2005

INSIGHTS INTO ENZYMATIC MANIPULATIONS OF NUCLEIC ACIDS

Rashada Corine Alexander
University of Kentucky, rcalex0@uky.edu

ABSTRACT OF DISSERTATION

Rashada Corine Alexander

The Graduate School

University of Kentucky

2004

INSIGHTS INTO ENZYMATIC MANIPULATIONS OF NUCLEIC ACIDS

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By

Rashada Corine Alexander

Lexington, Kentucky

Director: Dr. Stephen Testa, Professor of Chemistry

Lexington, Kentucky

2004

ABSTRACT OF DISSERTATION

INSIGHTS INTO ENZYMATIC MANIPULATIONS OF NUCLEIC ACIDS

This dissertation details three studies dealing with the manipulation of nucleic acids. In the first investigation, each of the four natural nucleobases were analyzed for the ability to serve as a universal template at the ligation junction of a T4 DNA ligase reaction. This resulted in the first instance of sequence-independent ligation catalyzed by any DNA ligase. Although all of the nucleobases display universal templating capabilities, thymidine and guanosine provided the most effective results. In addition, lowered $MgCl_2$ and ATP concentrations, as well as the inclusion of DMSO, also aided in the sequence-independent ligation reported here. In the course of these studies, current methods of removing urea from denaturing-gel purified nucleic acids proved cumbersome. Therefore, in the second study simple butanol extraction was examined as a means to eliminate urea from nucleic acid solutions. Stepwise butanol extraction was the most effective approach to solving this problem and provided a much needed technique for nucleic acid purification. This type of extraction also does not result in significant losses of nucleic acid sample. The third study exploits the molecular recognition and catalytic properties inherent in an autocatalytic group I intron to develop a ribozyme that can replace the 5' end of an RNA substrate with a different RNA. This 5' replacement splicing reaction can potentially repair mutations on the 5' ends of RNA transcripts that lead to a variety of genetic mutations. The model system was a common mutation in a small model mimic of the k-ras gene *in vitro*, which predisposes individuals to lung cancer. This 5' replacement splicing reaction occurred *in vitro* using this small model system; the reaction was also enhanced by the alteration of the molecular interactions involved. The results and implications of each of these studies are detailed in this dissertation.

KEYWORDS: universal template bases, T4 DNA ligase, urea extraction, group I intron, 5’ replacement splicing reaction

Rashada Corine Alexander

February 2, 2005

INSIGHTS INTO ENZYMATIC MANIPULATIONS OF NUCLEIC ACIDS

By

Rashada Corine Alexander

Stephen M. Testa, Ph.D.

Director of Dissertation

Mark S. Meier, Ph.D.

Director of Graduate Studies

February 2, 2005

Date

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

DISSERTATION

Rashada Corine Alexander

The Graduate School

University of Kentucky

2004

INSIGHTS INTO ENZYMATIC MANIPULATIONS OF NUCLEIC ACIDS

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Rashada Corine Alexander

Lexington, Kentucky

Director: Dr. Stephen Testa, Professor of Chemistry

Lexington, Kentucky

2004

This dissertation is dedicated to my family,
especially my grandfather, Steve Clayton
Alexander, Sr.

Acknowledgments

I would like to thank my advisor, Dr. Stephen Testa, for his guidance and support over the last five years. He has taught me a lot about science, and I have appreciated his honesty and help. Dr. Michael Bell, Ashley Johnson, Dana Baum, Joy Sinha, Patrick Dotson, and all of the Testa lab group members (past and present) also deserve my gratitude for their assistance and attitudes, both of which have made my time as a graduate student more enjoyable. In addition, they have all benefited my scientific development. I would also like to thank my doctoral committee members, Dr. Haley, Dr. Bachas, Dr. Ladipo, and Dr. Bailey, and my outside examiner, Dr. Martha Peterson, for their time, help, and consideration throughout this process. I am also grateful to Derrick Holeman, Marc Maynor and Marlon Jones for their social interaction and the good times I had lamenting the life of a graduate student with them. Dr. Bramwell also deserves my thanks for his encouragement and advice, as does my undergraduate advisor Dr. Sherri Lovelace-Cameron. I want to give my special thanks to all of the UK Chemistry Department Staff. Nancy Stafford, Joyce Cambron, Geri Gerke, June Smith, Jamie Robida, and Yuvonne Queen have all been very helpful and kind to me during my time at UK. I also want to thank Dr. Mark Meier for his advice and good nature.

Outside of the Chemistry-Physics Building, I have been blessed with a network of friends and mentors who have contributed greatly to the maintenance of my sanity and growth as a person. I want to thank all of the friends I have made during my academic career, especially (but not limited to) Kim Ford, Sharletta Mahone, Tiffaney Parkman, Megan Gallagher, Clella Hagans, Leslie Robertson Toney, Sarah Topper, the Farkas family, Mrs. Anna Allen-Edwards and the wonderful church family of Quinn Chapel AME Church. I am very fortunate to have met each of them and had them in my life. I also owe a debt of thanks to the Multicultural Office of the Medical Center for "adopting" and supporting me; the office is a great and necessary resource.

I want to thank my mother, Teresa Alexander, for her love, support, and care throughout my life. She is the strongest person I know, and the best mother I could have. My aunt, Karen Alexander, has been a great supporter of me in many ways, and I am very grateful to her. My brothers, Chris and Jacob, have been bright spots in my apartment and my life during my time here. My grandparents, Ruth and Clayton

Alexander, deserve my thanks for all they have done and taught me. I have had them the longest and am grateful for every minute and every lesson. Although my grandfather is gone and I miss him dearly, he is definitely not forgotten. I would also like to thank my father, John Brown, as well as Ruby, and my brothers Daniel and Tim for all of their support and love. To my "Auntie", Corine Brown, I give my thanks for instilling great values in me at an early age and showing me love at every turn. You are very important to me. I also thank my paternal grandparents, Ruby and Henry Brown, and wish I could have had more time with them. I also want to thank my aunts and uncles, Elaine, Kenneth, Steve, Elliott, Belinda and my cousin Jamie because they deserve my complete gratitude for their love and goodness throughout my academic career. I also thank my family's great friend and my surrogate father, A.Z. Perkins, for his support and care. I miss you very much. I also want to thank my namesake, Anna Corine Topper, for her innocence and presence in this world. I also thank the late Leola Goodloe for being a part of my life and a wonderful friend to my family.

Last and most important, I want to thank God for everything because that is precisely what He has done. I could not have done this without Him or the wonderful people and opportunities He has made available to me.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

The two types of nucleic acids, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), are mainly responsible for the storage and transmission of genetic information; however, the catalytic and regulatory capabilities of nucleic acids continue to be explored (1-7). In this dissertation, studies of nucleic acid manipulation via protein and RNA enzymes, or catalysts, are detailed.

Ligases are enzymes that are necessary for certain cellular processes, such as DNA replication and repair; DNA ligases join two neighboring nucleic acids that are base-paired to a nucleic acid template, forming a duplex (8-11). This site of ligation is called a nick. In general, complementary base-pairing in this interaction is assumed to be preferred for ligation; this preference for complementary base-pairing is termed fidelity (12). Since ligases play an important role in living organisms, a higher level of fidelity is imperative in the maintenance of genome integrity (12). Ligases with lowered fidelity, however, could be utilized for site-directed mutagenesis or sequence identification procedures (13-15). DNA ligases have been tested for their ability to ligate nicks in duplexes containing mispairs and were found to be tolerant of particular mispairs in the duplex; yet no ligase exhibiting complete sequence-independent ligation at or adjacent to the nick has been reported (13-14, 16-26). The first study reported here tested the ability of each of the natural DNA nucleobases (adenosine, cytosine, guanosine, and thymidine) to act as universal templates in ligation under specific reaction conditions. This universal templating ability would result in ligation independent of the sequence of the bases at or adjacent to the site of ligation. The work presents information on ligase fidelity, as well as nucleic acid abilities. This insight can be used to study or modify the behavior of either or both of these factors. The process used and the results obtained are explained in this dissertation.

In the process of dealing with nucleic acids, methods for their purification become necessary. Such methods often use high levels of urea, a common denaturant of biomolecules (3). Much of this urea remains in the purified nucleic acid samples. Unfortunately, the presence of urea in these samples can cause problems with the folding of nucleic acids and the resultant structures (27-28). Current ways to remove urea, such as ethanol precipitation and chromatography, can be ineffective and expensive. The second study in this dissertation focuses on the development of a simple and effective technique for the removal of urea in such

situations. Extractions of nucleic acids samples were performed with 2-butanol and evaluated for their ability to efficiently and reproducibly eliminate urea from the samples; the findings are elaborated upon here and present a more than suitable method for urea removal.

As stated above, nucleic acids contain and transmit genetic information; however, they have been and continue to be investigated with regard to catalysis and regulatory abilities in cellular processes (5-7, 29-30). Catalytic RNA was first reported in the 1980's with the discovery of a self-splicing group I intron and the characterization of the catalytic component of Rnase P, which is composed of RNA (3, 6). Group I introns reside in pre-rRNA transcripts and catalyze a self-splicing reaction in which they are removed from the transcript and the flanking exons are ligated together (1, 3, 5). These introns use molecular interactions (the P1, P9.0, and P10 helices) to catalyze this reaction (31-32). When the exons are eliminated from the transcript, the remaining intron can utilize those same interactions to catalyze new reactions on exogenous substrates (28, 33-34). The new catalytic entity is termed a ribozyme, or RNA catalyst. Such ribozymes have been derived from group I introns and reported to catalyze new reactions. These include reactions that result in the excision of sequences from an RNA substrate (trans excision-splicing) or the replacement of the 3' portion of an RNA transcript (trans-splicing or ribozyme-mediated 3' replacement) (29, 33). These reactions provide new methods to alter RNA sequences. These reactions also provide a possible means to correct genetic mutations that are linked to disease prior to the translation of RNA into protein (32). The replacement of the 5' portion of an RNA transcript via a ribozyme-catalyzed reaction has, however, not been reported. The third investigation included in this dissertation centers on the development and analysis of a new ribozyme-catalyzed reaction that can replace the 5' portion of an RNA substrate. Such a reaction would also possess therapeutic possibilities in cases of genetic mutations that are implicated in disease, but may not be corrected by trans-excision splicing or trans-splicing. This study involves the design of such a reaction system and its implementation in a small *in vitro* model system, as well as its further study with respect to the previously mentioned molecular interactions. The investigation imparts additional knowledge of the workings of group I intron-derived ribozymes and the opportunities for disease treatment they may provide.

Chapter 2 - Background

2.1 Nucleic acid composition

2.1.1 DNA

Deoxyribo**n**ucleic **a**cid, a biological polymer, consists of monomeric units termed nucleotides. Each of these units contains a phosphate group, a deoxyribose molecule (a sugar), and a nitrogenous base (1-4). The nucleotides of DNA are divided into two categories based on what type of nitrogenous base is present. Adenine and guanine are purines, while cytosine and thymine are pyrimidines (Figure 2.1). These designations refer to the base component of the nucleotide; however, the entire monomer units are called adenosine, guanosine, cytidine, and thymidine. Each nucleotide also has a single letter abbreviation as shown in Figure 2.1. Please note that *deoxy*ribose refers to the lack of a hydroxyl at the 2' position of the sugar (1-3).

When the nucleotides of DNA are linked, they form a chain, or strand. The nucleotides in the strand are covalently attached through phosphodiester bonds between the phosphate group of one nucleotide and the 3' OH of the adjacent nucleotide, as shown in Figure 2.2A (1-4). Since this linkage occurs from the 5' phosphate of one to the 3' OH of another, its direction is designated as 5' to 3'. DNA in its native form, however, exists as a double strand, or helix. This helix is formed through base-pairing, which are hydrogen-bonding interactions between the bases of the nucleotides. Complementary base-pairing results in A forming two hydrogen bonds with T, while G forms three hydrogen bonds with C (Figure 2.2B). The two strands in the helix are antiparallel, meaning that they run in opposite directions. One is oriented in the 5' to 3' direction, while the opposite strand is in the 3' to 5' direction (1-4). When shown as a two-dimensional structure, as in Figure 2B, the helix resembles a ladder. The sugar-phosphate linkages, or backbone, form the sides of the ladder. The bases, paired complementarily, form the rungs of the ladder. In a three-dimensional depiction, however, the double helix takes on a more coiled form (1, 3).

2.1.2 RNA

In terms of composition, RNA differs from DNA in that 1) the sugar in RNA has a 2' hydroxyl (hence, **r**ibo**n**ucleic **a**cid) and 2) it uses the pyrimidine uridine (U) in place of

thymidine (Figure 2.3A) (1-4). As for how RNA exists in nature, it is usually single-stranded. The strand can fold back on itself, resulting in a great deal of secondary structure, often in the form of helical regions containing base-paired stems and unpaired loops (Figure 2.3B) (3).

2.1.3 Nucleic acid terminology

For the purpose of clarity, common terms used in this dissertation when referring to nucleic acids will be explained here. While nucleotide refers to a phosphate group, a pentose molecule, and a purine or pyrimidine, the terms nucleoside and nucleobase refer to the connected pentose and base only (3, 35). An individual nucleotide in a strand of DNA or RNA can also be referred to as a residue (1-4, 35). Oligonucleotide or oligomer can refer to a single strand of DNA or RNA, which can be synthetic (35). The term DNA, itself, is sometimes used to refer to a single strand, as is RNA. Oligonucleotides may also be denoted as DNA or RNA molecules (2-4). When speaking of a DNA in the form of a double helix, the term duplex is also used (1-4).

2.1.4 Nucleic acids in molecular biology

Genetic information is encoded in specific sequences of DNA nucleotides. This information establishes and preserves the biological functions and processes of a living organism (3-4). The central dogma of molecular biology, shown in Figure 2.4, provides a general diagram for the procession of events leading from nucleic acids to proteins, which are macromolecules responsible for much of the cell's activity. The synthesis of DNA is termed replication and requires DNA-dependent polymerases, which are enzymes that catalyze the polymerization, or joining of nucleotides (1-4). Polymerases also play a part in the next part of the process, which is the transcription of DNA into RNA. Genes, which are particular sequences of DNA that compose an individual genetic element, can code for protein or for particular RNAs (1-3). There are three main types of RNAs, messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA) (1-4). In the progression of nucleic acids to protein, DNA is transcribed into mRNA, and this new RNA is called a transcript. This transcript, the pre-mRNA, undergoes posttranscriptional modification to become a mature RNA, prior to its translation into protein. This modification includes the removal of intervening regions not present in the mature RNA. These regions, introns, can be found in other types of RNA, but in the case of mRNA, they are spliced out and the remaining regions that will be expressed, or translated into protein (exons,

shown in gray) are joined to form a mature RNA transcript (1-4). This processing is done by the spliceosome, a complex composed of protein and RNA, (shown in Figure 2.5). Note that although the spliceosome construct is not shown, it mediates the splicing event (1-3). The exons are shown in black and the intron in gray. The first event in this type of splicing occurs through the attack of a conservd adenosine in the intron on the 5' end of the intron. The freed 3' OH of the terminal base on the 5' exon attacks the 3' exon. This results in exon ligation and the removal of the intronic regions (1-3).

The mature transcript is translated into protein at the ribosome, a ribonucleoprotein (made of protein and rRNA) complex. To understand translation, one must understand how proteins are coded in DNA and RNA. Codons are sequences of three nucleotides that specify a particular amino acid (Figure 2.6). Just as there are four nucleotides in the nucleic acid "alphabet", there are twenty amino acids that compose the protein "alphabet" (1-4). Three nucleotides per codon result in 64 combinations. Since only twenty amino acids are possible, the code is repetitious. Other than the single codon that signal the start of transcription and the 3 codons that signals only the end of transcription, each amino acid has multiple corresponding codons. This is called the redundancy of the genetic code (2). In the genetic code, the corresponding amino acid of a codon is determined mainly from the first two nucleotides. The third position in the codon, also termed the wobble position, can in most cases be any one of the four nucleotides without altering the amino acid identity of the codon (1-3). During translation, each codon in the mRNA pairs with three complementary nucleotides present in a tRNA molecule; this tRNA has the corresponding amino acid attached to it. The ribosome catalyzes the joining of the amino acids to yield the resultant protein product (1-3).

2.2 Nucleic acids and enzymes

Since genetic information is stored in DNA and transcribed into RNA prior to its expression into protein, a great deal of investigation into biological systems requires the manipulation of nucleic acids. This manipulation can be done to study DNA or RNA itself, as well as its interaction with other biomolecules. Nucleic acids are the natural substrates for many enzymes, such as polymerases and ligases (1-4). Many studies analyze or manipulate the behavior of the natural, or canonical, nucleobases in reactions that use these enzymes (1-4).

Some of these investigations include mutagenesis (altering a given genetic sequence), detection, and the study of enzymatic processes (1-4).

2.2.1 Enzymatic ligation

The process of ligation, in terms of nucleic acids, can occur chemically or enzymatically (3, 9-13, 37-42). The oligonucleotides ligated in the reaction can be DNA or RNA (1-3). The enzyme catalyzes the joining of an acceptor oligonucleotide, which contains a required 3' OH group, and a donor oligonucleotide, which contains a required 5' phosphate, as shown in Figure 2.7 (3, 9-12). This ligation junction between the oligonucleotides is referred to as a nick and the oligonucleotides being joined may need to be base-paired with a template molecule (3). As is the case with many enzymatic reactions, a cofactor is required; a cofactor is a nonprotein component that participates in the reaction. Cofactors can be metals or organic molecules and aid in reactions that proteins may not be suited for, such as oxidation-reduction reactions or the transfers of functional groups (3). Ligases, specifically, use ATP (adenosine triphosphate) or NAD (nicotinamide adenine dinucleotide) as cofactors. Which cofactor is needed depends on the source of the enzyme (3, 9, 43). Ligases generated in eukaryotic cells generally require ATP; those obtained from prokaryotic cells use NAD. The mechanism for the enzymatic ligation reaction, using ATP as a cofactor, is depicted in Figure 2.7. The reaction involves 1) the activation of the enzyme with the formation of a enzyme-AMP (adenosine monophosphate) construct with the attendant release of PPi 2) transfer of AMP to the donor oligonucleotide, creating an intermediate, and 3) ligation of the donor oligonucleotide to the acceptor oligonucleotide via a transesterification reaction, resulting in the release of AMP (1-3). Ligases also require divalent cations for their activity, usually Mg^{2+} (1-4, 9-12). DNA ligases can also catalyze the AMP-dependent reverse reaction, which creates nicks in DNA (43).

As stated above, the substrate for ligase can be DNA or RNA. While some of these ligases can ligate single-stranded oligonucleotides, those discussed here will be DNA ligases that require that the oligonucleotides are, at least partially, base-paired with a template molecule. The preference for this base-pairing to be complementary is referred to as the fidelity of DNA ligases. Since ligase is an essential enzyme *in vivo*, its fidelity is essential to maintaining genetic integrity in DNA repair, recombination, and replication (12). A low level of fidelity in such an enzyme could lead to high levels of mutagenesis. Ligases also play a role in a wide variety of *in vitro*

biotechniques that require nucleic acid manipulation, including standard cloning protocols, the Ligase Chain Reaction (LCR) (44), and the Ligase Detection Reaction (LDR) (14, 45-46). High fidelity in these reactions is important for experimental accuracy.

In alternate situations, however, ligases with low fidelity could also be useful for certain experimental strategies, including ligating together degenerate or unknown sequences for purposes such as site-directed mutagenesis or sequence identification (15-16). In this same vein, ligases are known to have the ability, at least *in vitro*, to ligate oligonucleotides that form certain mismatch base pairs with their templates, both at the ligation junction and at other positions in the resultant helices (14-15, 16-26). Ligases that show a complete lack of sequence specificity at and near the ligation junction, however, have not been reported. Chemical ligation, as opposed to enzymatic ligation, has also been studied; however, its fidelity has not been found to be any lower than that of ligases (37-42). Another way to secure such sequence-independent ligation would be to place universal bases (47) in the template molecule at the ligation junction.

Universal bases are nucleoside analogs that bind each of the four canonical bases equally, although not necessarily tightly (48-50). In this way, ligases could use the universal template region to direct the ligation of oligonucleotides regardless of the terminal sequences being joined. Many universal base analogs, however, intercalate with DNA helices and may not actually participate in base-pairing and formation of the helices (49). Nucleoside analogs that actually function through hydrogen-bonding interactions have not been shown to do so equally with all four natural nucleobases (47). Some have even been shown to lower the melting temperatures for the duplexes in which they are incorporated and lower the complementary base-pairing specificity of neighboring nucleotides (50). In addition, all available data indicates that known universal base analogs are not very effective in ligation reactions, especially at the ligation junction itself (48).

In lieu of nucleotide analogs that can behave as universal bases at ligation junctions, one additional avenue to explore in the quest for sequence-independent ligation would involve canonical bases. While no ligases capable of complete sequence infidelity at the ligation junction have been reported, the natural nucleobases could possess universal base capabilities. Specific reaction conditions have been shown to alter the activity of some enzymes. DMSO (dimethyl sulfoxide) has been shown to cause ‘star site’ activity in some restriction enzymes; this results in a relaxation in restriction site sequence requirements (51-52). PEG (polyethylene glycol) is

widely recognized as a molecular crowding agent and is often used to increase the efficiency of enzymatic reactions (53). SSB (single-stranded binding protein) is required in hexamer ligation (54-55). As for DNA ligase, temperature and salt concentration have been shown to affect its fidelity (43). Thus, ample evidence exists that altering reaction conditions can alter enzymatic activity. Therefore, the first investigation systematically analyzed whether any of the four canonical bases could act as universal template bases at, and immediately adjacent to, a centrally located ligation junction under specific experimental conditions. T4 DNA ligase was chosen because of its easy availability and previous reports of its ability to ligate molecules with some mismatched base pairs (14-18, 19-21).

2.3 Urea Extraction of Nucleic Acid Solutions

2.3.1 Necessity of Urea Extraction

The study of nucleic acids requires methods to isolate and purify them once they have been manipulated in the desired fashion. One of the most common analytical and purification methods for nucleic acids is PAGE, or polyacrylamide gel electrophoresis (56). The high resolution (down to a single nucleotide) that this method provides makes it an attractive one for purification (3). PAGE gels, however, can contain high concentrations of urea (up to 8M); urea is a commonly used denaturant and gels containing it are referred to as denaturing gels (56). Many approaches used to study aspects of molecular biology require pure nucleic acids, including methods for radiolabeling, hybridization, primer extension, and most anything dealing with RNA and DNA molecules. After running a denaturing PAGE gel, the desired nucleic acid band is excised from the gel and then extracted from the gel slice, usually by a soaking method that allows the nucleic acid to diffuse out of the gel and into solution (56). Unfortunately, the urea contained within the gel slice also diffuses into solution with the purified nucleic acid. The effects of urea on nucleic acid folding, function, and catalysis can be substantial (27-28). Therefore, removing as much urea as possible from the nucleic acid preparation is necessary and benefits subsequent applications.

2.3.2 Current Methods of Urea Removal

Two of the most common methods for removing urea from nucleic acid solutions can be problematic. Ethanol precipitation uses ethanol and an added amount of salt to cause the

precipitation of nucleic acids in solutions (56). The nucleic acids are pelleted by centrifugation of the ethanol/nucleic acid solution. The pellet is washed with a lesser concentration of ethanol to remove any residual salt and the pellet is dried and reconstituted for further experimental use (56). At such high concentrations of urea as those used in denaturing gels, however, ethanol precipitation is likely to also result in the precipitation of the urea with the nucleic acid. Chromatography, an additional method for urea removal, utilizes separation based on molecule polarity. In the case of urea extraction, the columns necessary can be costly. Additionally, the strong polarity of urea when in solution can lead to the column being overloaded or ineffective

Therefore, a simple and effective new strategy for removing urea from nucleic acid samples would be useful. The 2-butanol extraction of nucleic acid samples was investigated as an additional method to effectively extract urea from such solutions..

