. **1989,** *17,* 2261-2281.

84. Orgel, L. E., *Prebiotic Chemistry and the Origin of the RNA World.* Crit. Rev. Biochem. Mol. Biol. **2004,** *39,* 99-123.

85. Böhler, C., Nielsen, P. E. and Orgel, L. E., *Template Switching between PNA and RNA Oligonucleotides.* Nature **1995,** *376,* 578-581.

86. Piccirilli, J. A., *Origin of Life. RNA Seeks Its Maker.* Nature **1995,** *376,* 548-549.

87. Miller, S. L., *Peptide Nucleic Acids and Prebiotic Chemistry.* Nature Structural Biology **1997,** *4,* 167-169.

88. Schwartz, A. W., *Speculation on the RNA Precursor Problem.* J. Theor. Biol. **1997,** *187,* 523-527.

89. Herdewijn, P., *TNA as a Potential Alternative to Natural Nucleic Acids.* Angew. Chem., Int. Ed. Engl. **2001,** *40,* 2249-2251.

90. Joyce, G. F., *The Antiquity of RNA-Based Evolution.* Nature **2002,** *418,* 214-221.

91. Chaput, J. C., Ichida, J. K. and Szostak, J. W., *DNA Polymerase-Mediated DNA Synthesis on a TNA Template.* J. Am. Chem. Soc. **2003,** *125,* 856-857.

92. Wiberg, K. B., Morgan, K. M. and Maltz, H., *Thermochemistry of Carbonyl Reactions. 6. A Study of Hydration Equilibria.* J. Am. Chem. Soc. **1994,** *116,* 11067-11077.

93. Matteucci, M., *Hybridization Properties of a Deoxyoligonucleotide Containing 4 Formacetal Linkages.* Nucleosides Nucleotides **1991,** *10,* 231-234.

94. Gao, X., Brown, F. K. and Jeffs, P., *Probing Structural Factors Stabilizing Antisense Oligonucleotide Duplexes: NMR Studies of a DNA·DNA Duplex Containing a Formacetal Linkage.* Biochemistry **1992,** *31,* 6228-6236.

95. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* J. Am. Chem. Soc. **1995,** *117,* 5179 - 5197.

96. Case, D. A., Darden, T. A., T.E. Cheatham, I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S. and Kollman, P. A., *Amber 8.* 2004, University of California: San Francisco.

97. Cheatham, T. E., III and Kollman, P. A., *Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes.* J. Am. Chem. Soc. **1997,** *119,* 4805-4825.

98. Wiberg, K. B., Bader, R. F. W. and Lau, C. D. H., *Theoretical Analysis of Hydrocarbon Properties. 2. Additivity of Group Properties and the Origin of Strain-Energy.* J. Am. Chem. Soc. **1987,** *109,* 1001-1012.

99. Bengston, S., *Early Life on Earth*. Nobel Symposium No. 84. 1994, New York: Columbia University. 630 pp.

100. De Proft, F., Langenaeker, W. and Geerlings, P., *Acidity of Alkyl Substituted Alcohols - Are Alkyl-Groups Electron-Donating or Electron-Withdrawing.* Tetrahedron **1995,** *51,* 4021-4032.

101. Avetisov, V. and Goldanskii, V., *Mirror Symmetry Breaking at the Molecular Level.* Proc. Natl. Acad. Sci. USA **1996,** *93,* 11435-11442.

102. Cieplak, P., Cheatham, T. E., III and Kollman, P. A., *Molecular Dynamics Simulations Find That 3' Phosphoramidate Modified DNA Duplexes Undergo a B to a Transition and Normal DNA Duplexes an A to B Transition.* J. Am. Chem. Soc. **1997,** *119,* 6722-6730.

103. Joyce, G. F., Schwartz, A. W., Miller, S. L. and Orgel, L. E., *The Case for an Ancestral Genetic System Involving Simple Analogues of the Nucleotides.* Proc. Natl. Acad. Sci. U. S. A. **1987,** *84,* 4398-4402.

104. Kirby, A. J. and Jencks, W. P., *The Reactivity of Nucleophilic Reagents toward the P-Nitrophenyl Phosphate Dianion.* J. Am. Chem. Soc. **1965,** *87,* 3209-3216.

105. O'Brien, P. J. and Herschlag, D., *Functional Interrelationships in the Alkaline Phosphatase Superfamily: Phosphodiesterase Activity of Escherichia Coli Alkaline Phosphatase.* Biochemistry **2001,** *40,* 5691-5699.

106. Wiberg, K. B., *The Interaction of Carbonyl Groups with Substituents.* Acc. Chem. Res. **1999,** *32,* 922-929.

107. Angyal, S. J., *Haworth Memorial Lecture - Sugar-Cation Complexes - Structure and Applications.* Chem. Soc. Rev. **1980,** *9,* 415-428.

108. Parrish, F. W., Angyal, S. J., Evans, M. E. and Mills, J. A., *Complexes of Carbohydrates with Metal Cation. 6. Formation of Dimethyl Acetals of D-Glucose, D-Xylose, and L-Idose in Presence of Strontium Ions.* Carbohydr. Res. **1975,** *45,* 73-83.