2.4 Catalytic RNA

As mentioned previously, the first catalytic RNAs reported were a self-splicing group I intron from *Tetrahymena thermophilus* and the catalytic RNA subunit of Rnase P (6-7). RNase P, a ribonucleoprotein construct, cleaves the 5' end of tRNA transcripts to yield mature tRNAs (Figure 2.8A) (3, 7, 57). Note that RNase P is shown as a general complex. Since those discoveries, other catalytic RNAs have been reported. These include the hammerhead, hairpin, and hepatitis delta virus (HDV) ribozymes, group II introns, and additional group I introns (3, 58-61). The term ribozyme is often used to refer to an RNA entity with catalytic activity (36). The hammerhead (Figure 2.8B), hairpin (Figure 2.8C), and HDV (Figure 2.8D) ribozymes are all found in viruses (parasitic entities consisting of genetic material encapsulated by a protein coat) and they range in length from 40 nucleotides to 160, respectively (58-61). Like Rnase P, they all catalyze cleavage reactions, though they adopt different secondary structures (Figures 2.8B, 2.8C, and 2.8D). The substrate sequences are shown as dotted lines and the cleavage sites are denoted with arrows. The letter N denotes A, C, G, or U; D denotes A, C, or U. Each of these catalytic RNAs require metal ions for catalysis (58-61). In addition to these catalytic RNAs, there are also exist RNA molecules that possess regulatory and catalytic abilities, such as a glycine-dependent system recently reported (7).

The hammerhead, hairpin, and HDV ribozymes provide therapeutic options in gene inhibition. In cases of viruses or defective genes that are linked to disease, elimination of the

complete functional genes by cleavage could provide amelioration or eradication of their effects. Studies in this vein have been done and reported to be successful *in vivo* (30, 36, 62-64).

2.4.1 Intron self-splicing reactions

Group I and group II introns are both capable of self-splicing reactions that remove them from RNA transcripts and ligate the remaining exons together (7, 36, 65-66). General diagrams for each self-splicing reaction are shown in Figure 2.9; both of these reactions require metal ions for catalytic activity (65-66). Although these introns differ in size, they do fall into the range of 160-500 nucleotides in length (3, 59, 65-66). The self-splicing reaction of group I introns (Figure 2.9A) requires a guanosine cofactor, which is shown in the conserved guanosine-binding site. The 3' OH of the guanosine attacks at the 5' splice site, shown at the end of the P1 helix interaction, resulting in the addition of the guanosine to the 5' end of the intron (G-addition reaction). Once the guanosine cofactor has effected the 5' cleavage reaction, the conserved guanosine on the 3' end of the intron (ωG) takes its place in the conserved guanosine-binding site (7, 36, 65). The 3' OH of the 3' terminal base on the 5' exon is now free to attack at the 3' splice site, on the 3' end of the intron. This results in the ligation of the 5' and 3' exons and the excision of the intron. While some group I introns require proteins for this splicing *in vivo*, many can catalyze the reaction *in vitro* without proteins under physiological conditions (65).

The self-splicing reaction for group II introns (Figure 2.9B) involves a conserved adenosine within domain 6 (D6) that attacks the 5' splice site, identified by domain 1 (D1) (3, 36, 65-66). This results in a lariat or loop intermediate. As with the exon ligation step of group I intron self-splicing, the terminal base of the 5' exon attacks at the 3' splice site; the exons are ligated and the lariat intermediate is excised. This reaction bears similarities to the spliceosome-mediated removal of introns mentioned previously. These introns also seem to require proteins for efficient *in vivo* splicing, even though they can splice without proteins *in vitro*. This *in vitro* splicing, however, occurs under extreme conditions (high metal ion concentration or low pH) and at a slower rate than the *in vitro* splicing of group I introns (65-66).

2.4.2 Group I introns in detail

Group I introns are found in genes for messenger, ribosomal, and transfer RNA in various species; they have not yet been reported in vertebrates (3). As previously stated, three molecular

interactions are involved in this self-splicing reaction, the P1, P9.0, and P10 helices, which are shown in Figure 2.10 (7, 33). Each of these helices is associated with a specific recognition element (RE) located in the intron; these recognition elements are sequences complementary to a region in the intron or the flanking exons. The formation of these helices serves to move the molecule into the conformation necessary for efficient splicing, as well as define splice sites (33, 65). The aforementioned P1 helix involves base-pairing between RE1 (also termed the IGS or internal guide sequence) and the 5' exon, while the P10 helix forms between the 3' exon and RE3. The P9.0 helix forms intermolecularly between the RE2 region and the intron; this helix plays a role in defining the 3' intron boundary and thus the 3' splice site (33). There are two conserved and required components present in group I introns, the first being a conserved base-pair formed between a guanosine in the intron (RE1) and a 3' terminal uridine present in the 5' exon. This base-pair denotes the first catalytic site (5' splice site) and is found on the 3' end of the P1 helix. The second component is ωG at the 3' portion of the intron, defining the second catalytic site and marking the end of the intron, which immediately precedes the P10 helix (33, 36, 65, 67).

The self-splicing mechanism by which these introns operate is shown in more detail in Figure 2.10. Note that the exons are depicted as black lines and the intron is shown in gray. The first step involves the cleaving of the 5' exon from the transcript via nucleophilic attack by a required guanosine cofactor (7, 33, 36). The second step of exon ligation occurs through a second nucleophilic attack, this time by the 3' OH of the terminal uridine (of the 5' exon) on the 3' exon. This results in a new phosphodiester bond that links the 5' and 3' exons and the removed intron that has undergone the G-addition reaction (7, 33, 36).

The conserved structure of group I introns contains various helical regions, as previously mentioned. Figure 2.11 shows the secondary structure of the group I intron from *Pneumocystis carinii*, with the intron sequences in uppercase letters and the exon sequences in lowercase letters (32). As the structural elements are generally conserved in this category of catalytic RNA, this diagram serves as a general illustration of the structure of group I introns. The conserved nucleotides and interactions are found in the same general locations (36, 65). Two of these include the guanosine-binding site (in the P7 region) for the guanosine substrate and conserved adenosine residues that interact with the reaction site in the P1 helix (65, 67). Note also that the two catalytic sites are denoted with arrows; the terminal uridine of the 5' exon and the ωG are

circled. As for the elements of the structure and the roles they play in the self-splicing reaction, a great deal of the current knowledge comes from studies of the *Tetrahymena* group I intron. The conserved A residues located above the P4 helix (shown boxed in Figure 11) interact with the conserved G•U base-pair in the P1 helix. The A•A mispair interacts with the conserved G•U base-pair through hydrogen-bonding (67-69). This is implicated in the 5' splice site selection (67). As for 3' splice site selection, ωG appears to play a major role (33, 67, 70-71). Please note that this conservation occurs in terms of location with respect to the helical regions.

2.4.3 Trans-cleavage reaction

When the exons are removed from a transcript containing a group I intron, that remaining intron can, in turn, utilize the same molecular interactions to catalyze new reactions on exogenous substrates. This new construct, now referred to as a ribozyme, can be altered in terms of sequence-specificity to adapt to different targets (36, 72). The *Tetrahymena* intron is one of the most studied group I introns; one of the ribozymes derived from it catalyzes a trans-cleavage reaction (60, 73). The reaction, shown in Figure 2.12, requires a guanosine cofactor, as does the self-splicing reaction, and results in the cleaving of an exogenous substrate. The ribozyme is shown in gray and the substrate in black. The conserved G•U base-pair that defines the 5' cleavage site is present in the P1 helix interaction that targets the substrate (60). Once the P1 helix is formed, the 5' cleavage site is attacked by the guanosine cofactor, leaving the substrate severed (60, 73).

This type of cleavage reaction using a group I intron-derived ribozyme has also been proposed as an option in gene inhibition, similar to the smaller ribozymes mentioned previously (60). This intron-derived ribozyme reaction has not been explored in this arena as readily as the hammerhead and hairpin ribozymes (31, 36, 60). This is partially attributed to the expected frequency of possible cleavage sites and the likelihood that the ribozyme would also destroy non-targeted RNAs (60). Subsequent studies of group I introns, however, have led to additional ribozyme-catalyzed reactions with possible disease treatment capabilities (33-34).

2.4.4 Trans-splicing reaction

The trans-splicing (TS) ribozyme is derived from the *Tetrahymena* group I intron (30, 33). This ribozyme catalyzes a reaction that results in the replacement of the 3' portion of an

RNA transcript and can occur *in vitro* and *in vivo* (30, 64, 74-75). The reaction can also be called ribozyme 3' replacement splicing. A simple diagram of the reaction is shown in Figure 2.13; the molecular interactions previously mentioned play similar roles to the ones they play in the self-splicing reaction. Note that the trans-splicing ribozyme carries a new 3' exon on its 3' end. The ribozyme is shown in gray, with the new 3' exon depicted with an open line. The exogenous exon regions are shown in black. The P1 helix forms between the 5' exon of the exogenous substrate and the RE1 region of the ribozyme; the P9.0 helix forms intramolecularly within the ribozyme, similar to what occurs in the self-splicing reaction (33). The 5' cleavage reaction requires a guanosine cofactor, as does self-splicing. After the first step of 5' cleavage, which leaves the 5' exon of the exogenous substrate free, the P10 helix now forms between the new 3' exon that is attached to the ribozyme and the RE3 region; notice that the 3' exon from the exogenous substrate has undergone the G-addition reaction as the intron does in the self-splicing reaction (33). The second step of exon-ligation takes place much as it does in the self-splicing reaction, by way of a nucleophilic attack of the 3' OH of the terminal uridine, located on the 5' exon, on the new 3' exon that is carried by the ribozyme. This results in a new RNA transcript, containing the original 5' exon and a replacement 3' exon (30, 33). This reaction gave new insight to catalytic RNA and spawned many further studies on its mechanism and requirements. In cases of mutations located on the 3' end of an RNA transcript that are linked to disease or deficiencies that render the 3' portion of a transcript defective (truncations or shortened transcripts, for example), this reaction provides a repair method for these transcripts, one that can be explored from a therapeutic standpoint. This reaction has been implemented *in vivo* as a correction method with sickle cell globin, the p53 transcript (heavily linked to cancer), and various other disease-linked genes (64, 74-75).

2.4.5 Trans excision-splicing reaction

Many studies have and continue to utilize the *Tetrahymena* group I intron and its derived ribozymes; however, various other group I introns have been discovered and investigated, as have the ribozymes derived from them (76-78). The *Pneumocystis carinii* group I intron, reported in 1989, can be found in the rRNA of this fungal entity (76). It is an opportunistic pathogen that commonly causes problems in immunocompromised patients (76). A ribozyme derived from its group I intron, however, has been reported to catalyze a novel reaction with

therapeutic possibilities (33). This reaction, termed trans-excision splicing (TES), is shown in Figure 2.14 and it results in the excision of 1 or more bases from an RNA substrate (33). Like the group I intron from which it is derived, it also utilizes the P1, P9.0, and P10 interactions in its catalysis. Just as in the previous diagrams, the exons are shown in black and the ribozyme in gray. The P1 helix forms between the RE1 in the ribozyme and the 5' exon of the substrate; base-pairing between the RE3 region and the 3' exon forms the P10 helix. P9.0 forms between RE2 and the bridge region of the substrate (33). The first step of the reaction, the 5' cleavage, occurs via a nucleophilic attack from a free hydroxyl at the first catalytic site (after the aforementioned G•U base-pair). Note that this differs from the first step of trans-splicing in that no exogenous guanosine is required for the 5' cleavage step. The second step of exon ligation transpires as previously described, through a nucleophilic attack by the 3' OH of the terminal uridine on the 5' exon. This leaves the excised bridge region and a new RNA (33). While this novel reaction also provides new information on ribozyme-catalyzed reactions, its therapeutic potential involves insertion mutations or repeat regions of nucleotides (in cases like Muscular Dystrophy) that are linked to disease (1, 3).

2.4.6 Targets for group I intron-derived ribozyme reactions

In terms of RNA repair or alteration, both the TS and TES reactions can cater to a variety of needs. Some cases of genetic mutations exist, however, in which neither of these methods present the most viable strategy. One case in particular involves the k-ras gene. This gene belongs to the ras family of genes that are involved in aspects of GTP-binding and activation in cellular processes (1, 79-81). The main three types of ras genes are n-ras, h-ras, and k-ras. Dysfunctional variants of each can be linked to different cancers, such as colorectal, pancreatic, and lung (79). The ras protein products are all approximately 188-189 amino acids in length and differences in sequence usually do not appear until after amino acid 85 (79, 81). The mutation hotspots in this gene are codons 12, 13, and 61 and usually contain substitution mutations, with codon 12 being most commonly reported (80). The native glycine (GGT/GGU) of codon 12 is required for the proper coding function of this transcript. Its mutation to any other amino acid or its deletion or insertions at that site prove detrimental to the resultant gene product (79-81). This mutation can be found in patients with certain types of cancer and the presence of the mutations in cells results in oncogenic (cancerous) transformation when studied in cell cultures (79-80).

While all three maintain a significant degree of homology in terms of sequence and size, their roles in living systems appear to be quite delineated (81). Investigations into the necessity of each of the forms of ras have presented evidence that k-ras bears a significant role in living organisms (81-83). Transgenic animals can survive gestation and function with minimal problems when the h-ras or n-ras genes are knocked out, or deleted (82). The deletion of the k-ras gene, however, produces embryonic lethality in those animals; transgenic mice without k-ras do not survive past day 14 of gestation (81, 83). Thus, a functional k-ras transcript and protein product prove essential to animal systems. Dysfunctional k-ras transcripts, however, evidence themselves in 25-50% of lung cancer cases (84-85).

In terms of possible RNA repair methods, the TES and TS reactions do not provide the most ideal strategies for correcting mutations in codon 12 of the k-ras gene. Since removal of the codon proves just as deleterious as a substitution mutation, trans-excision splicing cannot serve as a repair method. The location of the mutation, close to the 5' end of the RNA transcript, renders ribozyme 3' replacement splicing, or trans-splicing, a less than facile choice for correction. Therefore, the development of a novel ribozyme-catalyzed reaction that can replace the 5' end of an RNA transcript would be a viable therapeutic option for the case of k-ras, as well as other disease-linked mutations, such as Huntington's Disease (86). It would also lend new insight into reactions catalyzed by group I intron-derived ribozymes.

By exploiting a 3' recognition element (involved in P9.0 formation), to target the transcript substrate, a ribozyme was engineered that can effectively replace the 5' end of a mutant RNA with a corrected version in a k-ras small model system. This reaction is termed the 5' replacement splicing (RS) reaction and a general diagram of it is shown in Figure 2.15; the ribozyme (derived from the *P. carinii* intron) is shown in gray with the new 5' exon depicted as an open line. The exogenous exons are shown in black. The P1 and P10 helices form with the new 5' exon and the 3' exon of the exogenous substrate, respectively. P9.0 forms between the 5' exon of the exogenous substrate and the RE2 of the ribozyme. The first step of 5' cleavage occurs through a free hydroxyl and separates the new 5' exon from the rest of the ribozyme. This differs from the TS reaction in that it does not need a guanosine cofactor for 5' cleavage. The second step of exon ligation proceeds as it does in the self-splicing reaction. This study resulted in the first report of a ribozyme-catalyzed 5' transcript replacement reaction, as well as the first ribozyme designed to utilize a 3' recognition element (RE2) as the primary determinant of

substrate targeting. The reaction was probed in terms of the effect of length dependence of the P9.0 and P10 helices, and the L1 loop. This 5' replacement splicing system was also compared with trans-splicing in terms of reaction rate and the effect of guanosine cofactors. These findings, as well as the full system design for the reaction will be fully discussed in this dissertation.

2.5 Experimental methods used for nucleic acids

2.5.1 Gel electrophoresis

The main method of analysis used in these studies is gel electrophoresis. Electrophoresis can be used to separate biomolecules (nucleic acids and proteins) on the basis of size; its medium is a gel matrix that can be agarose or polyacrylamide (3, 56). Theoretically, the molecules should move with respect to charge, size, and shape. With nucleic acids, charge and shape are consistent; thus size is the determining factor in migration. The molecules move courtesy of a constant electric field. These gels are "run" in this field vertically or horizontally. PAGE (polyacrylamide gel electropheresis) can be used with nucleotides when a difference in size down to single nucleotides needs to be seen. The gel is usually set between two glass plates and held there while it is run. The nucleic acid samples are applied to the gel at one end and migrate under the influence of the electric field (3, 56). Smaller molecules move faster and so move farther down the gel, while larger molecules remain higher up on the gel. One of the most common methods for visualizing and analyzing PAGE gels is autoradiography.

2.5.2 Autoradiography

Autoradiography utilizes two components, radioactive specimens and photographic emulsions (56). The specimen is usually radiolabeled with radioisotopes such as ^{35}S, ^{33}P, and ^{32}P, which emit β-particles as the unstable atoms decay (3, 56). The emulsions are suspensions of silver halide (AgCl, AgFl, AgBr, AgI) in gelatin. These emulsions usually undergo a process that places them on a thin surface, like X-ray film. After the radioactive specimens are fixed on a two-dimensional surface, the emulsion is exposed to it. Each β-particle emitted from the specimens converts some of the silver ions to silver atoms. This removes them from the crystal lattice of the emulsion. This creates an image, or a hole in the lattice corresponding to that area containing radioactive specimens (2-3, 56). The atoms of silver are likely to lose their electrons and return to the lattice, so the image is likely to fade. Thus, it must be developed before the

image is lost. The development process results in images or bands corresponding to the orientation of the specimens on the gel (56).

Traditional autoradiography requires X-ray film; however, the process of phosphorimaging provides an additional method of visualizing radioactive reactions. Plates, or screens, are coated with a light-responsive phosphor. The screen is exposed to the gel and the energy from the β-particles is stored in a europium-based coating (56). The screen is scanned by a laser and photons are released, and collected and used to form an image. This image is presented on a computer screen. This requires a more expensive instrument and specific software that can manipulate the image and allow for quantification of the bands. While more expensive than film, the time to acquire images is reduced 80-90% and a darkroom is not necessary (56). The images of PAGE gels shown in these studies have been acquired using this method.

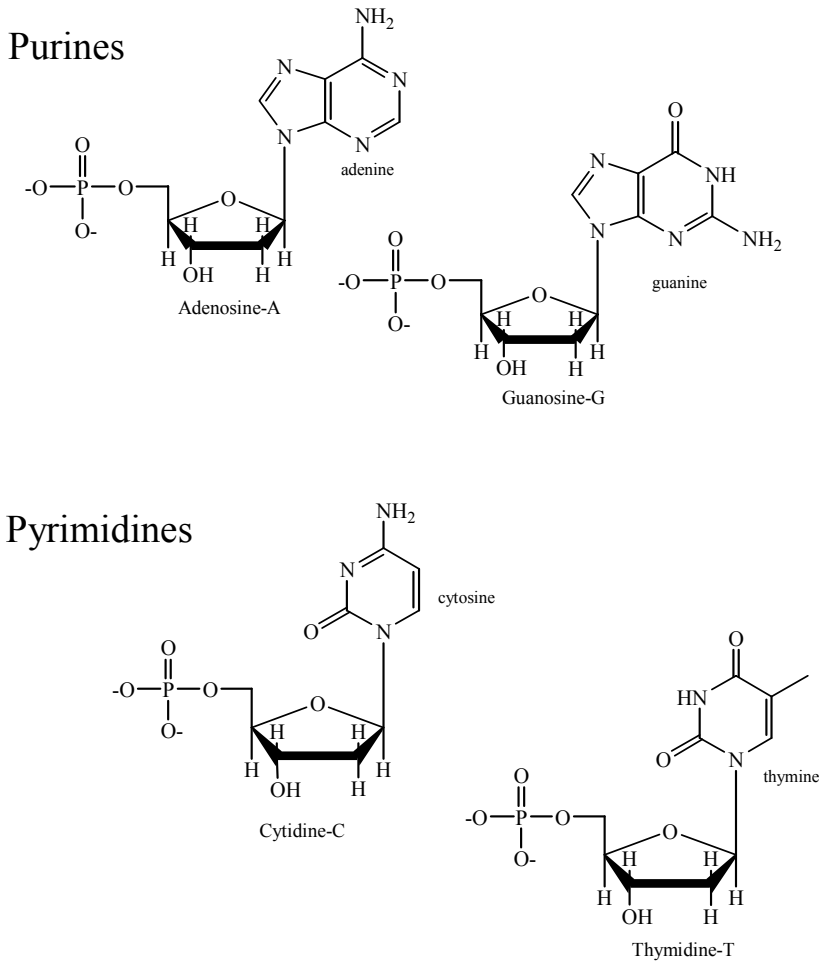

Figure 2.1. DNA nucleotides. Each of the nucleotides contain a nitrogrenous base, which can be a purine or a pyrimidine, as well as a sugar (ribose) and a phosphate group. The nucleotides names and single-letter abbreviations, shown below each structure, corresponds with the identity of the purine or pyrimidine (which is located next to the base).

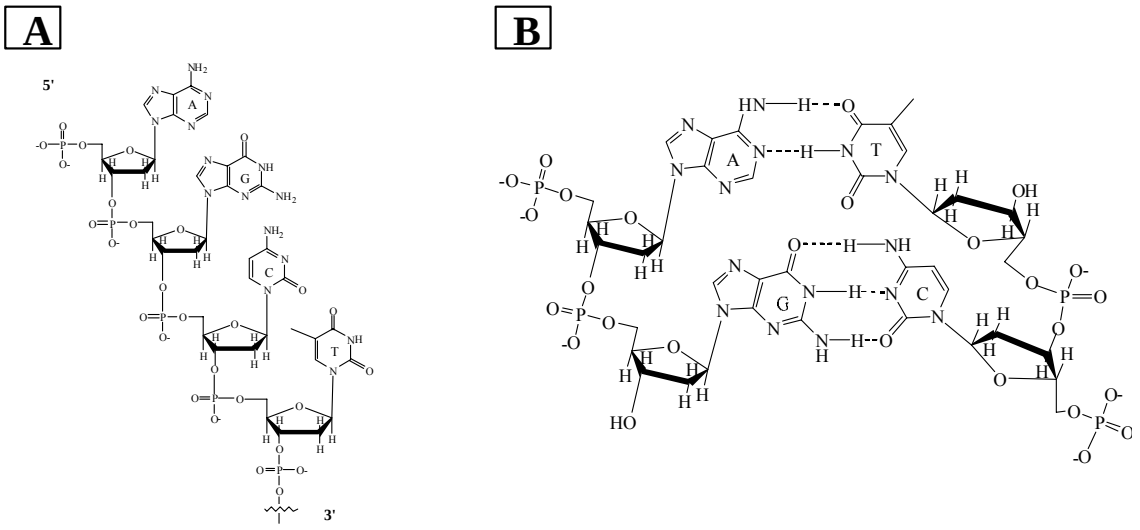

Figure 2.2. DNA structure. **(A)** A single strand or chain of DNA is shown. The nucleotides are linked through the phosphate group of one to the 3' OH of another. The directionality of this linkage is termed 5' to 3'. **(B)** A double strand or DNA duplex is depicted; the nucleotides interact through hydrogen-bonding between the purines and pyrimidines. This interaction is termed base-pairing; A pairs with T and C with G. Note that the strands are antiparallel, with one in the 5' to 3' direction and the other in the opposite 3' to 5' direction.

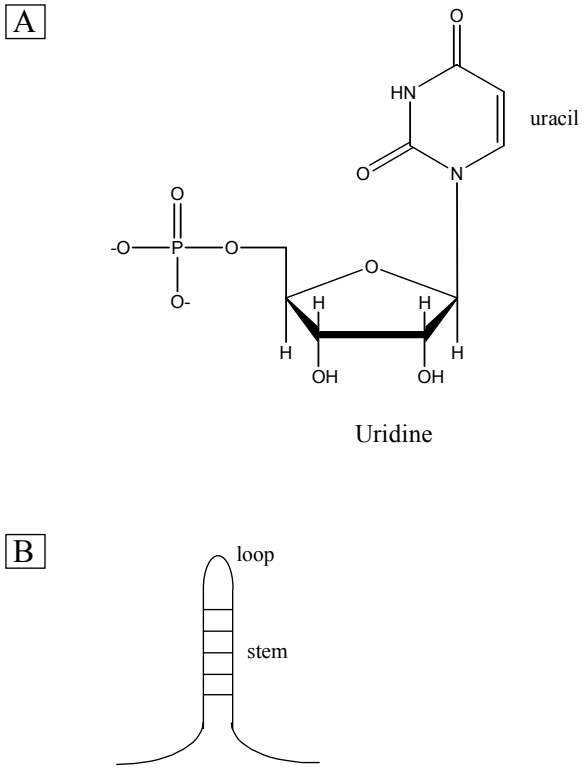

Figure 2.3. RNA structure. **(A)** The nucleotide uridine, which replaces thymidine in RNA, is shown. Note that in RNA, the 2' position of the sugar has a hydroxyl group, whereas DNA has only a hydrogen at the corresponding position. **(B)** A common secondary structure for RNA is depicted as a line diagram, with a base-paired stem region and an unpaired loop region.

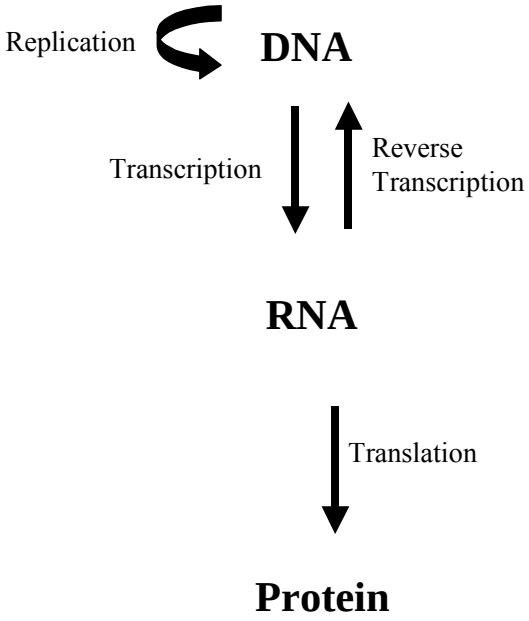

Figure 2.4. Central dogma of molecular biology. The progression from DNA to RNA to protein is shown. The replication of DNA is synthesis; the DNA can be transcribed into RNA (RNA can also serve as a template for the reverse transcription of DNA). The RNA is then translated into protein.

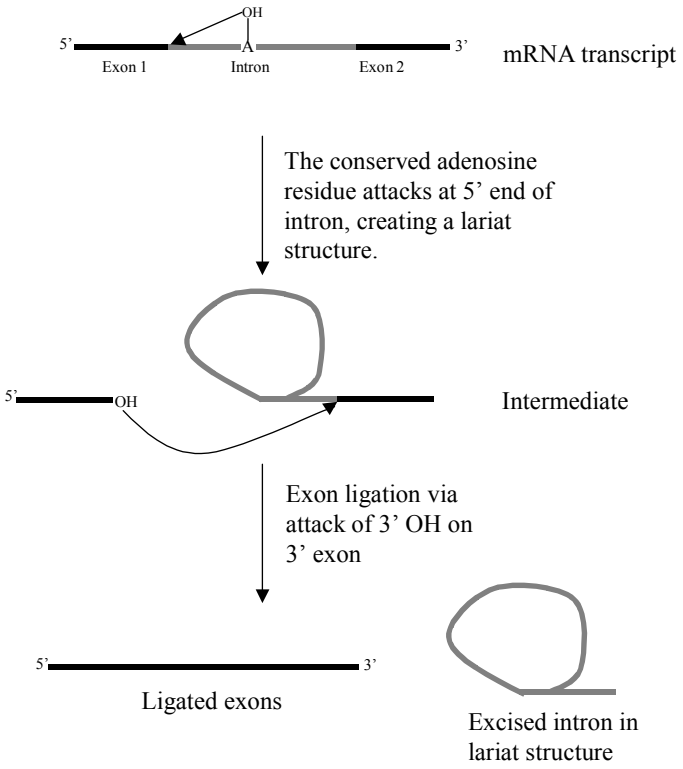

Figure 2.5. Removal of introns via spliceosomal splicing. The exons are shown as black lines, while the intron is shown as a gray line. The spliceosome (a ribonucleoprotein construct), though not shown, mediates the splicing. In the first step, a conserved adenosine attacks the 5' end of the intron, resulting in a lariat, or loop structure. The exons are ligated through the attack of the 3' OH on the 3' exon. The exons are joined, leaving begind the excised intron.

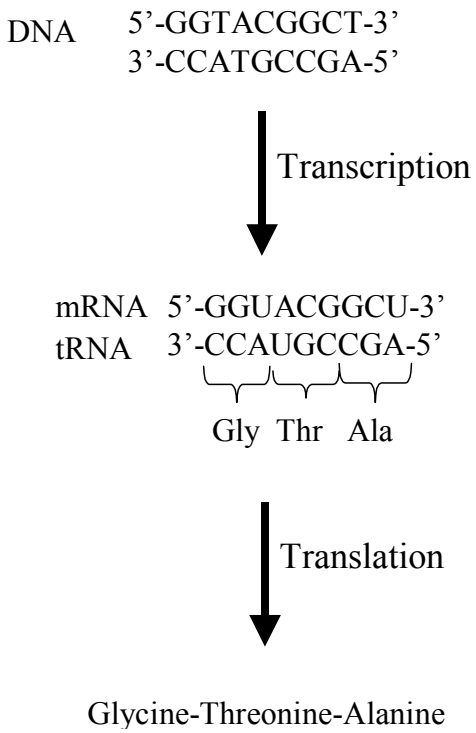

Figure 2.6. Transcription and translation. A given sequence of double-stranded DNA is transcribed into mRNA (messenger RNA); this sequence contains sequences of three nucleotides that specify a particular amino acid. These sequences are also termed codons. During translation, the codons in the mRNA pair with their respective complements, which are present in the tRNA (transfer RNA). The tRNA also has the amino acid that the codon codes for attached to it. The ribosome (a ribonucleoprotein construct not shown) catalyzes the joining of the amino acids to give the resultant protein.

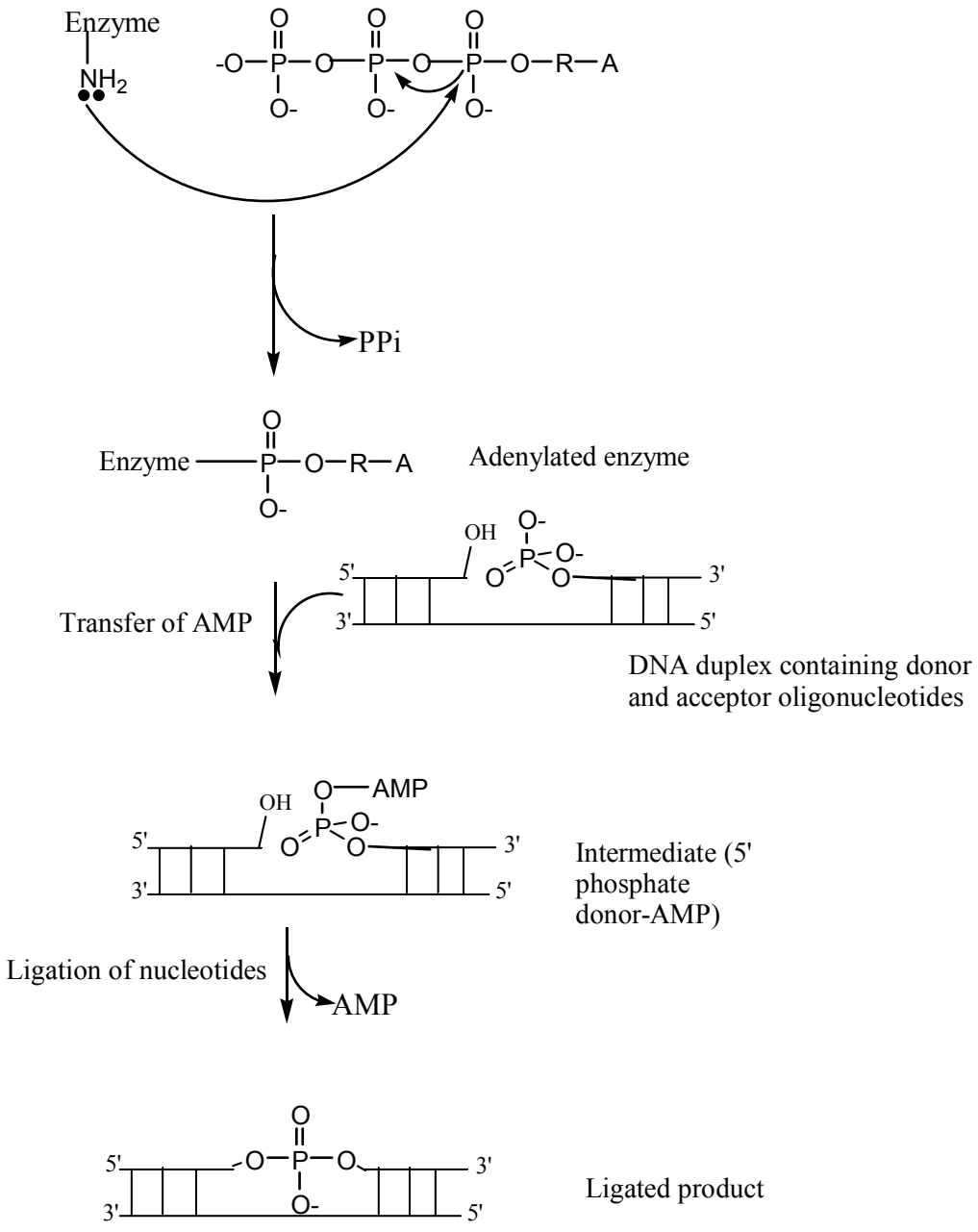

Figure 2.7. DNA ligation mechanism. The active lysine of this enzyme contributes an amine group, which acts as a nucleophile in the first step of enzyme adenylation. The cofactor, ATP, contributes AMP, which is joined to the enzyme and PPi is released. In the next step, this AMP is transferred to the donor oligonucleotide, thus forming an intermediate. The site of ligation shown between the donor and acceptor oligonucleotides is also termed the nick. In the last step, the donor and acceptor oligonucleotides are ligated together and AMP is released.

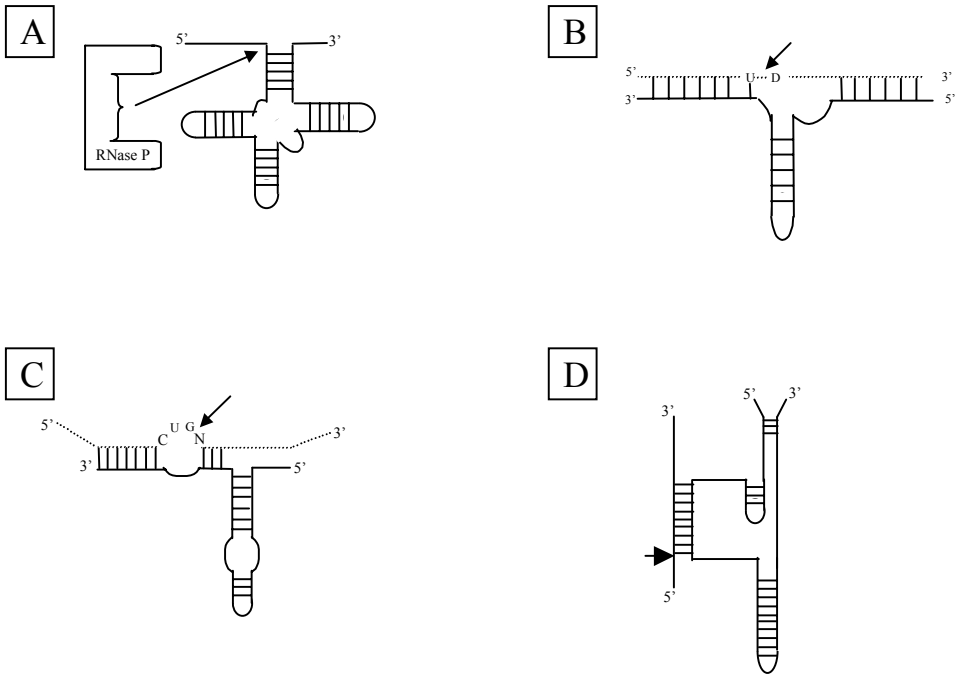

Figure 2.8. General secondary structures of and cleavage schemes for catalytic RNAs. The substrate sequences for each ribozyme are shown as dotted lines, while the sites of cleavage are specified with arrows. N denotes A, C, G, or U; D specifies A, C, or U. **(A)** RNase P (composed of proteins and RNA) cleavage **(B)** Hammerhead ribozyme cleavage **(C)** Hairpin ribozyme cleavage **(D)** Hepatitis delta virus cleavage.

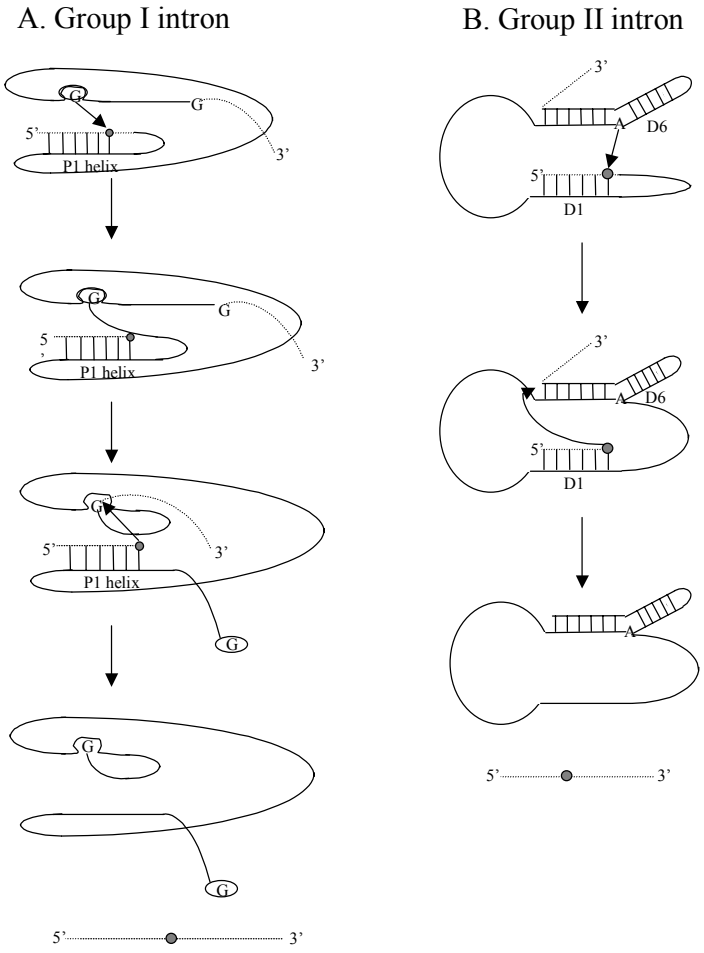

Figure 2.9. Simple schematic of group I and II intron splicing. The exons are shown as dotted lines and the intron as a solid line. **(A)** The self-splicing reaction of group I introns requires a guanosine cofactor that is depicted in the conserved guanosine-binding site (GBS). This guanosine attacks at the 3' end of the P1 helix and is added to the 5' end of the intron (G-addition). The conserved guanosine on the 3' end of the intron (ωG) moves into the GBS. The 3' OH of the terminal base on the 3' end of the 5' exon attacks the 3' exon, resulting in exon ligation and the excision of the intron. **(B)** The self-splicing reaction of group II introns entails a conserved A within domain 6 (D6). This adenosine attacks at the 5' splice site, (defined by domain 1, or D1), yielding a lariat intermediate. The second step of exon ligation proceeds as it does in group I intron self-splicing, with an attack from the terminal base of the 5' exon on the 3' exon.

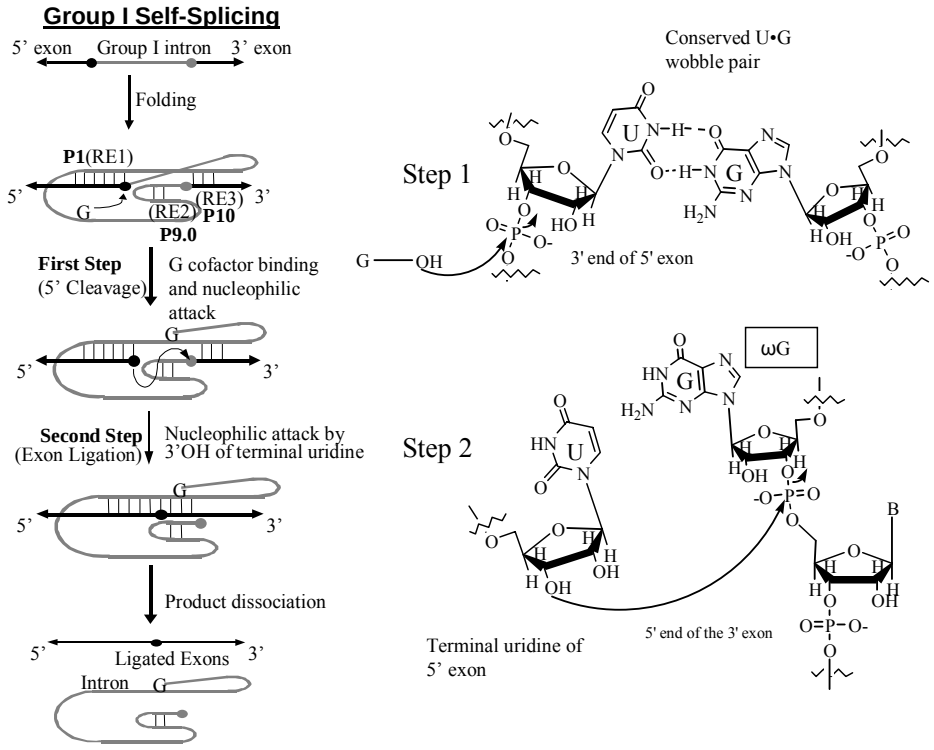

Figure 2.10. Group I intron self-splicing diagram. The intron is depicted as a gray line, while the exons are black lines. P1, P9.0, and P10 are helices that result from the recognition elements RE1, RE2, and RE3 (respectively) of the intron base pairing with the exons and itself. P1 forms between RE1 and the 5' exon, P10 forms between the RE3 and the 3' exon, and P9.0 forms intramolecularly in the intron between the RE2 and the 3' intron. A conserved guanosine-uridine base-pair (located at the end of the P1 helix) helps to define the first catalytic site, which is the 5' splice site. This step of 5' cleavage occurs via a nucleophilic attack by the required guanosine cofactor at the 5' splice site and is shown in more detail on the right. The second step of exon ligation involves the conserved guanosine at the 3' end of the intron, also termed ωG; this guanosine helps to define the 3' splice site. The 3' OH of the 3' terminal uridine on the 5' exon attacks at the 3' splice site, resulting in exon ligation and intron excision.

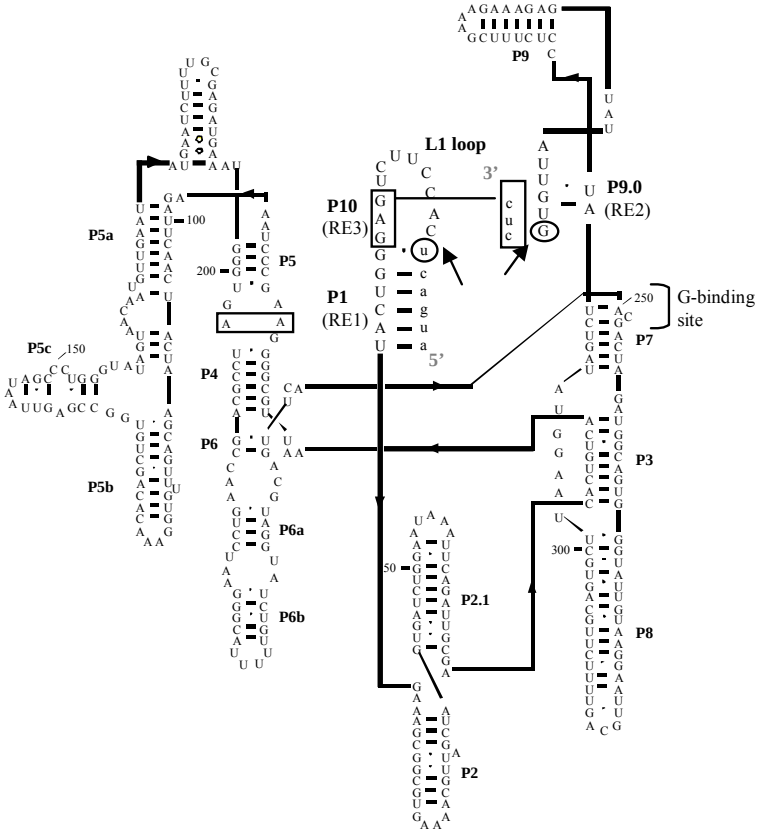

Figure 2.11. Secondary structure of *Pneumocystis carinii* group I intron. The sequence of the intron is shown in uppercase letters and that of the exons in lowercase. The uridine of the conserved guanosine-uridine base-pair and the ωG are circled, with the splice sites denoted with arrows. The guanosine-binding site is also noted.

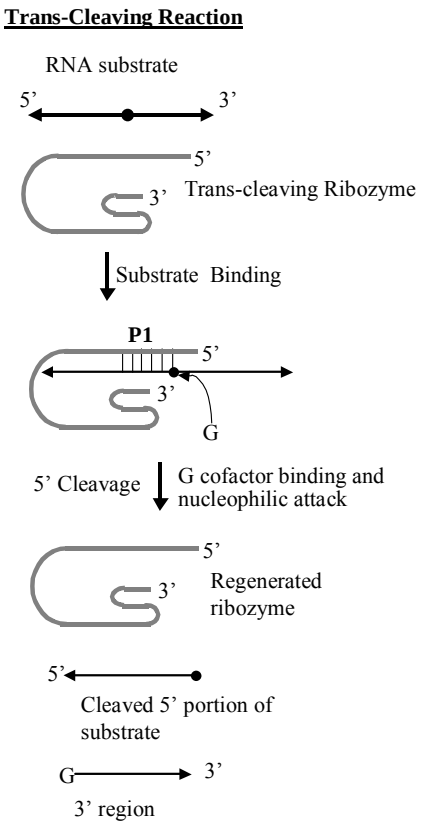

Figure 2.12. Trans-cleaving reaction diagram. The catalytic RNA is represented by a gray line, the 5' and 3' regions are black lines. The black circle represents the uridine at the 5' splice junction. This reaction also utilizes the P1 helix mentioned in the self-splicing reaction of group I introns. P1 forms between RE1 and the 5' portion of the substrate. A guanosine cofactor is required for the trans-cleavage. The reaction results in the substrate being cleaved at the 3' end of the P1 helix and the ribozyme is left unchanged.

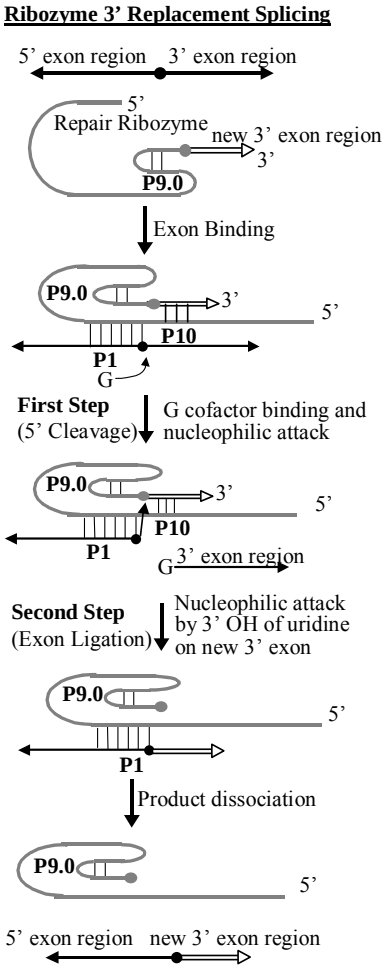

Figure 2.13. Trans-splicing or ribozyme 3' replacement reaction diagram. The catalytic RNA is represented by a gray line, the 5' and 3' exons are black lines, and the replacement exon carried by the ribozyme is represented by an open arrow. The black circle in the 5' exon represents the uridine at the 5' splice junction and the gray circle adjacent to the 3' exon represents a guanosine at the 3' splice junction. This reaction utilizes those same interactions and elements associated with the self-splicing reaction of group I introns. P1 forms between RE1 and the 5' exon of the exogenous substrate and P9.0 forms within the intron, as it does in the self-splicing reaction. A guanosine cofactor is required for the first step of 5' cleavage. The P10 helix then forms between RE3 and the new 3' exon. Exon ligation preceeds as it does in the self-splicing reaction, resulting in the generation of a new RNA strand composed of the original 5' exon and the new 3' exon

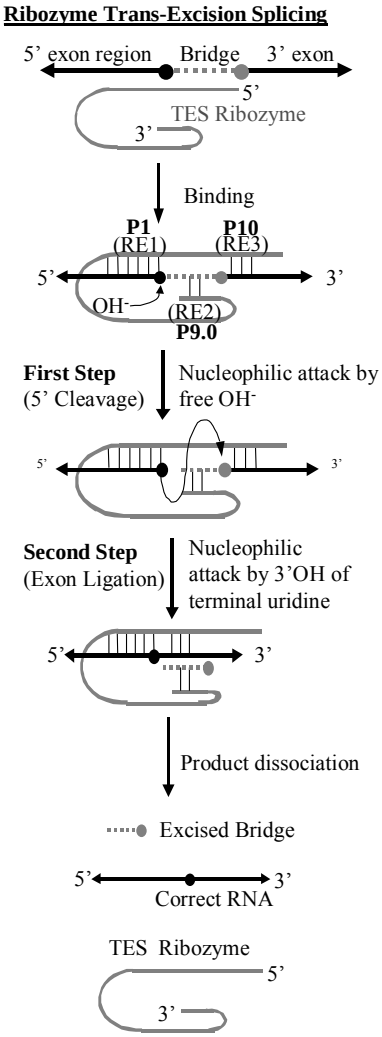

Figure 2.14. Trans excision-splicing reaction diagram. The catalytic RNA is represented by gray lines, the 5' and 3' exons are black lines, and the bridge region to be excised is shown as a dotted line. The black circle in the 5' exon represents the uridine at the 5' splice junction and the gray circle adjacent to the 3' exon represents a guanosine at the 3' splice junction. This reaction utilizes the same interactions and elements associated with the self-splicing reaction of group I introns. P1 forms between RE1 and the 5' exon, P10 forms between the RE3 and the 3' exon, and P9.0 forms between the RE2 and the bridge region. The 5' cleavage step does not require a guanosine cofactor; it utilizes a free hydroxyl in the nucleophilic attack. Exon ligation occurs as it does in the self-splicing reaction of group I introns.

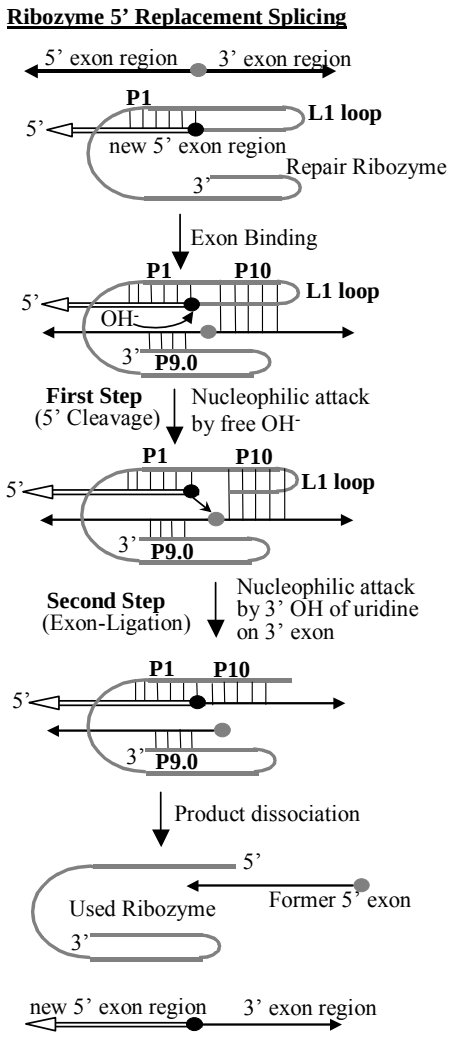

Figure 2.15. 5' Replacement splicing reaction diagram. The catalytic RNA is represented by gray lines, the 5' and 3' exons are black lines, and the replacement exon is represented by open arrows. The black circle in the 5' exon represents the uridine at the 5' splice junction and the gray circle adjacent to the 3' exon represents a guanosine at the 3' splice junction. This reaction utilizes those same interactions and elements associated with the self-splicing reaction of group I introns. P1 forms between RE1 and the new replacement 5' exon, P9.0 forms between RE2 and the 5' exon of the exogenous substrate, and P10 forms between RE3 and the 3' exon. The 5' cleavage step does not require a guanosine cofactor; it utilizes a free hydroxyl in the nucleophilic attack. Exon ligation occurs as it does in the self-splicing reaction of group I introns.

Chapter 3 - Study of Canonical Bases as Universal Templates with T4 DNA Ligase

3.1 Materials and Methods for the Universal Template Study

3.1.1 Oligonucleotide Synthesis and Preparation

The oligonucleotides used were synthesized either on an Applied Biosystems 380B DNA Synthesizer or purchased from Integrated DNA Technologies (IDT, Coralville, IA). Oligonucleotides were purified either by reverse phase HPLC (IDT) or Thin Layer Chromatography, as described (31, 87-88) Designated oligonucleotides were 5' end radiolabeled by incubating 0.4 μM DNA, 0.43 μM [γ-^{32}P] rATP (Amersham Pharmacia, Piscataway, NJ), 70 mM Tris HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM EDTA, 0.1 μM rATP, 10% glycerol, and 10 units of T4 polynucleotide kinase (New England Biolabs, Beverly, MA) in a reaction volume of 10 μL for 45 minutes at 37 ºC. The products were isolated and purified via a 17% native polyacrylamide gel, using 0.25x TBE as the running buffer (1x TBE is 100 mM Tris, 90 mM boric acid, and 1 mM EDTA at pH 8.4). The products were then isolated from the gel matrix via overnight pulverization in 2 mL sterile water with a stir bar, followed by centrifugation to partition the gel matrix from the solution. The radiolabeled product solutions were then evaporated to a final concentration of 8 nM.

3.1.2 Ligation Reactions

All ligation reactions contained 3.1 pmol template, 2.5 pmol non-radiolabeled oligonucleotide (the acceptor molecule), 70 fmol radiolabeled oligonucleotide (the donor molecule), and 3 units T4 DNA ligase (Promega, Madison, WI.). In all ligation reactions the template and donor molecules were mixed in an appropriate buffer solution and denatured at 90 ºC for 30 sec. After a 5 min annealing period at room temperature, the ligase and the acceptor molecules were added to the reaction mixture (10 μL total volume). This was then incubated for 18 h at an appropriate temperature (listed in the figures and tables). The reactions were terminated by adding 7 μL stop buffer (10 mM urea, 0.1x TBE, 3 mM EDTA). Initial reactions were conducted in Promega's standard recommended conditions, which consisted of 30 mM Tris-HCl (pH 7.8), 10 mM DTT, 10 mM $MgCl_2$, and 1 mM rATP at 30 ºC. These conditions were subsequently optimized to test for universal templating abilities by altering the $MgCl_2$ concentration (varied from 1 mM to 100 mM), the ATP concentration (varied from 0.25 μM to

10 mM), time (varied from 1 hr to 24 h), temperature (varied from 4 ºC to 44 ºC), and by including additives such as dimethyl sulfoxide (DMSO; Malinckrodt), polyethylene glycol-8000 (PEG; Sigma-Aldrich), and single-stranded binding protein (Promega; from 5% to 50%). The final optimized buffer conditions were 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 μM ATP, and 20% DMSO at either 30 ºC or 22 ºC for 18 h. In certain cases, 1 unit of calf intestinal alkaline phosphatase (CIAP; Promega) was added to individual reactions. Reaction products and intermediates were isolated using 11% denaturing polyacrylamide gels and the reactions were quantified using a STORM 860 Molecular Dynamics Phosphorimager. Observed rate constants were quantified by fitting the time-dependent data to a simple exponential function.

3.1.3 Thermal denaturation experiments

Thermal denaturation experiments were run using a standard thermodynamic buffer [20 mM sodium cacodylate, 1 M NaCl, and 0.5 mM Na_2EDTA at pH 7.0] on a Beckman DU 650 UV-VIS Spectrophotometer using a 15 μM total strand concentration (89-90). Measurements were taken using a wavelength of 260 nM.

3.2 System Design

The test system for this project was designed to quantify ligation as a function of the sequences being ligated. In the diagram of the system shown in Figure 3.1, the oligonucleotides to be ligated are decamers that begin and end with the same nucleotide. Phosphorylated (represented by 'p') decamers are radiolabeled (represented by an asterisk) and are in italics. The names of the decamers to be ligated are represented by lowercase lettering according to what nucleoside (from 5' to 3') is paired with the universal template positions (in the grey box). The template names are in uppercase and represent the central universal template positions (underlined). On the left side, studies with the *n* position (at the actual nick) are shown. On the right, studies with the *n-1* position (adjacent to the nick) are illustrated. The acceptor molecules contain a 3' OH group and the donor molecules contain a radiolabeled 5' phosphate. Once the decamers bind to a template, T4 DNA ligase catalyzes the formation of a covalent bond between the 5' phosphate and 3' OH group (See Figure 2.7 and Example in Figure 3.1). By having only one decamer that can act as a donor molecule and using an excess of the acceptor molecule

(which contained only hydroxyl groups on both ends), the sequence of oligonucleotides being ligated was controlled. Template oligonucleotides possessed a centrally located region with of a single type of nucleoside in tandem, which flanked the ligation junction when paired to the donor and acceptor molecules. Thus, ligation was analyzed as a funtion of template sequence. All oligonucleotides are written in the 5' to 3' direction, save when denoted otherwise. In this report, the focus was on the sequences at the ligation junction and the *n-1* positions.

3.3 Results

3.3.1 Investigation of the n Position - Standard Conditions

All 16 dinucleotide *n* position combinations were explored with 4 different templates; these templates had either A, C, T, or G as the template nucleosides at the site of ligation. For this, T4 DNA ligase (Promega) was used under the standard manufacturer's conditions. A ratio of 1/0.8/0.02 template, non-radiolabeled acceptor, and radiolabeled donor, respectively, gave the best results (data not shown). A representative gel, using the T2 template is shown in Figure 3.2. The standard reaction conditions were 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.125 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h. P represents product, I denotes adenylated intermediate, and S represents starting material; these designations are located on the right side of the gel. The reaction components are shown above each lane. The first lane contains a generic 25 base size control, the next four lanes are the four unreacted donor molecules, and the next two lanes are the synthetic (a + $_p$a) and (c + $_p$a) size controls. The next four panels show the 16 possible reactions. The next three lanes show control assays with the ligation of (a + $_p$a) without template, with template and treated with calf alkaline intestinal phosphatase (CIAP) after the reaction has gone to completion, and without ligase. All experimental indications inferred that these assays used were functioning properly, evidenced by the number and size of the products. The ligation of (c + $_p$a) yielded an intriguing product. The product ran noticeably lower than the other ligation products, as shown in Figure 3.2. This trend also occurred with each of the four templates (data not shown). A synthetic size control of the (c + $_p$a) product was run on the gel and also migrated faster than other products. Apparently, this specific sequence exhibited atypical behavior in terms of its migration. Observe that the (a + $_p$a) and (c + $_p$a) size controls ran marginally faster than the actual ligation products, which was attributed to the presence of 5' phosphates in the

controls that were absent in the actual ligation products. Further indications that the assays were working correctly were evidenced by the results of the control assays. The formation of both products and intermediates happened in a template-dependent manner. Moreover, dephosphorylation of the product band did not produce any notable reduction in intensity. This would be expected if the product was only internally radiolabeled. In other words, two radiolabeled donor molecules were not being ligated together.

Table 3.1 lists the results of duplicate trials with standard deviations using each of the four templates. The data shows percent product formation (in bold) and percent intermediate formation (in parenthesis). The horizontal heading denotes template used and vertical heading denotes the new ligated product. Refer to Figure 3.1 for designations of acceptor, donor, and template oligonucleotides. When complete complementarity occurred in the duplex, the ligation extent was 95% or more, which was anticipated. When one mismatch occurred between the universal template region and the ligation junction, most ligations resulted in greater than 80% product (the few exceptions still reacted over 45%). When no complementary base pairing occurred at the ligation junction, the results varied widely; some dinucleotide combinations were ineffective (for example, 11% for (a-$_p$c)/A2), while some were quite effective (for example, 77% for (c-$_p$a)/A2). Each of the canonical nucleosides displayed at least some ability to template the ligation of all 16 dinucleotide combinations. When judged by the average product generated as a function of template, the effectiveness of each template reacting with each of the 16 possible sequence combinations followed the order G2 (63%) ≅ T2 (62%) > A2 (58%) > C2 (47%).

3.3.2 Investigation of the n Position - Optimized Conditions

The previous results illustrated that T4 DNA ligase is a sequence-dependent enzyme when used under its standard conditions. In cases of low product formation, however, adenylated intermediate formation was usually substantial. For product formation to increase, only the second step of the reaction had to be boosted (the ligation step in Figure 3.1). Changes in three components of the reaction resulted in augmented product formation. The optimization data is shown in Figure 3.3; product is represented by squares and intermediate by circles. As evidenced in panel A, the optimum concentration of 3 mM $MgCl_2$ was found using the previously mentioned reaction conditions. Subsequently, it was implemented in the ATP-dependent assay in panel B. Both of these factors were optimized using (c+$_p$c)/T2 as a test reaction. The data from

these assays showed that the optimized $MgCl_2$ and ATP concentrations enhanced product formation of (c + $_p$c)/T2 from 12% (Table 3.1) to greater than 90%. Another ligation reaction with low product formation, (c + $_p$c)/C2, was also tested with the optimal $MgCl_2$ and ATP concentrations. Yet, (c + $_p$c)/C2 ligated no better using these conditions (data not shown). Additives to the reaction were then evaluated for their effect in improving product formation. DMSO was tested using 3 mM $MgCl_2$ and 10 μM ATP and its inclusion in the reaction (c+$_p$c)/C2 improves the ligation from less than 10% to greater than 90% (Figure 3.3C). PEG, a molecular crowding agent, was also considered, but was not found to be beneficial. The addition of SSB also had no beneficial effect on the reaction (data not shown). Time-dependence assays were run using (c+$_p$c)/T2 in 3 mM $MgCl_2$, 10 μM ATP, and 20% (w/v) DMSO. Figure 3.3D shows this time-dependence plot of the reaction. This did indicate that no product breakdown transpired over time.

Figure 3.4 shows a representative gel under the final optimized conditions with the T2 template; the optimum reaction conditions were 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 μM ATP, 20% DMSO (w/v), 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.125 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h. The same designations used in labeling the reactions and bands in Figure 3.2 were also used in Figure 3.4. Table 3.2 shows the results of duplicate trials and the standard deviation values of all 64 combinations of decamers and template under the optimized conditions. The horizontal and vertical heading designations for Table 3.1 are also used here. The percent of product formed is shown in bold, while the percent of intermediate formed is in parenthesis. These optimized conditions considerably augmented the ligation of mismatched nucleosides at the site of ligation. Moreover, they did not decrease the effectiveness of ligation with matched base pairs. The average for the 16 dinucleotide combinations with each template is T2 (97%) ≅ G2 (96%) > A2 (93%) ≅ C2 (92%). Under these optimized conditions, all four of the canonical nucleosides directed the effective ligation of all 16 possible sequence combinations at the *n* positions. A-T pairs at the *n-1* positions were also tested for effective ligation and were not found to hinder the ligation extent (data not shown). Therefore, C-G pairs were not required for successful ligation.

3.3.3 Investigation of the n-1 Position – Optimized Conditions

The next step in analysis focused on whether these optimal experimental conditions could allow ligation of decamers that are mismatched at the positions adjacent to the ligation junction (the *n-1* positions in Figure 3.1) when the *n* position was base paired. For this investigation, the templates contained a central tandem segment of 4 identical nucleosides. The results, displayed in Table 3.3 in duplicate, showed that these optimized experimental conditions did encourage effective ligation when mismatches at positions adjacent to the ligation junction were present. The percent product formation and percent intermediate formation are shown as in previous tables, along with the standard deviations. The horizontal and vertical headings again denote template used and the new ligated product, respectively. The x shown in the vertical heading is defined in the horizontal heading. Refer to Figure 3.1 for naming designations. Nearly all combinations worked well, but 4 out of 64 yielded less than 50% product. Lowering the reaction temperature from 30 ºC to 22 ºC (denoted with ♦ in Table 3.1), in these instances significantly improved these reactions. The reaction (gt-$_p$ta)/A4 went from 6.2% to 83.1%, (gt-$_p$tc)/A4 went from 24.6% to 87.2%, (cg-$_p$gc)/C4 went from 34.7% to 90.9%, and (tg-$_p$gc)/C4 went from 42.2% to 97.4%. This drop in the reaction temperature conceivably benefited the hybridization of the decamers to the template and was not presumed to appreciably hamper the other reactions. The effectiveness of the template nucleosides was similar to what is seen in the *n* position; T4 (92%) ≅ G4 (92%) > A4 (80%) ≅ C4 (79%).

3.3.4 Investigation of Mismatches at Both the n and n-1 Positions

Mismatches at both the *n* and *n-1* positions (either in the donor molecule, the acceptor molecule, or both) underwent scrutinization. Analyzing all 256 possibilities (using a homogeneous universal template region) was not feasible, so a subset of 16 representative sequence combinations for each template served as general test reactions (Table 3.4). For this system the reaction temperature was 22 ºC, as it appeared to work better than 30 ºC in cases of low product formation and was assumed to be a good initial temperature for a double mismatch study.

When no mismatches occurred (dark gray background in Table 3.4), the extent of ligation was 89%-96%, as anticipated. The same designations used in previous tables are also used in Table 3.4; the results from duplicate trials are shown with the standard deviation values. When 2

successive mismatches occurred either with the donor and template or the acceptor and template (light gray background), ligation generally proceeded well (5% to 97%), except for (gg-$_p$cc)/C4 (31%). In cases of mismatches at all four positions, the extents of reaction varied widely. Using the A4, C4, and G4 templates, some donor-acceptor combinations yielded low amounts of product, less than 3% ligation product. With the T4 template, however, substantial ligation for all sequence combinations occurred. Reaction temperatures lower than 22 ºC did not boost product formation in these reactions further. Decreased ligation effectiveness as the number of mismatches increased was not unexpected, as each mismatch probably contributed to helix destabilization and cases "fraying" at the ends of the duplexes. Regardless, these results showed that up to four consecutive mismatches flanking the ligation junction can be effectively ligated.

In these assays, mismatches and/or overhangs occurred at the termini not being ligated (refer to Figure 3.1). To ascertain whether or not this was a factor in the reactions, systems that mimic (ta-$_p$at)/T4 and (tt-$_p$tt)/T4, except that the non-ligating terminal ends of each substrate were shortened and were completely complementary with the template were examined. The non-ligating terminal mismatches did not significantly affect our results, as the extent of ligation for (ta-$_p$at)/T4 went from 95.6% in the end-matched system to 96.8% in the end-mismatched system. The reaction (tt-$_p$tt)/T4 went from 24% in the end-matched system to 29.4% in the end-mismatched system. This provided evidence that the presence of terminal mismatches does not play an obvious role in the experimental results.

3.4 Time Studies and Melting Curves

The ligation reactions in this report were quantified at 18 h, but it was supposed that the rates of ligation varied from one combination to another. Substrate-template combinations that were more complementary were presumed to possess higher rate constants. Confirmations of this trend were obtained via time-dependent ligation assays run at 22 ºC using the representative systems shown in Figure 3.5. The data points shown are from two independently run assays with standard deviation values under 10%. The completely matched system, (gg-$_p$gg)/C4, reacted swiftly, with ligation being complete within 1 min. The duplex (tt-ptt)/C4, reacted the slowest, probably as a result of its four mismatches. Complete ligation was obtained in approximately 12 h. The two systems that possessed intermediate observed rate constants both had mismatches at the *n-1* positions. Refer to Table 3.3 and the ligation of (cg-$_p$gc)/C4 and (tg-$_p$gt)/C4, respectively.

The first reaction displayed unsuccessful ligation at 30 ºC and was more than 10-fold slower at 22 ºC than (tg-$_p$gt)/C4, which ligated well at 30 ºC. No product breakdown during the reactions indicated that 18 h appeared to be an acceptable endpoint for these representative systems, which provided another important fact about the reaction.

In addition, melting curves were obtained for the duplexes used in the time studies via thermal denaturation experiments using a standard thermodynamic buffer and shown in Table 3.5. A standard buffer was used because attempts to determine thermal stabilities of the various pre-ligated duplexes under our reaction conditions were unsuccessful due to fact that the melting curves did not show a sigmoidal structural transition. The values reported were the average of two independent assays. As could be expected, the duplexes that had higher rate constants also possessed higher melting temperatures (T_m) than those duplexes that ligated slower.

3.5 Discussion

This report imparts new information about T4 DNA ligase and its ability to ligate oligonucleotides irrespective of the sequence of the terminal ends of the oligonucleotides being joined. Mismatches only at the *n* position resulted in no less than 80% of ligation product formed. Mismatches at the *n-1* position yielded no less than 50% ligation product (following final temperature optimization). Though instances with 2 or more mismatches at the ligation junction generated product formation above 50%, the extent of ligation was inversely proportional to the number of mismatches.

3.5.1 Reaction Conditions

The optimization experiments demonstrated that the fidelity of T4 DNA ligase began to lower in a relatively narrow range of $MgCl_2$ concentration (between 2 and 4 mM, Figure 3.2A). Though the reason for such a narrow range was not known, it was not suspected to be attributed to the inactivity of the ligase at higher $MgCl_2$ concentrations, as intermediate was steadily formed up to 10 mM $MgCl_2$. Moreover, since the formation of intermediate depended on the formation of the duplex (Figure 3.5), this effect, in turn, did not appear to be due to the presence or absence of duplex formation at higher $MgCl_2$ concentration. It could, however, have related to how stable the helix was once formed. It has been previously noted that ligation of mismatches occurred most effectively when the T_m of the pre-ligated duplex was slightly below the reaction

temperature, and that only transient in-line attack was required for ligation (12, 25). Perhaps only the narrow range of divalent cation permitted these steric conditions to occur. If so, then ligating molecules that form less secure duplexes may have required the adjustment of the $MgCl_2$ concentration.

The optimal range of ATP concentration was 10 to 100 μM. Too little of this cofactor proves ineffective as it does not allow for sufficient binding of ATP to the enzyme (91), leading to a lack of enzyme adenylation (Step 1 of Figure 2.7). At high levels of ATP, intermediate formation was established, which indicated enzyme activity and the formation of the DNA duplex. Product formation was low, however, in these cases. Conceivably, the enzyme is consistently adenylated at high concentrations of ATP. Once the adenylated enzyme transfers the AMP to the donor molecule, the enzyme can then separate from the complex. If re-adenylation occurs the before the enzyme rebinds to the adenylated intermediate, it cannot catalyze the final ligation step.

The supplementation of the reaction with 10% to 30% DMSO greatly benefited the ligation of some mismatched oligonucleotides, mainly through an increase in the ligation step of the reaction. Perhaps, then, nucleoside analogues developed to behave as universal bases may show an increase in ligation function when DMSO is incorporated into the reaction. While the origin of the DMSO effect was unknown, it has been reported that DMSO can diminish the specificity of nucleoside-binding restriction enzymes (51-52). The results of this study echoed those reports. This effect is termed star site activity. On the other hand, DMSO might not have acted directly on the protein, since ligation of fully base paired nicks did not occur any faster in the presence or absence of DMSO (data not shown). DMSO may have been interacting with the DNA backbone in a manner that was favorable to the orientation of the mismatched components for the reaction.

The effectiveness of the ligation reactions at 30 ºC for the single and double mismatches and 22 ºC for the quadruple mismatches showed that the drop in temperature permitted more effective hybridization of the oligonucleotides to the template. Dropping the temperature even further proved ineffective, in all probability because the duplex increased in rigidity or the enzyme activity decreased. Higher ligation temperatures also did not help, because the resultant drop in hybridization strength would have been expected to increase the fidelity of the ligation reaction (42, 92). Useable melting curves were obtained using a standard thermodynamic

analysis buffer (Table 3.5). For the pre-ligated representative duplex systems shown in Figure 3.5, those systems that ligated most effectively in general had the highest melting temperatures. Those systems appeared to be more stable (with respect to hybridization) and more effective in terms of reactivity, which suggested that hybridization strength played an important role in reactivity.

3.5.2 Canonical Nucleosides as Universal Template Bases

All of the four canonical nucleosides acted as universal template bases at both of the *n* positions or both of the *n-1* positions. This also occurred when mismatches were present in either the donor or the acceptor (Table 3.4, light gray boxes). These mismatches in the donor molecule when the template was C4 were less effective than other double mismatch combinations. When mismatches occurred at all 4 positions, however, ligation efficiencies decreased notably.

A straightforward association between the predicted thermodynamics of the individual mismatches (92-97) and the effectiveness of ligating oligonucleotides that contain such mismatches (at the ligation junction) did not appear to exist. With one or two mismatches, all types of mismatches worked well under the optimized conditions, regardless of whether the individual mismatches were predicted to be stable or structured. When four mismatches occurred and were expected to be stable, only (gg-$_p$gg)/G4 and (tt-$_p$tt)/G4 ligated well, whereas (aa-$_p$aa)/G4 and (gg-$_p$gg)/T4 did not. When four mismatches occurred and were expected to be unstable, (tt-$_p$tt)/C4 and (cc-$_p$cc)/T4 worked well, whereas (cc-$_p$cc)/C4 did not. Thus, the presence of structure at these positions was not a factor in determining ligase activity, or the different reaction conditions utilized between these reports prohibited a direct comparison of the results.

In general, sequence-independent ligation occurred using any of the canonical nucleosides in the template flanking the *n* position. These template nucleosides may or may not have formed hydrogen bonds with the corresponding nucleosides in the donor or the acceptor molecule. It may be that the lack of any specific and required interactions permitted the sequence-independent ligase activity that was observed. Previous reports show that T4 DNA ligase can ligate oligonucleotides even if there is a single nucleotide gap or a residue that is missing a purine or pyrimidine (19) at the ligation junction. When considered alongside the above data, it appears that the requirements for T4 DNA ligase do not include stable base pairing

at the site of ligation. As suggested previously (25), transient proximity is expected to be sufficient.

As a template nucleobase, thymidine consistently performed as well as or better than the other nucleobases. This was underscored by the fact that the G4 template directed the ligation of tt to $_{p}$tt poorly (5%) as compared with the T4 template directing the ligation of gg to $_{p}$gg (87%). There was no hint that guanosines ligate together more effectively than thymidines. Thymidine must have participated in a distinctive role, but only in the template positions, and not when it was in the donor and acceptor terminal positions. Pyrimidines did not appear to perform better in template positions, since cytidine was the poorest universal template base. The low tendency of thymidine to form rigid structures combined with its relatively small size may have given it the flexibility to permit the corresponding nucleobases in the opposite strand to adopt a transitory conformation for reactivity. Since most mismatches are expected to be destabilizing, however, the number of canonical universal template bases that a given template can include will likely be restricted.

3.6 Comparison with Previous Reports

The general idea throughout the literature is that mismatches, when tolerated, are more permissible at the terminal donor position than the terminal acceptor position (13-14, 20, 22-24, 26). With our experimental conditions, this effect was not evident with *n* or *n-1* mismatches, because mismatches appeared to ligate similarly.

Although enzymatic methods were used here for analyzing low fidelity ligation reactions, evidence shows that chemical ligation methods (37-39) can be utilized to ligate oligonucleotides that contain mismatches (38-42). One advantage to using chemical methods is a faster rate of ligation than that of enzymatic methods. Furthermore, chemical methods can be effective at lower temperatures (down to 0 ºC). This permits more effective oligonucleotide hybridization. Enzymes, on the other hand, are largely inactive at lower temperatures. In the case of a single mismatch at the ligation junction, the efficiency of chemical ligations drops, while the fidelity appears to be relatively high in comparison with the results presented here (38, 40, 42).

3.7 Implications

The results presented in this study should advocate caution when performing experiments that rely on the exclusive ligation of matched pairs using T4 DNA ligase. As shown here, its fidelity can be lowered fairly easily. Low fidelity conditions consisted of relatively low amounts of $MgCl_2$ (3 mM) and ATP (10-100 μM). The addition of 10-30% DMSO (w/v) enhances this effect.

Bear in mind that while low fidelity can be shown *in vitro*, such results are not anticipated *in vivo*. This directly relates to the importance of the genomic processes of recombination, replication, and repair, since a low fidelity ligase would result in high mutagenic rates *in vivo* and a loss in genome integrity.

Canonical nucleosides can serve as universal bases in ligation templates, at least at and adjacent to ligation junctions. This could lead to new procedures that permit the ligation of oligonucleotides with unspecified or degenerate terminal ends. A template molecule with thymidines in succession at the ligation junction could order the ligation of various oligonucleotides. When the sequence of the substrate region flanking the ligation junction is identified, simple PCR reactions can be used to amplify ligated products for their subsequent sequencing. In its most straightforward form, this strategy, along with similar methods, could be useful for sequencing the terminal ends of oligonucleotides.

Table 3.1 Results of *n* position investigation under standard conditions.

	A2	C2	G2	T2
a-$_p$a	**12.8 ± 1.1** (72.8 ± 0.6)	**42.1 ± 2.5** (51.0 ± 0.5)	**12.8 ± 2.8** (31.4 ± 3.0)	**97.3 ± 0.6** (1.6 ± 0.1)
c-$_p$a	**76.9 ± 0.5** (14.2 ± 5.1)	**9.1 ± 2.2** (70.6 ± 2.9)	**81.7 ± 6.8** (6.9 ± 0.1)	**45.9 ± 8.3** (47.9 ± 11.2)
g-$_p$a	**17.5 ± 0.4** (27.2 ± 0.3)	**83.0 ± 5.9** (9.1 ± 2.1)	**6.5 ± 0.9** (42.5 ± 2.3)	**96.6 ± 0.1** (2.3 ± 0.1)
t-$_p$a	**86.8 ± 2.1** (7.7 ± 1.8)	**29.4 ± 4.0** (64.0 ± 0.5)	**84.9 ± 1.4** (11.1 ± 2.3)	**97.4 ± 0.7** (1.5 ± 0.5)
a-$_p$c	**11.4 ± 3.0** (65.7 ±14.8)	**23.9 ± 4.3** (70.3 ± 5.1)	**88.1 ± 0.2** (11.1 ± 0.2)	**92.3 ± 2.9** (5.1 ± 2.1)
c-$_p$c	**72.9 ± 1.4** (23.9 ± 0.7)	**8.1 ± 1.4** (28.7 ± 0.2)	**98.3 ± 0.1** (1.3 ± 0.1)	**11.8 ± 4.6** (65.1 ± 1.1)
g-$_p$c	**27.4 ± 1.1** (18.7 ± 1.0)	**56.1 ± 7.8** (38.0 ± 9.1)	**97.1 ± 0.3** (2.2 ± 0.2)	**6.8 ± 0.4** (73.4 ± 2.4)
t-$_p$c	**95.3 ± 2.0** (2.2 ± 0.6)	**25.5 ± 0.4** (65.7 ± 8.8)	**92.0 ± 2.3** (4.3 ± 2.4)	**38.8 ± 1.8** (48.9 ± 0.1)
a-$_p$g	**9.1 ± 0.1** (63.1 ±10.0)	**86.2 ± 12.8** (6.4 ± 3.1)	**21.1 ± 7.3** (33.7 ± 3.6)	**94.4 ± 0.3** (4.1 ± 0.4)
c-$_p$g	**73.4 ± 1.9** (23.9 ± 1.1)	**47.9 ± 2.6** (43.7 ± 3.1)	**78.8 ± 6.9** (12.8 ± 3.8)	**32.1 ± 1.8** (58.7 ± 1.0)
g-$_p$g	**25.4 ± 0.1** (20.9 ± 0.3)	**94.7 ± 1.2** (1.7 ± 0.1)	**10.5 ± 5.7** (47.2 ± 3.5)	**55.2 ± 3.9** (36.6 ± 6.7)
t-$_p$g	**94.7 ± 2.4** (2.8 ± 0.6)	**93.1 ± 3.9** (4.7 ± 2.0)	**49.1 ± 8.8** (16.4 ± 3.8)	**86.8 ± 4.7** (7.9 ± 2.6)
a-$_p$t	**76.9 ± 7.0** (21.7 ± 6.6)	**45.4 ± 3.3** (46.0 ± 0.3)	**38.2 ± 1.9** (48.5 ± 5.6)	**92.1 ± 0.2** (5.1 ± 0.4)
c-$_p$t	**97.5 ± 0.1** (1.5 ± 0.1)	**7.2 ± 5.3** (32.6 ± 10.8)	**95.0 ± 0.1** (3.5 ± 0.2)	**18.6 ± 1.1** (75.1 ± 0.6)
g-$_p$t	**53.1 ± 2.3** (41.3 ± 2.7)	**80.1 ± 0.8** (12.1 ± 7.5)	**53.3 ± 3.6** (43.1 ± 3.6)	**34.5 ± 0.7** (52.3 ± 4.6)
t-$_p$t	**97.8 ± 1.1** (1.5 ± 0.9)	**15.8 ± 6.1** (69.6 ± 7.8)	**87.7 ± 7.8** (4.9 ± 2.3)	**85.2 ± 2.2** (11.4 ± 1.1)
Average	**58.1** ± **34.8** (25.6 ± 23.5)	**46.7** ± **32.1** (38.4 ± 25.6)	**62.2** ± **34.0** (20.1 ± 17.1)	**61.6** ± **34.3** (31.1 ± 28.6)

The horizontal headings denote the template used, while the vertical labels refer to the newly ligated product. The percent of product formed is shown in bold, with the amount of intermediate shown in parenthesis. Both values are shown with standard deviations. The averages for each template are also shown in the same format. The reactions were run with 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h.

Table 3.2 Results of *n* position investigation under optimized conditions.

	A2	C2	G2	T2
a-$_p$a	**96.2 ± 2.7** (1.4 ± 1.0)	**97.0 ± 0.3** (1.5 ± 0.1)	**95.5 ± 1.2** (1.4 ± 1.0)	**97.8 ± 0.6** (0.6 ± 0.1)
c-$_p$a	**95.9 ± 1.9** (1.4 ± 0.6)	**94.4 ± 0.8** (2.9 ± 0.5)	**91.8 ± 6.1** (1.7 ± 0.1)	**95.9 ± 2.2** (0.6 ± 0.1)
g-$_p$a	**93.4 ± 5.0** (1.5 ± 0.5)	**96.0 ± 1.4** (2.0 ± 0.8)	**96.5 ± 1.8** (1.3 ± 0.8)	**98.5 ± 0.1** (0.3 ± 0.1)
t-$_p$a	**93.5 ± 6.2** (1.3 ± 0.8)	**96.9 ± 6.2** (1.5 ± 0.7)	**97.5 ± 1.6** (0.9 ± 0.7)	**98.5 ± 0.3** (0.3 ± 0.1)
a-$_p$c	**97.3 ± 0.2** (1.3 ± 0.2)	**92.2 ± 6.5** (4.5 ± 3.6)	**94.0 ± 7.1** (2.4 ± 2.8)	**97.8 ± 0.3** (0.8 ± 0.4)
c-$_p$c	**85.7 ± 4.5** (5.7 ± 1.6)	**95.0 ± 2.4** (2.9 ± 1.8)	**97.1 ± 0.6** (1.6 ± 0.4)	**97.7 ± 0.9** (0.6 ± 0.1)
g-$_p$c	**86.1 ± 2.0** (6.3 ± 0.1)	**86.4 ± 1.6** (9.2 ± 4.2)	**92.8 ± 7.6** (3.2 ± 3.1)	**98.4 ± 0.2** (0.5 ± 0.1)
t-$_p$c	**97.2 ± 1.0** (1.7 ± 0.7)	**88.0 ± 1.0** (7.7 ± 2.8)	**97.9 ± 0.9** (1.1 ± 0.6)	**98.5 ± 0.1** (0.4 ± 0.1)
a-$_p$g	**91.7 ± 3.5** (1.8 ± 0.1)	**90.0 ± 3.5** (7.4 ± 7.0)	**93.1 ± 2.2** (1.5 ± 0.9)	**95.4 ± 0.2** (0.5 ± 0.1)
c-$_p$g	**90.8 ± 0.7** (1.9 ± 0.1)	**95.4 ± 0.7** (2.1 ± 0.5)	**95.4 ± 0.7** (1.8 ± 0.1)	**92.2 ± 0.4** (0.8 ± 0.2)
g-$_p$g	**88.3 ± 1.6** (1.9 ± 0.1)	**94.3 ± 3.2** (2.6 ± 0.6)	**92.5 ± 0.8** (2.4 ± 0.1)	**93.4 ± 3.0** (2.8 ± 3.0)
t-$_p$g	**92.9 ± 1.4** (1.9 ± 0.2)	**95.1 ± 4.1** (2.3 ± 0.8)	**94.1 ± 1.3** (2.4 ± 0.3)	**94.7 ± 2.4** (0.6 ± 0.3)
a-$_p$t	**97.7 ± 0.1** (1.3 ± 0.3)	**85.3 ± 0.5** (8.2 ± 1.3)	**96.7 ± 5.3** (1.6 ± 0.6)	**97.6 ± 0.7** (1.1 ± 0.1)
c-$_p$t	**89.6 ± 0.1** (0.8 ± 0.1)	**80.2 ± 0.3** (16.8 ± 3.0)	**98.1 ± 2.6** (0.8 ± 0.1)	**97.6 ± 0.2** (1.1 ± 0.5)
g-$_p$t	**94.6 ± 0.6** (2.8 ± 1.6)	**96.2 ± 0.1** (1.8 ± 0.2)	**97.6 ± 0.3** (1.1 ± 0.1)	**97.8 ± 0.2** (1.3 ± 1.0)
t-$_p$t	**89.5 ± 0.3** (5.8 ± 4.1)	**96.5 ± 5.8** (1.7 ± 0.1)	**97.7 ± 0.3** (1.0 ± 0.2)	**97.8 ± 0.7** (1.4 ± 0.8)
Average	**92.5 ± 3.9** (2.4 ± 1.8)	**92.4 ± 5.0** (4.7 ± 4.2)	**95.5 ± 2.2** (1.6 ± 0.7)	**96.9 ± 2.0** (0.9 ± 0.6)

The horizontal headings denote the template used, while the vertical labels refer to the newly ligated product. The percent of product formed is shown in bold, with the amount of intermediate shown in parenthesis. Both values are shown with standard deviations. The averages for each template are also shown in the same format. The reactions were run with 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 µM ATP, 20% DMSO, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h.

Table 3.3 Results of *n-1* position investigation under optimized conditions.

	A4 x = t	C4 x = g	G4 x = c	T4 x = a
ax-$_p$xa	**83.6 ± 0.2** (10.2 ± 0.2)	**81.9 ± 1.9** (7.9 ± 1.0)	**89.4 ± 5.7** (8.1 ± 5.5)	**84.0 ± 9.1** (3.5 ± 1.4)
cx-$_p$xa	**93.7 ± 1.5** (2.7 ± 1.0)	**76.5 ± 0.9** (13.3 ± 2.3)	**97.9 ± 0.7** (1.1 ± 1.1)	**93.6 ± 0.2** (1.8 ± 0.4)
gx-$_p$xa	**6.2 ± 0.3** ♦ (33.2 ± 1.4)	**67.6 ± 4.7** (21.0 ± 3.6)	**78.9 ± 5.9** (10.8 ± 3.9)	**86.5 ± 4.9** (3.0 ± 0.2)
tx-$_p$xa	**95.9 ± 0.5** (1.2 ± 0.2)	**83.1 ± 0.7** (8.4 ± 0.8)	**93.6 ± 1.0** (2.0 ± 0.7)	**87.7 ± 0.3** (4.3 ± 1.1)
ax-$_p$xc	**94.5 ± 1.6** (2.8 ± 1.4)	**51.9 ± 1.4** (38.5 ± 1.5)	**98.3 ± 0.2** (0.8 ± 0.1)	**97.8 ± 0.6** (1.0 ± 0.3)
cx-$_p$xc	**97.4 ± 1.1** (1.1 ± 0.4)	**34.6 ± 1.7** ♦ (51.3 ± 4.5)	**98.4 ± 0.4** (0.7 ± 0.2)	**98.0 ± 0.2** (0.9 ± 0.2)
gx-$_p$xc	**24.6 ± 9.1** ♦ (41.1 ± 3.6)	**73.9 ± 2.6** (8.2 ± 0.3)	**97.3 ± 0.6** (1.0 ± 0.1)	**81.7 ± 4.3** (11.1 ± 0.6)
tx-$_p$xc	**89.2 ± 7.3** (0.8 ± 0.2)	**42.2 ± 8.2** ♦ (40.7 ± 6.7)	**97.8 ± 0.3** (0.8 ± 0.1)	**86.1 ± 2.0** (7.2 ± 4.9)
ax-$_p$xg	**94.7 ± 1.7** (1.9 ± 0.8)	**96.3 ± 0.5** (2.1 ± 1.0)	**87.0 ± 1.9** (8.6 ± 0.1)	**97.0 ± 1.3** (1.8 ± 0.7)
cx-$_p$xg	**84.4 ± 1.6** (1.9 ± 0.8)	**95.4 ± 1.5** (3.3 ± 2.1)	**94.8 ± 0.9** (2.3 ± 0.6)	**97.0 ± 0.8** (0.9 ± 0.2)
gx-$_p$xg	**85.8 ± 5.0** (6.9 ± 2.6)	**96.4 ± 0.1** (2.2 ± 0.8)	**97.2 ± 0.3** (0.9 ± 0.1)	**96.6 ± 0.1** (1.5 ± 0.1)
tx-$_p$xg	**90.9 ± 5.8** (2.7 ± 1.7)	**95.7 ± 1.6** (3.1 ± 2.0)	**97.5 ± 0.1** (0.8 ± 0.1)	**95.4 ± 0.5** (2.3 ± 0.1)
ax-$_p$xt	**79.7 ± 5.8** (9.0 ± 3.3)	**93.4 ± 0.1** (3.2 ± 0.1)	**91.5 ± 1.2** (4.0 ± 0.8)	**97.7 ± 0.4** (1.1 ± 0.4)
cx-$_p$xt	**96.4 ± 1.6** (1.7 ± 0.8)	**94.9 ± 0.2** (3.0 ± 0.7)	**94.6 ± 0.2** (2.0 ± 0.1)	**92.6 ± 5.1** (0.6 ± 0.1)
gx-$_p$xt	**78.2 ± 1.4** (9.3 ± 2.5)	**92.5 ± 2.1** (2.6 ± 0.5)	**95.0 ± 0.1** (1.8 ± 0.1)	**97.1 ± 0.1** (1.1 ± 0.1)
tx-$_p$xt	**98.0 ± 1.4** (0.5 ± 0.1)	**92.6 ± 0.2** (4.1 ± 0.1)	**94.8 ± 0.3** (1.9 ± 0.1)	**96.8 ± 0.4** (1.3 ± 1.3)
Average	**80.3 ± 26.5** (7.9 ± 11.9)	**79.3 ± 20.4** (13.3 ± 16.0)	**92.8 ± 6.9** (3.0 ± 3.2)	**92.9 ± 5.7** (2.7 ± 2.8)

The horizontal headings denote the template used, while the vertical labels refer to the newly ligated product. The x given in the vertical label is defined in the horizontal heading. The percent of product formed is shown in bold, with the amount of intermediate shown in parenthesis. Both values are shown with standard deviations. The averages for each template are also shown in the same format. The reactions were run with 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 µM ATP, 20% DMSO, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h.

Table 3.4 Results of *n* and *n-1* mismatch investigation under optimized conditions.

	A4	C4	G4	T4
aa-$_p$aa	**2.7 ± 1.7** (56.2 ± 7.5)	**2.8 ± 0.8** (47.7 ± 6.5)	**8.8 ± 1.8** (15.8 ± 4.1)	**92.6 ± 1.4** (3.7 ± 0.8)
cc-$_p$aa	**4.3 ± 0.1** (69.8 ± 9.2)	**4.8 ± 3.8** (69.2 ± 1.6)	**86.6 ± 3.3** (5.5 ± 1.6)	**91.7 ± 0.6** (3.8 ± 0.3)
gg-$_p$aa	**2.4 ± 1.7** (46.5 ± 8.4)	**54.8 ± 4.9** (20.4 ± 1.1)	**3.0 ± 0.8** (15.4 ± 6.0)	**91.7 ± 0.4** (3.6 ± 0.4)
tt-$_p$aa	**84.6 ± 2.6** (5.1 ± 0.3)	**4.2 ± 0.9** (45.9 ± 10.2)	**7.3 ± 1.7** (22.5 ± 7.9)	**88.5 ± 0.2** (5.9 ± 0.7)
aa-$_p$cc	**0.8 ± 0.3** (16.2 ± 6.5)	**0.5 ± 0.1** (17.7 ± 8.7)	**75.8 ± 2.3** (14.3 ± 3.7)	**79.4 ± 1.7** (11.7 ± 1.1)
cc-$_p$cc	**1.1 ± 0.1** (25.3 ± 5.0)	**2.2 ± 0.2** (14.8 ± 3.2)	**88.7 ± 5.6** (2.8 ± 0.3)	**65.7 ± 2.4** (14.6 ± 2.1)
gg-$_p$cc	**1.0 ± 0.1** (21.4 ± 8.4)	**31.2 ± 6.3** (19.7 ± 4.8)	**91.1 ± 1.1** (5.2 ± 0.7)	**22.6 ± 0.4** (62.6 ± 0.4)
tt-$_p$cc	**89.9 ± 0.1** (4.3 ± 1.0)	**3.1 ± 0.2** (24.5 ± 3.4)	**96.1± 0.1** (2.1 ± 0.3)	**87.6 ± 1.5** (6.9 ± 0.8)
aa-$_p$gg	**2.1 ± 0.1** (75.2 ± 15.8)	**89.1 ± 1.5** (6.4 ± 1.6)	**77.1 ± 2.2** (14.8 ± 0.6)	**96.1 ± 1.1** (2.1 ± 0.7)
cc-$_p$gg	**92.2 ± 0.4** (4.1± 0.1)	**91.4 ± 4.5** (5.2 ± 3.3)	**95.5 ± 1.8** (3.5 ± 0.3)	**76.2 ± 0.8** (14.8 ± 0.4)
gg-$_p$gg	**2.9 ± 0.3** (88.3 ± 1.1)	**93.2 ± 0.1** (1.2 ± 0.3)	**47.1 ± 1.1** (42.0 ± 2.1)	**87.4 ± 0.1** (9.7 ± 0.3)
tt-$_p$gg	**95.2 ± 1.3** (3.2 ± 1.0)	**92.4 ± 3.5** (4.5 ± 2.6)	**72.4 ± 5.6** (16.4 ± 1.0)	**39.4 ± 4.4** (55.1 ± 5.0)
aa-$_p$tt	**63.7 ± 5.2** (22.8 ± 6.4)	**8.3 ± 1.8** (24.5 ± 6.5)	**7.9 ± 4.8** (12.7 ± 9.1)	**88.9 ± 2.1** (6.9 ± 2.3)
cc-$_p$tt	**95.6 ± 0.3** (1.4 ± 0.3)	**20.8 ± 1.9** (27.0 ± 4.4)	**89.7 ± 0.8** (46.6 ± 0.1)	**27.5 ± 7.1** (64.2 ± 10.0)
gg-$_p$tt	**83.8 ± 3.1** (8.6 ± 8.3)	**55.2 ± 2.1** (16.3 ± 1.2)	**2.4 ± 0.9** (32.4 ± 1.1)	**7.6 ± 1.1** (85.8 ± 0.8)
tt-$_p$tt	**96.1± 0.4** (1.1 ± 0.3)	**52.4 ± 0.2** (7.0 ± 2.2)	**5.2 ± 0.8** (14.9 ± 0.4)	**29.4 ± 2.8** (62.1 ± 3.7)
Average	**44.9 ± 44.8** (28.1 ± 29.5)	**37.8 ± 37.3** (22.0 ± 18.4)	**53.4 ± 39.8** (14.0 ± 11.1)	**67.0 ± 30.5** (25.8 ± 28.8)

The horizontal headings denote the template used, while the vertical labels refer to the newly ligated product. The percent of product formed is shown in bold, with the amount of intermediate shown in parenthesis. Both values are shown with standard deviations. The averages for each template are also shown in the same format. The reactions were run with 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 µM ATP, 20% DMSO, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 h.

Table 3.5 Melting curve data for duplex systems using C4 template shown in Figure 3.5

Acceptor/donor combination	T_m (°C)
gg-$_p$gg	56.7
tg-$_p$gt	38.5
cg-$_p$gc	26.8
tt-$_p$tt	30

The thermal denaturation experiments were run using a standard thermodynamic buffer composed of 20 mM sodium cacodylate, 1 M NaCl, and 0.5 mM Na_2EDTA at pH 7.0 using a 15 μM total strand concentration. Measurements were taken using a wavelength of 260 nM. Each T_m is the average of two independent denaturation experiments.

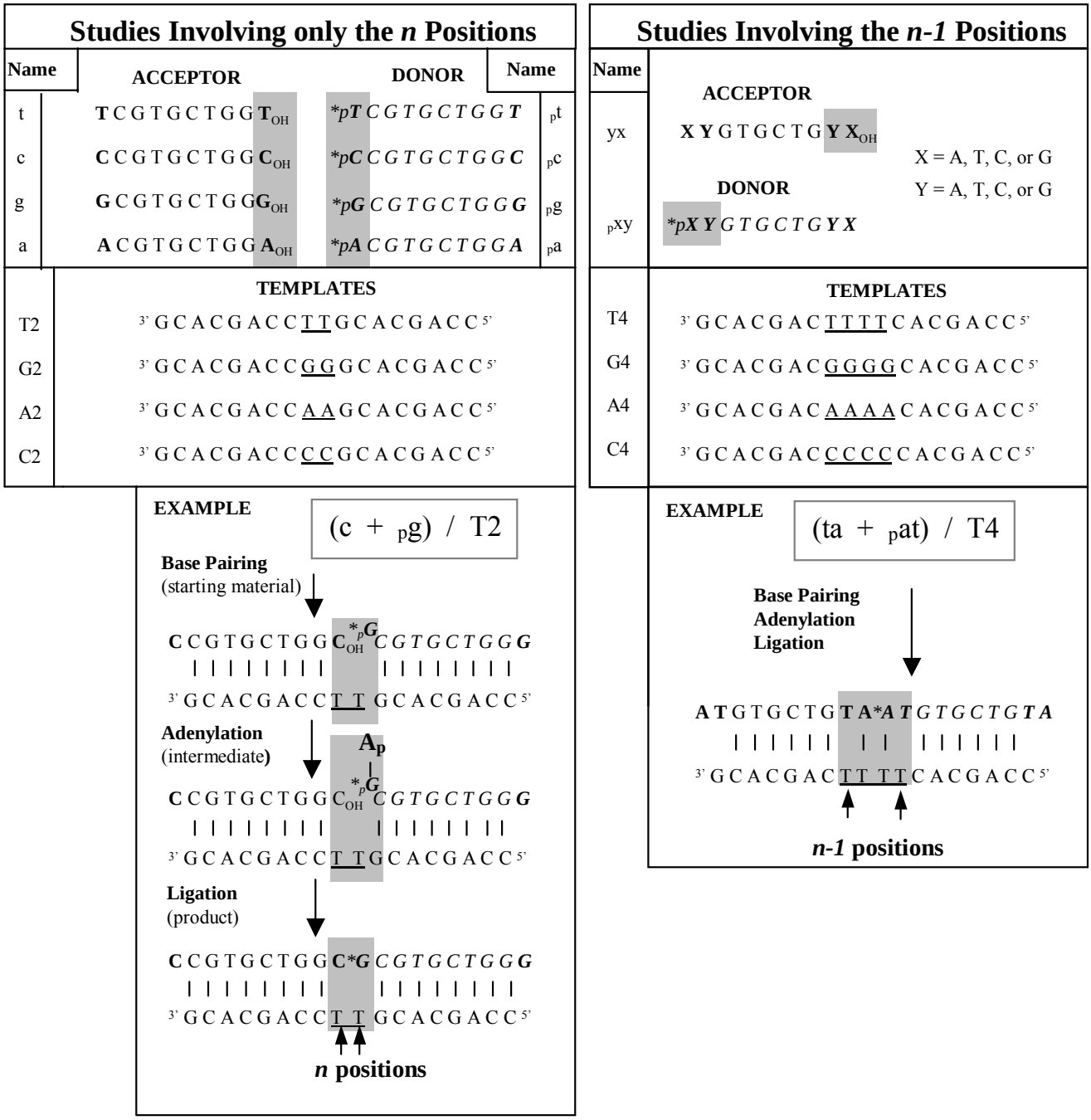

Figure 3.1. Diagram of investigation system for universal template study. On the left side, the *n* position system is shown, while the right side illustrates the *n-1* position system. The acceptor and donor oligonucleotides are decamers, shown in lowercase lettering. Each one is named for the nucleotide with which it begins and ends. The donor decamers are phosphorylated via radiolabeled and shown in italics. The templates are in uppercase and denote the central universal template regions. The steps of the ligation reaction (adenylation and intermediate formation, and ligation) are shown. Once the decamers bind to the template, T4 DNA ligase catalyzes the formation of the covalent bond between the 3' OH and the 5' phosphate of the acceptor and donor, respectively.

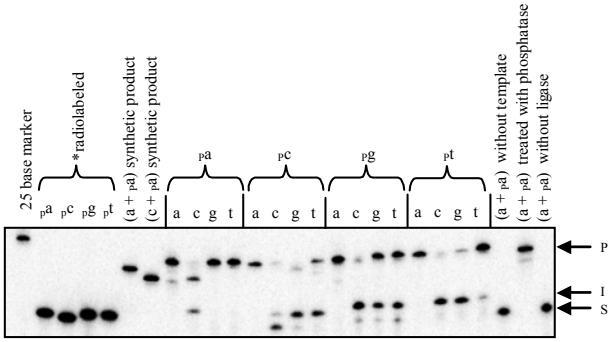

Figure 3.2. Representative gel of ligation reactions run under standard conditions with the T2 template. The standard reaction conditions are 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 hours. The designations for P (product), I (intermediate), and S (starting material) are shown on the right side of the gel. The acceptor/donor combinations are shown above each lane. The first lane contains a 25-base size control, while the next four lanes are the donor oligonucleotides (unreacted); the following two lanes are the synthetic (a + $_p$a) and (c + $_p$a) size controls. The sixteen possible reactions are shown in the next four panels. The next three lanes are control reactions with (a + $_p$a) without template, with template and treated with calf alkaline intestinal phosphatase (CIAP) after the reaction is completed, and without ligase.

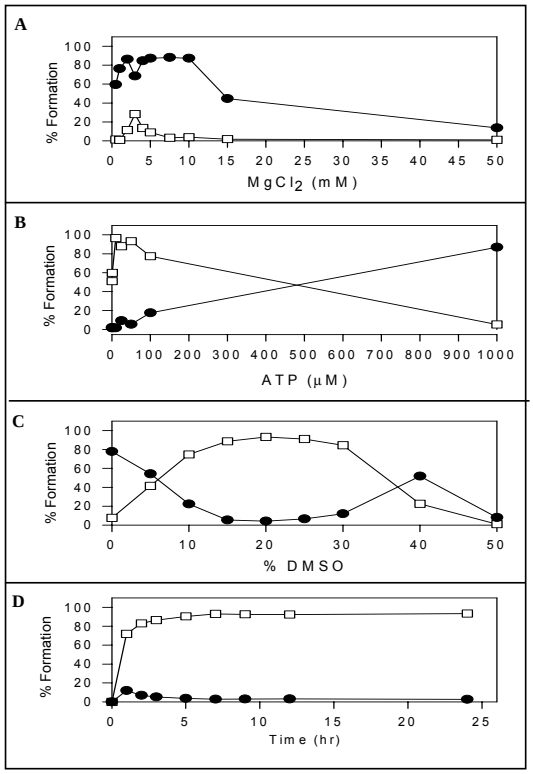

Figure 3.3. Optimization of ligation conditions. Squares represent product and circles denote intermediate. The optimum $MgCl_2$ concentration of 3 mM was determined using (c + $_p$c)/T2 as a test reaction. It was then used in the subsequent ATP-dependent assay shown in panel B. The inclusion of DMSO was tested in panel C using 3 mM $MgCl_2$ and 10 μM ATP with the test reaction (c + $_p$c)/C2. Time-dependence tests were run using (c + $_p$c)/T2 in 3 mM $MgCl_2$, 10 μM ATP, and 20% DMSO. Each data point was the average of two independent assays.

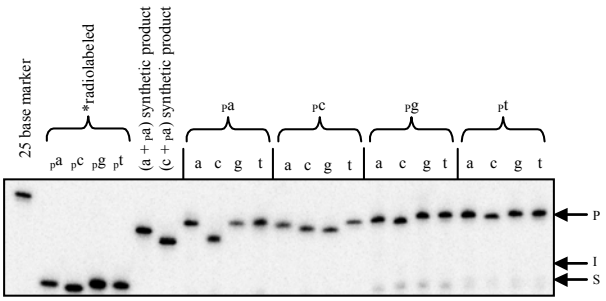

Figure 3.4. Representative gel of ligation reactions under optimized conditions with the T2 template. The standard reaction conditions are 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 μM ATP, 20% DMSO, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase at 30 ºC for 18 hours. The designations for P (product), I (intermediate), and S (starting material) are shown on the right side of the gel. The acceptor/donor combinations are shown above each lane. The first lane contains a 25-base size control, while the next four lanes are the donor oligonucleotides (unreacted); the following two lanes are the synthetic (a + $_p$a) and (c + $_p$a) size controls. The sixteen possible reactions are shown in the next four panels.

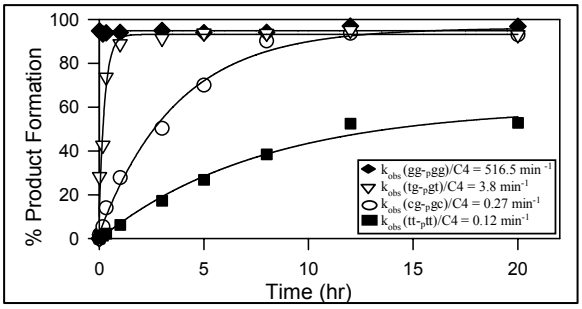

Figure 3.5. Time-dependence of ligation reactions. Each reaction was run under the optimized conditions, which were 30 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 10 μM ATP, 20% DMSO, 2.5 pmol acceptor molecule, 70 fmol donor molecule, 3.1 pmol T2 template, and 3 units T4 DNA ligase, at 22 °C. The ligation systems used were (gg-$_p$gg)/C4, (cg-$_p$gc)/C4, (tg-$_p$gt)/C4, and (tt-$_p$tt)/C4. The symbol used for each system, along with its resultant k_{obs}, is shown at the bottom right of the graph. Each data point was the average of two independently run assays.

Chapter 4 - Removal of Urea via 2-Butanol Extraction

4.1 Materials and Methods for urea extraction via 2-butanol extraction

4.1.1 Extraction and quantification procedure

To generate an 8 M urea solution, approximately 150 mg of urea (Sigma-Aldrich) was dissolved in 2.5 mL water (at pH 7). This approximates the typical amount of urea in an 8 M urea denaturing gel slice and the typical volume of water used to diffuse the nucleic acid out of the gel slice. Therefore, in these experiments the initial aqueous phase contained 1 M urea. To each successive extraction an equal volume of 2-butanol (Fisher Scientific) was added, starting with 2.5 mL. Each extraction consisted of vortex-mixing the 2-butanol/water mixture for at least 30 seconds and then separating the phases by centrifuging for 5 minutes at 3,000 rpm. Since water partitions into the 2-butanol phase during each extraction (50), successive equal volume extractions (pH 7) resulted in successively smaller 2-butanol and aqueous solution volumes, and were discontinued once less than 100 μL of aqueous solution remained (culminating in a total of five extractions). To determine the effect of 2-butanol volume on extraction properties, a large volume extraction series (at pH 7) was conducted. In these experiments, 1.5 times (1.5X) more 2-butanol than aqueous solution was added in each extraction. Because more 2-butanol is added, more water can partition into the 2-butanol phase per extraction. Therefore, the water volume is reduced to less than 100 μL in 3 (rather than 5) extractions. The 2-butanol phase of each extraction was then dried completely in a Juoan RC10.10 centrifugal evaporator. The amount of urea that partitioned into the 2-butanol phase for each extraction was quantified by weighing the urea-containing tube and subtracting the weight of the tube. The sum of the urea extracted and that remaining in the aqueous phase weighed within 10% of the starting amount of urea for each extraction series, indicating the accuracy of the experimental method. Each extraction series reported is the average of three independent experiments.

4.1.2 Extractions with nucleic acid

Extractions on samples with nucleic acid had approximately 1 pmol of 5’ end radiolabeled Oligo1, d(CTGTGCTGAC), added into the initial 2.5 mL aqueous solution. These extractions were conducted using samples with or without urea. Oligo1 was purchased from and HPLC purified by Integrated DNA Technologies (Coralville, IA) and was radiolabeled as

previously described (31). The equal volume extraction strategy outlined above was followed. The amount of radiolabeled Oligo1 remaining in each 2-butanol extraction and in the aqueous phase was quantified by blotting 10 µL (out of 200 µL total) of each dried and resuspended extraction on a Baker Si250F thin layer chromatography (TLC) plate and analyzing the TLC plate radiation with a Molecular Dynamics STORM 860 Phosphorimager.

4.2 System Design

The purpose of this study was the investigation of 2-butanol extraction as an effective method to remove urea from aqueous nucleic acid samples. In the extractions, an 8M urea solution was generated and extracted with equal and large (1.5X) volumes of 2-butanol until a minimal amount of solution (100 µL) remained. The amount of urea removed with each extraction was measured and the results were evaluated with respect to the volume used in the extraction and the concentration of urea in both the aqueous and butanol phases. The extractions were also duplicated with an addition of radiolabeled oligonucleotide to observe whether or not the urea removal left the nucleic acid behind in the aqueous phase. The levels of radioactivity were quantified to establish if urea had an effect on the partitioning of the nucleic acid into either the aqueous or the butanol phase.

4.3 Results

Figure 4.1 depicts the percent of urea that remained in the aqueous solution (at pH 7.0) with respect to the number of 2-butanol extractions. All of the data points shown were the results of three separately run assays with standard deviation values shown as error bars. When an equal volume of butanol was used, five extractions were performed (squares). In cases of large volume extractions, three extractions were done (circles). Regardless of the volume of the extraction used, approximately 3% of the urea originally in solution remained. This reduction of 97% (150 mg to 4.5 mg) was further evidenced by the considerable amount of urea eliminated in each extraction. These results indicated a reproducible and quantitative removal of urea via 2-butanol extraction.

Figure 4.2 shows the molar concentration of urea remaining in each phase after each extraction. Again, equal volume extractions are denoted with squares and large volume with circles and the values shown are from three independent assays. The error bars correlate with the

standard deviation values. The solid symbols correspond to the urea concentration remaining in the aqueous phase, while the open symbols represent the urea concentration in the 2-butanol phase. Urea concentrations in both the butanol phase and aqueous phases (0.29 M ± 0.04 and 0.91 M ± 0.12, respectively) after each of the equal volume extractions were relatively constant. Results in the case of the large volume extractions were in essence the same (0.26 M ± 0.04 and 0.89 M ± 0.12, respectively).

In order to answer the question of how the urea was being carried into the 2-butanol phase, the data was charted as a function of the water that reallocated into the 2-butanol phase (Figure 4.3). This differed from Figure 4.2, which took into consideration both phases; this analysis took into account only the water that is taken into the butanol phase. The designations for equal and large volume extractions mirrored those previously mentioned and the data shown is from three independent experiments, with the standard deviations again shown as error bars. It was apparent from the results that after the first extraction, the values remained relatively constant. The concentration of urea and water in the 2-butanol phase was 0.82 M ± 0.07 for the equal volume extractions and 0.84 M ± 0.07 for the large volume extractions. These results fell within the standard deviations of the urea concentrations in the respective aqueous phases shown in Figure 4.2.

Nucleic acids partitioned poorly into the butanol phase; henceforth, 2-butanol extraction provides an attractive method for the concentration of nucleic acids. The next step in the investigation examined whether or not the presence of urea in aqueous nucleic acid solutions affects the partitioning of nucleic acids between the two phases. Equal volume 2-butanol extractions were performed as above with samples that contained urea and samples that did not; these samples additionally contained a trace amount of radiolabeled Oligo1. The presence or absence of urea did not affect the volume of water redistributed per extraction (data not shown). The amount of radiolabeled nucleic acid that was carried into the 2-butanol phase, as well as what remained in the aqueous phase, was quantified and is presented in terms of the nucleic acid lost per 2-butanol extraction (Table 4.1). The equal volume extractions were performed in duplicate. In the case of the urea-containing solution, the extractions resulted in an average loss of 6.2% per extraction. For the solution without urea, the extractions produced an average loss of 7.7% per extraction.

4.4 Discussion

Extraction with 2-butanol yielded a consistent and significant elimination of urea from aqueous solutions, as shown in Figure 4.1. The results obtained with the equal and large volume extractions provide evidence that the concentration of urea in each phase was not dependent on the volume of 2-butanol added (Figure 4.2). Moreover, the urea concentration in each phase did not depend on the volume of aqueous solution remaining, when comparing the different volumes in each of the extractions. It seemed that a concentration equilibrium was established between the two phases. The aqueous phase possessed roughly a 3-fold higher concentration of urea than the butanol phase. This suggested that the urea is not very soluble in the 2-butanol, if at all.

As for how the urea was removed from the solution, it seemed that most, if not all, of the urea carried into the 2-butanol phase by the redistributed water. Apparently, the urea did not dissolve in the 2-butanol directly to any great extent. The equilibrium that prevailed between the phases, as well as the amount of urea extracted, were functions of the amount of water that reallocated into the 2-butanol phase. Thus, the large volume extractions removed the same amount of urea in fewer extractions, as more water partitioned into the 2-butanol phase.

In every extraction series performed, merely 10% to 20% of the total aqueous volume redistributed from the aqueous phase into the 2-butanol phase in the first extraction (data not shown). On the contrary, each subsequent extraction removed approximately 60% of the remaining aqueous volume. This reflected the differences in urea concentrations shown in Figure 4.3. This implied a difference between the first extraction and those following it. Yet, the concentration of urea in the 2-butanol phase was the same for every extraction, including the first extraction (Figure 4.3). One probable explanation for such results was that the 2-butanol saturates the aqueous phase. This led to additional volume in the aqueous phase, which established an equilibrium in the first extraction. This intimated that more water partitioned into the 2-butanol phase in the first extraction than what could accurately be determined. Perhaps, some of it was being replaced in the aqueous phase by 2-butanol. This might have explained the higher urea concentration in the first extractions that is shown in Figure 4.3.

The extraction data obtained using a solution containing nucleic acids illustrated that the presence of urea in the aqueous solution did not substantially affect the partitioning of the nucleic acid between the two phases. As evidenced from the data, the average amount of nucleic

acid lost per extraction differed only slightly depending upon whether or not urea was present (6.2% versus 7.7%).

4.5 Implications

These results demonstrated that the extraction of aqueous nucleic acid solutions utilizing 2-butanol removes urea. In addition, the presence of urea when concentrating such solutions by 2-butanol extraction did not affect the partitioning of the nucleic acids or increase the amount that may be lost per extraction. This study offered a simple method for removing urea from nucleic acid samples, including those purified from denaturing polyacrylamide gels. Water could be added so that additional 2-butanol extractions could be done if further urea extraction is desired. On the other hand, butanol extraction removed enough urea such that other forms of purification (i.e. ethanol precipitation or column chromatography) could be more successful. In terms of the volume of 2-butanol to add for each extraction, both the equal and large volume extractions yielded a $\geq$ 95% elimination of urea in aqueous solutions. Large volume extractions, however, have the advantage of decreasing the total number of extractions required, which will probably reduce the amount of nucleic acid lost to the 2-butanol phase. Unfortunately, using an even larger volume of 2-butanol increases the likelihood that all of the water, nucleic acid, and urea will partition into the 2-butanol phase.

To conclude, this study has proven that the extraction of urea from nucleic acids via 2-butanol is an effective method in cases of high urea concentrations. In these situations, other methods routinely fail. In addition, it is shown here that the urea does not dissolve into the 2-butanol phase directly, but is “taken into” the 2-butanol phase with the water that is redistributed into it. This simple, yet efficient new strategy for purification is useful in the enhancement nucleic acid purity for use in a myriad of molecular biology applications.

Table 4.1 Loss of nucleic acid per equal volume 2-butanol extraction with and without urea

Extraction #	With urea	Without urea
1	5.6%	5.7%
2	12.5%	8.7%
3	2.1%	0.3%
4	2.8%	13.5%
5	8.2%	10.3%
Average loss per extraction	6.2%	7.7%

Each extraction set was performed with 1 pmol of radiolabeled Oligo1 in a 2.5 mL aqueous solution. In those extractions with urea, the concentration of urea was approximately 8 M. The amount of nucleic acid remaining was quantified by blotting 10 μL of a 200 mL solution onto a TLC plate and comparing the radioactive intensity. All of the percentages were obtained by averaging two separate assays.

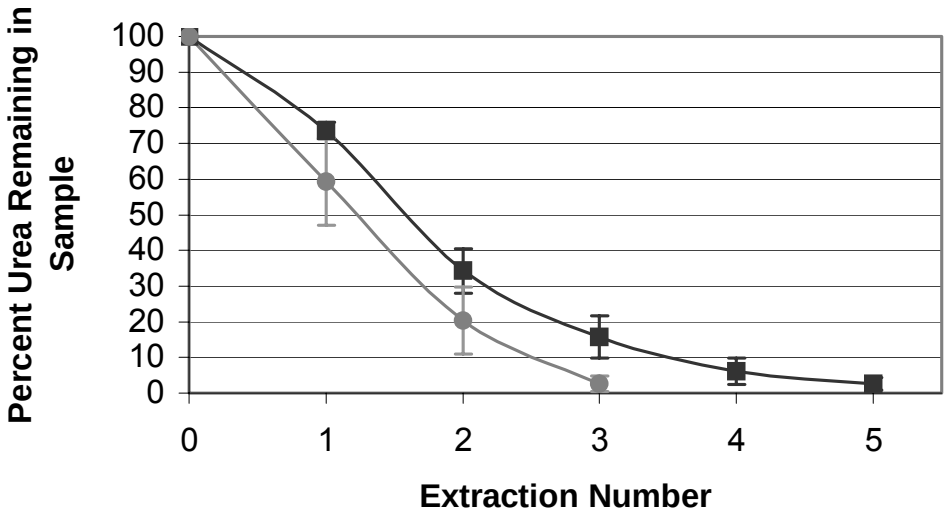

Figure 4.1. The percent urea remaining in the original aqueous solution (at pH 7.0) versus the number of 2-butanol extractions. Equal volume extractions are shown with squares, while the large volume extractions are denoted with circles. Each data point was obtained by triplicate extractions with the standard deviations shown as error bars.

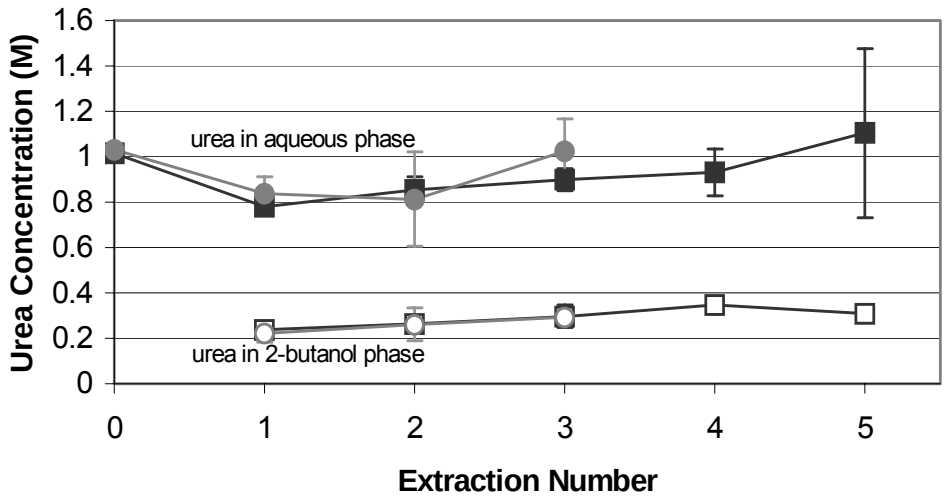

Figure 4.2. The concentration of urea (in M) remaining in each phase after each extraction. Equal volume extractions are shown with squares, while the large volume extractions are denoted with circles. The concentration of urea shown in each data point takes into account the remaining volume of each phase. Each data point was obtained by triplicate extractions with the standard deviations shown as error bars.

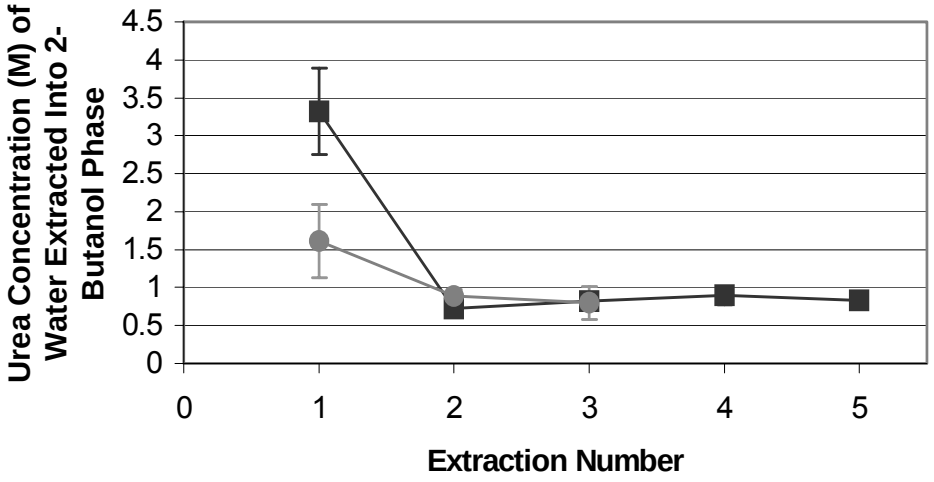

Figure 4.3. The concentration of urea (in M) per volume of water (in mL) that redistributes into the 2-butanol phase for each extraction. Equal volume extractions are shown with squares, while the large volume extractions are denoted with circles. This graph takes into account the volume of water taken into the butanol phase for each extraction. Each data point was obtained by triplicate extractions with the standard deviations shown as error bars.

Chapter 5 - RNA Replacement via Ribozyme-Catalyzed 5' Transcript Replacement

5.1 Materials and Methods for the 5' Replacement Splicing Study

5.1.1 Oligonucleotide Synthesis and Preparation

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). RNA oligonucleotides were purchased from Dharmacon, Inc. (Boulder, CO) and subsequently deprotected with the manufacturer's protocol. RNA substrates were radiolabeled at the 3' termini using modifications to a two-step procedure previously described (98). First, cytidine monophosphate (C_p) was radiolabeled at the 5' termini by incubating 10 μM C_p, 0.85 μM [γ-^{32}P] rATP (Amersham Pharmacia, Piscataway, NJ), 70 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.1 mM rATP, 10% glycerol and 10 units T4 Polynucleotide kinase (New England Biolabs, Beverly, MA) in a reaction volume of 10 μL for 1.5 h at 37 °C. The kinase was heat-inactivated at 65°C for 15 minutes. Next, 2 μM of the RNA substrate was ligated to the prepared *pCp in a reaction with 0.28 μM *pCp, 50 mM Tris-HCl (pH 7.4), 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM rATP, 0.1mM EDTA, 10% dimethyl sulfoxide (DMSO), and 20 units of T4 RNA ligase (New England Biolabs) at 4 °C for ≥ 16 h. The products were gel purified as described (31).

5.1.2 Plasmid Construction and Synthesis

The *P. carinii* ribozyme plasmid precursor, P-8/4x, was generated as previously described (31). The plasmid containing the initial 5' replacement splicing ribozyme was derived from the P-8/4x plasmid and designated PKA. A number of alterations were required to generate this plasmid (Figure 5.1). First, an Age I site was engineered into the P2 helix region of the plasmid containing the ribozyme (Step 1 of Figure 5.1). This new and unique restriction site was used in conjunction with an existing Sfi I site upstream of the T7 promoter to create an insertion site for a double-stranded oligonucleotide fragment containing a T7 promoter, the new 5' exon, 6 bases of the L1 loop native to the *P. carinii* intron, and the RE1 and RE3 recognition elements (Step 2 of Figure 5.1). This insert replaces the T7 promter and the IGS (RE1 and RE3) of the P-8/4x of the ribozyme (31). Last, the RE2 region was modified specifically for the targeted portion of the k-ras gene (Step 3 of Figure 5.1).

A full description of the methods used follows. A unique Age I site was created in the starting plasmid (P-8/4x) via site-directed mutagenesis using primers 5' CATGAAACCGGTGTGAAAACGTTAGGTAGTG 3' and 5' CACTACCTAACGTTTTCACACCGGTTTCATG 3'. The amplification reaction contained 15 pmol of each primer, 25 ng of parental plasmid, 2.5 units of Pfu DNA Polymerase from Stratagene (La Jolla, CA) and 0.5 μM dNTPs in a reaction buffer of 10 mM KCl, 10 mM $(NH_4)_2S0_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1 % Triton X-100, and 0.1 mg/mL BSA; the total reaction volume was 50 μL. After initial denaturation for 30 seconds at 95 ºC, the mixture underwent 15 cycles of 95 ºC for 30 seconds, 55 ºC for 2 minutes, and 68 ºC for 6 minutes. The parental plasmid was digested with 20 units of DpnI from New England Biolabs in 4.1 μL of the manufacturer's buffer for 1 hour at 37 ºC. Transformation of *E. coli* DH5α competent cells from Invitrogen (Rockville, MD) was performed with 3 μL of this reaction according to the manufacturer's specifications. Purification of the vector utilized the QIAPrep Miniprep Kit (QIAGEN, Valencia, CA). The resulting plasmid was digested with 25 units of Age I (New England Biolabs) in 7.5 μL of the manufacturer's buffer at 37 ºC to fully cleave 25 μg of the plasmid; the newly linearized plasmid was purified with the QIAquick PCR Purification Kit (QIAGEN). The linear plasmid was digested a second time with 40 units of SfiI (New England Biolabs) in 20 μL of the manufacturer's buffer and 2 μg of BSA at 50 ºC. The plasmid was run on a 1.5% agarose gel. The band containing the plasmid was excised and purified using the QIAquick Gel Extraction Kit (QIAGEN). To prevent recircularization without the insert, the 15 μg of plasmid was dephosphorylated by incubation with 4 units of calf intestinal alkaline phosphatase (CIAP) in 6 μL of the manufacturer's buffer at 37 ºC for 1 hour. Purification of the plasmid was done using the QIAquick PCR Purification Kit (QIAGEN).

Next, 300 pmols of each oligonucleotide of the insert were phosphorylated by incubation with 1 mM rATP and 10 units of T4 Polynucleotide kinase (New England Biolabs) in 5 μL of the supplied buffer at 37 ºC for 1 hour; the total reaction volume was 50 μL. The reactions were purified using the QIAquick Nucleotide Removal Kit (QIAGEN). Approximately 9 pmols of each strand of the deoxyoligonucleotide fragment were combined and preannealed at 80 ºC for 5 minutes and added to 10 μg of the double-digested plasmid with 21 units of T4 DNA ligase (Promega, Madison, WI) with 2 μL of the supplied buffer. Ligation proceeded overnight at 22

ºC. Transformation of *E. coli* DH5α cells with 5 μL of the ligation reaction using the manufacturer's specifications resulted in a vector purified with the QIAPrep Miniprep Kit (QIAGEN). Two more successive rounds of site-directed mutagenesis were performed with the following sets of primers to alter the RE2 region: 1) 5' GCTTAAGGTATAGTCTGTCAAAAACCTCTTTCG3' and 5' CGAAAGAGGTTTTTGACAGACTATACCTTAAGC3'; 2) 5'GGTATAGTCTGTCAAGGCCCTCTTTCGAAAG3' and 5' CTTTCGAAAGAGGGCCTTGACAGACTATACC 3'. This new plasmid was designated PKA-6/8/6 to reflect the base plasmid (PKA) and the length of the L1 loop/RE2 region/RE3 region. The plasmid was sequenced by Davis Sequencing (Davis, CA) to confirm the changes. Any further changes to test the length dependence of RE2, RE3, or the L1 loop were performed via the site-directed mutagenesis technique described previously with the primers specific for the changes desired.

5.1.3 Transcription

The plasmid was linearized using Xba I (Invitrogen) and purified using the QIAquick PCR Purification Kit (QIAGEN) prior to runoff transcription. The transcription reaction was run as previously described (31) and purified with the QIAGEN-tip 100 anion-exchange columns (QIAGEN) using the following protocol: the column was first equilibrated with 4.0 mL Buffer I (750 mM NaCl, 50 mM MOPS (pH 7.0), 15% ethanol, and 0.15% Triton X-100). The transcription reactions were subsequently loaded on the column and the column washed with 7.0 mL of Buffer I. Finally the transcription products were eluted with 4.0 mL of Buffer II (1.0 M NaCl, 50 mM MOPS (pH 7.0), and 15% ethanol). The eluted products were isopropanol precipitated, ethanol precipitated, and reconstituted in sterile water and quantified with a Beckman UV-VIS DU-650 spectrophotometer.

5.1.4 5' Replacement Splicing Reactions

Reactions were run in HxMg buffer composed of 50 mM HEPES (25 mM Na^+), 135 mM KCl, and x mM $MgCl_2$ at pH 7.5, where x is from 0 to 200. Prior to each reaction, the ribozyme was added to a solution of sterile water and the appropriate buffer at a 100 nM concentration (0-400 nM ribozyme in concentration-dependent assays) and preannealed at 60 ºC for 5 minutes.

The ribozyme solution was allowed to cool to the reaction temperature listed in the figures. In the studies analyzing the effect of guanosine monophosphate (GMP) on the 5' replacement splicing reactions, GMP concentration ranged from 0-300 μM and the specific amount of GMP being tested was added to the ribozyme solution prior to preannealing. Temperatures tested during optimization ranged from 22-60 ºC. The reaction was initiated by the addition of a 1 μL aliquot of a solution containing radiolabeled substrate in the appropriate buffer at a 8 nM concentration to give a final volume of 6 μL. The reaction time was tested from 1-24 h. The reactions were terminated by adding an equal volume of stop buffer (10 M urea, 3 mM EDTA, and 0.1X TBE), followed by denaturation at 90 ºC for 1 minute. Products and reactants were separated on a 14% polyacrylamide/8M urea gel. The gel was transferred to chromatography paper (Whatman 3MM CHR) and dried under vacuum. The bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. The results shown are the average of 2 independent assays. In all cases, the standard deviations were in the range of ±10% of the average of the 2 assays. Observed rate constants were quantified by fitting the data from the time dependence studies to a simple exponential function.

5.1.5 Sequencing Protocol for 5' replacement splicing reaction product

The 5' replacement splicing reaction was run using the 12:1 ribozyme to substrate ratio described above (optimized with respect to the ribozyme concentration), but with non-radiolabeled substrate and 5 μL of the reaction was added to 5 pmol of a template oligonucleotide with the sequence 5' GCACAGGGACCTGTATGTCAAGGCACTCGTCCCTTCCG 3', 3.3 pmol of linker 2A (5' TGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC 3'), 3.3 pmol of linker 1B (5' GTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA 3') and 6 units of T4 DNA ligase (Promega) under the manufacturer's conditions for 18 h at 4 ºC. The ligation reaction was purified using QIAquick Nucleotide Removal Kit (QIAGEN) to isolate the 5' replacement splicing product with both linker oligonucleotides attached. Next, 5 μL of the purified ligation product was used in an RT-PCR reaction utilizing the Access RT-PCR kit from Promega under the manufacturer's conditions. The RT-PCR primers used were P1 (5' TGCCAGGTCACTCAAGTCGGTCATTCATGTCA 3') and P2 (5' TGCTCAGGCTCAAGGCTCGTCTAATCACAGT 3'). The amplified product was purified via

the QIAquick Gel Extraction kit (QIAGEN) and ligated into the pDrive Cloning Vector from QIAGEN. The ligation was run under the manufacturer's conditions for 1.5 h at 4 ºC. Last, *E. coli* DH5α cells were transformed with 5 μL of this ligation reaction using the manufacturer's specifications; the resultant vector was purified with the QIAPrep Miniprep Kit (QIAGEN). The plasmid was sequenced by Davis Sequencing to provide confirmation of the 5' replacement splicing product.

5.2 Principle of Ribozyme-Catalyzed 5' Transcript Replacement

5.2.1 Proposed Mechanism

The purpose of this study was the engineering of a ribozyme that could replace the 5' end of a mutant RNA with a corrected version in a small model system. Note that this differs from 3' replacement splicing, which uses 5' recognition elements as primary determinants of substrate targeting. This ribozyme utilizes three recognition elements in the reaction: RE1, RE2 and RE3. RE1 base pairs with the new 5' exon (covalently attached to the ribozyme) to form the P1 helix. RE2 base pairs with the 5' end of the substrate to form the P9.0 helix (Figure 2.15), while RE3 forms the P10 helix with the 3' end of the substrate. Connecting the new 5' exon and the remainder of the ribozyme is a hairpin loop (labeled L1) that is native to the *P. carinii* intron (31).

5.2.2 Design of the 5' Replacement Splicing System

The system design is shown in Figure 5.2. In both panels A and B, the 5' replacement splicing ribozyme is in blue uppercase lettering, the 16-mer substrate is in green lowercase lettering. In panel A, the exon regions to be ligated are in bold. The ribozyme recognition element RE1 base pairs with the 5' exon-ribozyme conjugate to form the P1 helix (an intramolecular interaction). The ribozyme recognition elements RE2 and RE3 base pair with the substrate to form the P9.0 and P10 helices, respectively. The P10 helix is boxed. The first design principle is used for this system is shown in Step 1 of Figure 5.2; it consists of the conserved base-pair formed by a guanosine (gray) in the ribozyme and a uridine (blue) in the substrate, or the 5' replacement exon in this case. This interaction defines the first catalytic site and is denoted with an arrow. The second principle is the required ωG, in the substrate (red) that defines the second catalytic site; this site is also denoted with an arrow. These first two design principles

involve the previously mentioned conserved and required elements of group I intron-derived ribozymes.

The sequence of the gene k-ras was used as a target because of the location and nature of the previously mentioned mutation hotspot, codon 12 (which is heavily linked to lung cancer). The small model substrate is a short synthetic portion of the k-ras gene that is 15 nucleotides in length and contains codons 18-22 (Figure 5.2). Codon 20 best suited the design principles necessary for a splice site and it is located downstream of codon 12. Therefore, the 5' replacement splicing reaction will replace everything upstream of (or 5' to) codon 20 (including codon 12, which is not included in this small molecule model system). The new 5' exon is attached to the ribozyme via a normal phosphodiester bond (to produce a 5' exon-ribozyme conjugate) and contains the last base in codon 16 through the last base in codon 20 of the k-ras gene. The last base of codon 20 in the new 5' exon is a uracil, which replaces the guanosine at this position in the original transcript. This will not change the amino acid sequence of the resultant protein product (Figure 5.2B). This concerns the third principle of design, which is the targeting of guanosines within a sequence that could yield silent mutations when the guanosine is replaced with uridine in the final product. Panel B of Figure 5.2 displays this with the amino acid sequences of both substrate and product.

The 5' replacement splicing reaction in this small model system was expected to result in a product 19 bases in length. The 3' radiolabeling procedure performed on the substrate allowed visualization of the reaction; it also resulted in an additional cytosine on the 3' end of the substrate (16 bases) and product (20 bases).

5.3 Results

5.3.1 Initial System and Optimization

The initial test ribozyme contained the new 5' exon, 6 bases of the L1 loop found in the *P. carinii* intron (31), a 6-base RE1, an 8-base RE2, and a 6-base RE3 (shown in Figure 5.2). The lengths of the recognition elements were derived from previous work on the TES reaction that indicated longer P9.0 regions to be beneficial, as well as P10 regions up to 6 bases in length (99).

The reaction resulted in the first-reported ribozyme-catalyzed replacement of the 5' portion of an RNA substrate in an optimized yield of 33.1% ± 2% (the average of four independently run assays) (representative polyacrylamide gel in Figure 5.3A). Reactions were

run using 100 nM rPKA-6/8/6 ribozyme, 8 nM radiolabeled substrate, and 50 mM $MgCl_2$ at 44 °C for 20 h. The first lane contains the synthetic reaction product; the second lane contains the 5' replacement splicing reaction with the 16-mer substrate yielding the 20-mer product. The third lane shows an alkaline digest of the synthetic product, and the fourth lane shows the 16-mer starting material under reaction conditions with no ribozyme. Shown to the right of the gel are the sequences and the sizes of the substrate and product. Note the radiolabeled cytosine on the 3' end of both the substrate and the product. In these studies, the $MgCl_2$ concentration, reaction temperature, and reaction time were optimized (Figure 5.3B), as well as the effect of the concentration of the 5' exon-ribozyme conjugate. The $MgCl_2$ dependence was tested from 0 to 200 mM, with 50 mM corresponding to the highest product yield. The reaction was analyzed at temperatures ranging from 0 to 60 °C; 44 °C yielded the highest amount of product. Time dependence reactions ranged from 0 to 24 h; maximum product yield was obtained at 20 h. All data points shown result from two separate assays and have standard deviation values under 10%. Under optimized conditions, a k_{obs} value of 1.58×10^{-3} min^{-1} was obtained. The concentration dependence of the 5' exon-ribozyme conjugate was tested from 0 to 400 nM; maximum product formation could be obtained with a concentration of 100nM. Each factor was investigated at 50 mM $MgCl_2$, 44 ºC, and 20 h using a 12-fold excess (100 nM) of ribozyme over radiolabeled substrate (8 nM), changing only the variable being tested, as shown in Figure 5.3B.

5.3.2 Length Dependence of P9.0

The length of the P9.0 region was investigated as it relates to the product yield of the reaction. The region was tested in increments of 2, resulting in P9.0 helices of 2, 4, 6 and 8 base pairs (Figure 5.4A). The reactions were run with 100 nM ribozyme, 8 nM substrate, and 50 mM $MgCl_2$ at 44°C for 20 h. All data points have been done in duplicate; the standard deviation for all points is less than 10%. Both RE3 and the L1 loop were each at 6-base lengths. As the length of the P9.0 region increased, so did the percentage of product formed. From these tests, it was determined that a longer length for the P9.0 helix was beneficial. The length of the small model substrate did not allow for the testing of a P9.0 longer than 8 bases; however, longer lengths could prove even more beneficial.

5.3.3 Length dependence of P10 and the native loop

The length of the P10 region was also tested in terms of its effect on product yield. Due to the position of P10 and the adjacent L1 loop, however, the analysis had to consider the effects of the length of both regions simultaneously. Decreasing the number of bases in the RE3 region would not necessarily result in a shortened P10 helix, as bases in the L1 loop could participate in P10 formation (see Figure 5.2A). For example, a 4-base RE3 ($^{5'}$GUAU$^{3'}$) would still result in a 6 base RE3 and a 6-base pair P10 helix, with a U•G base-pair at the 5' end of the P10 helix. The additional 2 bases would be from the 3' terminal uridines in the L1 loop, resulting in a 4-base loop ($^{5'}$CACC$^{3'}$). Thus, the effect of both factors was investigated simultaneously, using a 6-base pair and a 1-base pair P10 (Figure 5.4, B1 and B2). Each of the reactions were run under the same conditions the P9.0 length dependence used; these data points are also the average of two independently run assays with standard deviations under 10%. In each case, the P9.0 length was 8 base pairs. With each P10 region, analysis was done on loop lengths from 2-14, advancing in increments of 2. The longest length is shown to the right of each graph. Uridines were used when lengthening the loop because they are the least likely base to form intermolecular and intramolecular structures. As can be seen from Figure 5.4B1 and 2, a longer P10 alone was beneficial in terms of product yield. When the P10 region consisted of only 1 base pair, reaction percentages did not reach over 10%, regardless of loop length. At various loop lengths with a 1-base pair P10, it appeared that any effect loop length could have on product formation was likely undetectable due to the detrimental effect of decreasing the length of P10. Once P10 was increased to 6 base pairs, the product percentages also increased as loop length increases. Note, however, that this was not a precise trend, as a 10-base loop paired with a 6-base RE3 provided the optimal level of product formation for the 5' replacement splicing reaction.

5.3.4 Optimized Construct Investigation

From this data, an 'optimized' ribozyme construct (with a 10-base L1 loop, an 8-base RE2, and a 6-base RE3) was developed. It was scrutinized to find out if it reacted differently to changes in $MgCl_2$, temperature, time and 5' exon-ribozyme conjugate concentration (Figure 5.5) in comparison to the initial ribozyme construct. A representative gel of the reaction is shown in Figure 5.5A. The reactions used 100 nM ribozyme, 8 nM radiolabeled substrate, and 50 mM $MgCl_2$ at 44 °C for 20 h. The first lane contains an alkaline digest of the synthetic product; the

second lane contains the 5' replacement splicing reaction with the 16-mer substrate yielding the 20-mer product. The third lane shows the synthetic reaction product; and the fourth lane shows the 16-mer starting material only under reaction conditions without ribozyme. To the right of the gel are the sequences and sizes of the substrate and product. Note again the radiolabeled cytosine on the 3' end of both the substrate and the product. The optimized reaction yielded 73.6% ± 0.1 product (average of four independently run assays), which is twice as much as seen with the initial ribozyme construct. Magnesium chloride, temperature, and 5' exon-ribozyme conjugate concentration dependence studies on this system yielded the same values for optimum product formation (50 mM, 44 °C, 20 h). The reactions were run as described above except for the changing variable; all data points have been done in duplicate with standard deviations below 10%.

The results from the time dependence study of the 5' exon-ribozyme conjugate did, however, yield different results from the initial system. From this time study, a k_{obs} value of 9.17 x 10^{-3} min^{-1} was obtained. This rate constant was almost 6-fold greater than that of the first construct. With the exception of the rate constant increasing, the remaining reaction parameters did not change. This implied that they were inherent traits of the catalytic core of this ribozyme construct as opposed to related to the recognition elements or loop region.

5.3.5 Sequencing Confirmation

To sequence the 5' replacement splicing product, a sequencing protocol used with small RNAs was modified (100-101). With this protocol (shown in Figure 5.6), linker oligonucleotides are ligated to the product to be sequenced. The DNA linker molecules and the DNA template are shown in black, while the 5' replacement splicing product is shown as an open line. The template was complementary to the 5' replacement splicing product, as well as the 3' end of the 2A and the 5' end of the 1B linker. The template did not, however, contain the complement to the new uridine that is present in the product. As a result, there was a 1-nucleotide bulge when all three molecules were bound. This ensured that if the new uridine was present in the sequenced 5' replacement splicing product, it could only have been obtained from the correct reaction product, as opposed to non-specific interaction with the template molecule. The linkers and reaction product bound to the template and were ligated together. These linker oligonucleotides added length to the product that enabled RT-PCR amplification of it using RT-PCR primers (shown as

solid lines) that corresponded only to the portions of the linkers that did not interact with the DNA template. Because there was no complementary base for the new uridine found on the 5' replacement splicing product, any amplified RT-PCR product containing the new uridine could only have come from the reaction product and not from non-specific interactions with the template. The amplified ligated product was purified and subsequently underwent ligation into a plasmid. Sequencing of the resultant plasmid revealed the expected 5' replacement splicing product sequence, and confirmed that the reaction occurred as expected.

5.3.6 Effect of GMP on the 5' replacement splicing reaction

After optimization of the new ribozyme construct, the next step was the investigation into the effect of GMP on the reaction. This reaction does bear similarities to ribozyme 3' replacement, which requires a guanosine cofactor (Figure 2.13). In addition, the G-addition reaction could occur in the first reaction step instead of the hydrolysis reaction (in a cell, for example). Therefore, it was important to ascertain how the addition of such a cofactor would affect this reaction. These reactions were run just as the optimization reactions; the only difference is the addition of GMP to the reaction. As seen in Figure 5.7, the addition of various concentrations of GMP clearly lowers the product levels in the analyzed reaction. Even at a concentration of 1 μM, the amount of 5' replacement splicing product decreases. Each of the data points were from duplicate assays with standard deviations under 10%.

5.4 Discussion

This study describes a novel ribozyme-catalyzed 5' transcript replacement reaction. By utilizing the 3' recognition elements (P9.0 and P10) of a group I intron as the primary recognition element (Figure 2.15) between the ribozyme and the transcript substrate, this first-reported 5' replacement reaction did, in the small model system, occur. As mentioned previously, there are various types of reactions that are catalyzed by group I intron-derived ribozymes. Each of these has potential in as treatment options for genetic mutations; however, this is the first reaction catalyzed by a ribozyme that has been specifically engineered for 5' RNA transcript replacement. In terms of RNA replacement or repair, this novel reaction offers a promising catalytic option.

5.4.1 Length Dependence of P9.0

From these results, a longer P9.0 helix provided more product formation than a shorter P9.0 helix. The RE2 region binds to the 5' exon that will be replaced, forming the P9.0 helix (Figure 5.2A). This interaction targets the 5' end of the exogenous substrate and moves it into its necessary orientation for the 5' replacement splicing reaction. A stronger P9.0 helix (longer RE2 length) means more stable targeting of the substrate and a decrease in the possibility of substrate and intermediate dissociation, both of which advance the potential for reaction completion. The beneficial effect of a longer P9.0 helix in this system echoed previous results shown in a *P. carinii*-derived ribozyme catalyzing an excision-ligation reaction (99).

5.4.2 Length Dependence of P10

This investigation also showed that a longer P10 helix also enhanced product yield. It has been previously reported that 3' exon binding for a ribozyme derived from the *P. carinii* intron is weaker than that of the 5' exon for the same system (31). This would correspond with a lengthier RE3 region (longer P10 helix, thus stronger interaction) being more helpful in binding and holding the 3' exon region for exon ligation and product formation. An extended RE3 region maintains the P10 helix interaction that keeps the 3' exon in place for the second step of the reaction. This lengthened interaction could augment the rate of the reaction by having the 3' exon component 'pre-organized' for exon ligation. This finding also corresponded to previous work done in the *P. carinii*-derived ribozyme system mentioned above; an increased P10 helix length resulted in greater product formation, with an optimum length of 5 base pairs (31). The correlation between these two studies suggests that these length preferences are fundamental to reactions catalyzed by the *P. carinii* group I intron-derived ribozymes.

5.4.3 Length Dependence of L1 Loop Region

In this system, the original purpose for the sequence of the loop region was to connect the new 5' exon to the ribozyme. It obviously, however, played a role in determining the reaction rate, evidenced by a 5-fold increase in the observed rate constant when comparing the original ribozyme construct (with a 6-base L1 loop) and the optimized construct (with a 10-base loop). The original ribozyme construct differed from the more effective construct at only the loop length. The P9.0 and P10 helices were the same. What can also be seen from the change in the

reaction rate is that a longer L1 loop also increased the extent of product formation. Tdwo possible theories could apply to why the reaction benefited from a longer loop region. The first concerns flexibility; perhaps the 10-base loop region provided ample 'maneuvering room' for the portion of the ribozyme containing the 5' exon and relieved any strain in the P1 or P10 helices. With a loop of 12 or 14 bases, the relieved strain could have turned into a 'floppy', unstable confirmation that would explain the decrease in product as the loop length increased. The second theory has as its basis the similarity between the 10-base loop region and the actual sequence (CACCUUCU) of the native L1 loop found in the *P. carinii* intron in nature. Beyond the first 6 bases in both loops, the remaining bases were pyrimidines, and perhaps lended a structural similarity of the natural conformation of the intron to the optimized ribozyme construct. Changing the sequence of the L1 loop to more closely resemble the native intron in this way could have served to bring the ribozyme closer to the native state of the intron, and thereby increased its activity.

5.4.4 Effect of GMP on 5' replacement splicing

The study of GMP's effect on the 5' replacement splicing reaction revealed a lowering of product levels as the concentration of GMP increased. This obviously differs from ribozyme 3' replacement splicing, which needs GMP for the first step of the reaction. In that system, GMP is at a concentration of 100 μM in the *in vitro* reaction; our investigation shows a 4-fold decrease in product at 100 μM (Figure 5.7). Further testing is required to ascertain whether this decrease in product is due to direct GMP inhibition of the reaction or to GMP-mediated degradation of the substrate and/or reaction product. The TES reaction (derived from the same group I intron), is also inhibited by GMP; therefore, evidence exists of a possible inherent trait of GMP-mediated inhibition or decrease in product accumulated with the reactions from this class of ribozymes. On the flip side of the coin, trans-splicing (ribozyme 3' replacement splicing) was reported for the first time in 1994; yet we report 5' transcript replacement catalyzed by a ribozyme for the first time here. Thus, an impediment to the formation and retention of reaction product in the form of GMP inhibition or degradation could also explain why the 5' replacement splicing reaction has not been presented before now, and why it is presented in a system other than those reported with ribozyme 3' replacement splicing. From a practical standpoint, however, this effect could hinder

the success of the reaction if it is implemented *in vivo*, due to the presence of such cofactors (ranging from 200-700 μM concentrations) in cellular systems (102).

5.5 Comparison with previous work

The previously mentioned trans-splicing (TS) reaction, or ribozyme 3' replacement, focuses on the correction (by replacement) of mutant 3' end of an RNA transcript (29). Therefore, trans-splicing ribozymes carry a replacement 3' exon. This reaction also requires a guanosine cofactor. Trans-splicing has been shown to work *in vitro* and *in vivo* in a variety of systems (29, 33, 73-74, 103-108). The reaction rate of the TS reaction *in vitro* is 0.17 min^{-1} (109), as compared to the optimized rate of 9.15 x 10^{-3} min^{-1} of the 5' replacement splicing reaction.

Trans-splicing does, however, possess its own potential limitations. Mutations located on the 3' end of a transcript would be the most reasonable targets for repair. Repair of mutations on the 5' end of a transcript may be attempted, but problems may arise depending on the amount of bases that could be effectively carried by the delivery plasmid in an *in vivo* system. The 5' replacement splicing reaction, in contrast, is specifically designed for mutations on the 5' end of a transcript, providing an alternate avenue for this type of repair. In addition, 5'RS does not require a guanosine cofactor.

As for recognition element investigations, the P10 helix has been implemented in TS reactions and found to be beneficial (32). The length of this helix was tested from 0 to 6 base pairs, with 6 base pairs as the normal construct. Those P10 helices with 0 and 2 base pairs resulted in a drop in ribozyme activity *in vivo* as compared to the activity shown with 4 or 6 base pair helices. These results bear similarity to our findings of lengthier P10 regions proving advantageous in the 5' replacement splicing system, specifically a 6 base-pair P10 helix.

Considering studies on the effect of L1 loop length, there is a loop region in trans-splicing ribozymes similar in placement to the one in 5' replacement splicing ribozymes. One study has reported on the effect of its length on cis-splicing activity in mammalian cells (110). L1 loop insertions from 4-84 nucleotides were tested for their ability to enhance the splicing activity of a *Tetrahymena* ribozyme in cells. The loop was altered from the native 14 bases up to 98 bases in length. Loop lengths of 18, 37, 38 bases were the most productive. The study did not examine loop lengths shorter than the native loop, however, the finding of an expanded L1 loop

region benefiting the reaction correlates with what we report here. The authors of the previous report cited changes in ribozyme-substrate flexibility to explain these results.

Another type of transcript modification, spliceosome-mediated RNA trans-splicing, or SMaRT, has been exploited with both 3' and 5' exon replacement (111-114). It utilizes a pre-trans-splicing molecule (PTM), which includes a binding domain and the correct replacement RNA sequence for the transcript. It might suffer, however, from certain potential limitations. For example, the available splice sites are exclusively at intron-exon junctions. As a result, nuclear introns are required for these types of systems. The 5' replacement splicing reaction, on the other hand, does not require that the splice sites reside at intron-exon junctions. Thus its targeting capabilities are not limited to only replacing entire intron-exon segments of a transcript or restricted from repairing just a portion of an exon.

5.6 Implications

For the first time, a ribozyme-mediated 5' replacement splicing reaction has been developed that can and does occur. Through investigation of the effect of the P9.0 on the 5' replacement splicing reaction, an 8-base region was shown to result in increased product yields. The P10 region can also improve this reaction through elongation; this effect is most apparent when paired with a lengthier L1 loop region. A 6 base-pair P10 in conjunction with a 10-base loop increases product formation more than 2-fold and the reaction rate over 5-fold. This finding provides evidence of a significant role for the loop region, one that can be exploited in future studies. The design of this ribozyme reaction is somewhat unique in its reliance on a 3' recognition element of a group I intron as the primary recognition element for the substrate. This relationship between reaction success and the molecular interactions creates a better understanding of how to manipulate the factors in the reaction to enhance its chances for success.

The reaction rate is low, compared to the rates of other reactions for the *P. carinii* system. The rate of the native self-splicing intron is 0.9 min^{-1}, while the rate of its suicide inhibition is more than 10-fold lower at 0.081 min^{-1} (115). The TES reaction, which provided some of the foundation for this design, has 0.05 min^{-1} as its lowest rate (33). As low as this rate may be compared to the others mentioned here, it is still 5-fold faster than the optimized 5' replacement splicing reaction. Although the rate is relatively low for a group I ribozyme-catalyzed reaction, it is different in design from what has previously been done with the *P. carinii*-derived ribozyme.

Another obstacle for this reaction could be effective binding of the 3’ exon. As mentioned previously, weaker 3' exon binding has been found in a *P. carinii*-derived ribozyme; this could result in problematic dissociation of the 3’ exon. The 5’ replacement splicing reaction obviously requires that the 3' exon remain bound until reaction completion, as its ligation to the new 5’ exon creates the reaction product. Thus weaker binding and 3’ exon dissociation may play a part in the slower reaction rate and greater requirement for magnesium. Perhaps these are tradeoffs needed for a successful reaction. Previous studies present evidence, however, that the efficiency of group I intron-derived ribozymes can be increased by insertions and mutations (116-117). Some of the possible limitations of this 5’ replacement splicing reaction system could be overcome with random mutagenic studies to determine beneficial alterations to the ribozyme.

The findings discussed here provide additional insight on the inherent aspects of group I-intron-derived ribozymes, as well as the potential applications of such systems. The most obvious and hopeful of the applications for this system deals with the therapeutic pathway. As previously mentioned, this reaction system may be examined as a possible treatment of mutation-related diseases.

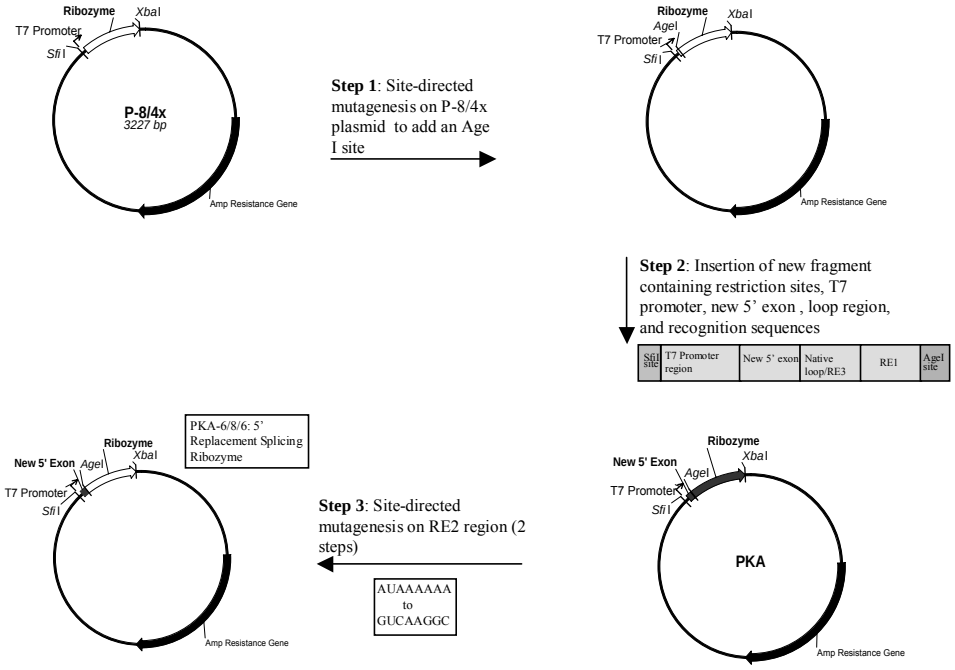

Figure 5.1. 5' Replacement splicing ribozyme plasmid design. 5' Replacement Splicing Plasmid Design. The starting plasmid, P-8/4x was modified with an Age I restriction site in the region containing the ribozyme (Step 1). Step 2 utilized an existing SfiI restriction site and the new AgeI site to insert a double-stranded fragment containing the necessary elements of the 5' replacement splicing ribozyme. Step 3 required 2 rounds of site-directed mutagenesis to alter the RE2 region. The finished plasmid is designated PKA; the digits following the base plasmid name describe the L1 loop/RE2 region/RE3 region lengths.

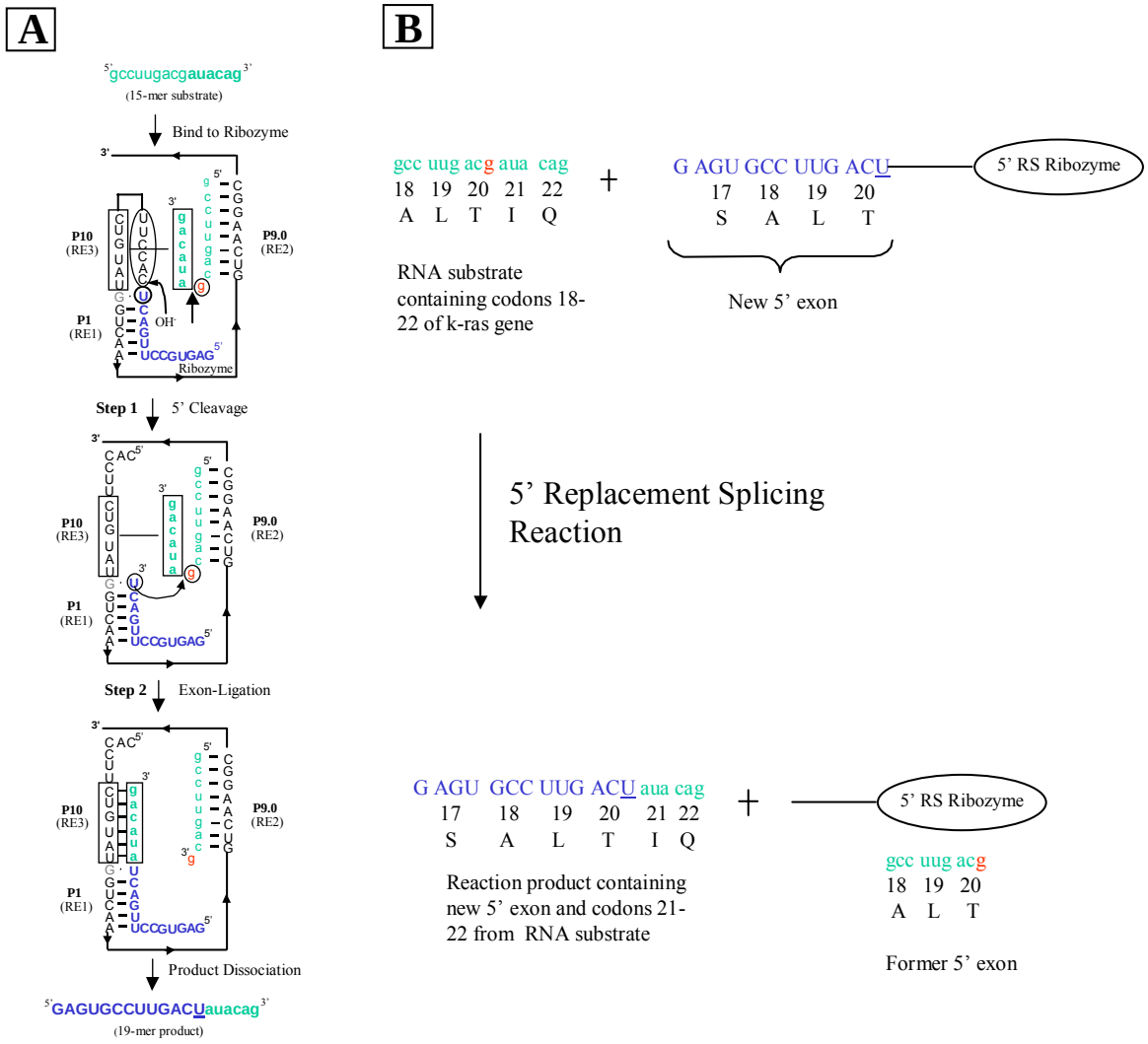

Figure 5.2. 5' Replacement splicing system design. **(A)** The 5' replacement splicing ribozyme is in uppercase lettering, with the exon in blue, while the 16-mer substrate is in green lowercase lettering and the exon regions to be ligated are in bold. The ribozyme recognition element RE1 base pairs with the 5' exon-ribozyme conjugate to form the P1 helix (an intramolecular interaction). The ribozyme recognition elements RE2 and RE3 base pair with the substrate to form the P9.0 and P10 helices, respectively. Arrows in the uppermost diagram represent the catalytic sites for the 5' cleavage and exon-ligation steps. The P10 helix is boxed and the L1 loop is circled. The 5' uridine (blue) and 3' guanosine (gray) that help define the catalytic sites are circled. **(B)** The small model substrate contains codons 18-22 and is denoted with green lowercase letters. The replacement 5' exon attached to the ribozyme is in blue and denoted with uppercase letters; it contains the last base of codon 16 through the last base of the new codon 20. This last base is a uridine (underlined) in the replacement exon, altering the nucleotide sequence, but not the amino acid sequence. The reaction results in a product containing the entire replacement exon and two codons from the substrate. Note that the product is 4 bases longer than the substrate, which aids in identification.

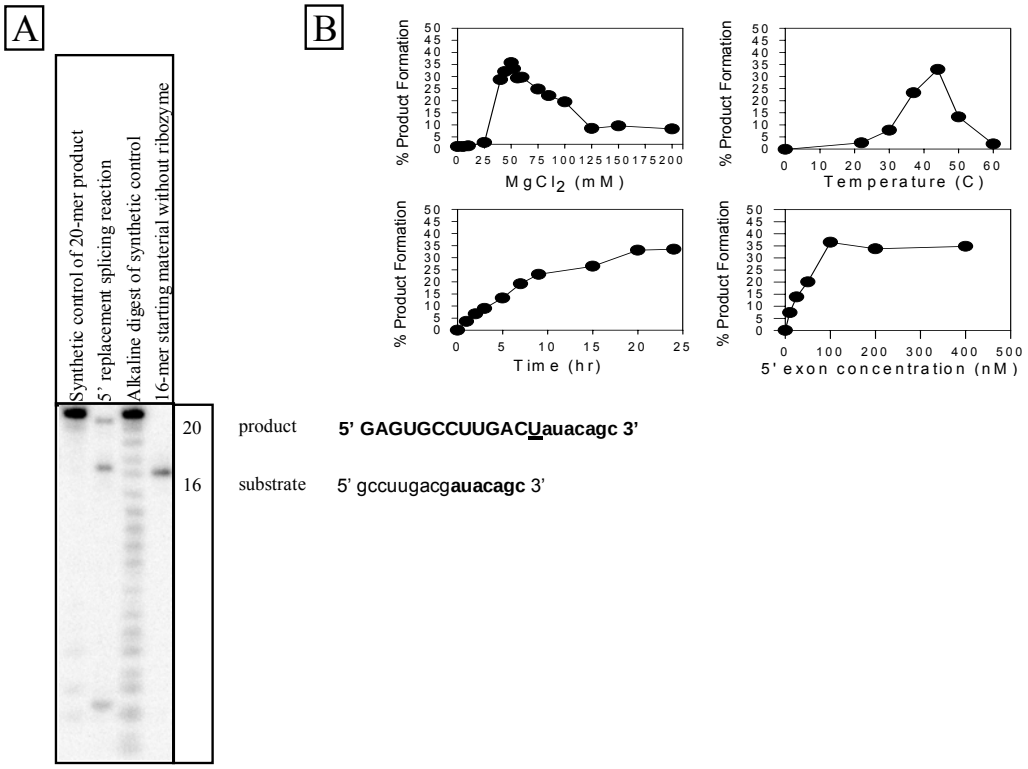

Figure 5.3. Representative gel and optimization data with rPKA-6/8/6. (**A**) Polyacrylamide gel showing product and substrate from the 5' replacement splicing reaction. The plasmid from which the ribozymes are transcribed is PKA; the ribozyme itself is denoted as rPKA-6/8/6 (base plasmid-L1 loop/RE2/RE3 lengths). Reactions were run using 100 nM rPKA-6/8/6 ribozyme, 8 nM radiolabeled substrate, and 50 mM $MgCl_2$ at 44 °C for 20 hours. The first lane contains the synthetic reaction product; the second lane contains the 5' replacement splicing reaction with the 16-mer substrate yielding the 20-mer product. The third lane shows an alkaline digest of the synthetic product, and the fourth lane shows the 16-mer starting material under reaction conditions with no ribozyme. Shown to the right of the gel are the sequences and the sizes of the substrate and product. Note that there is a radiolabeled cytosine on the 3' end of both the substrate and the product. (**B**) Optimization of the replacement splicing reaction with rPKA-6/8/6. The reactions were run as described above except for the changing variable. Each data point is the average of two independently run assays. The standard deviation for all points is less than 10%.

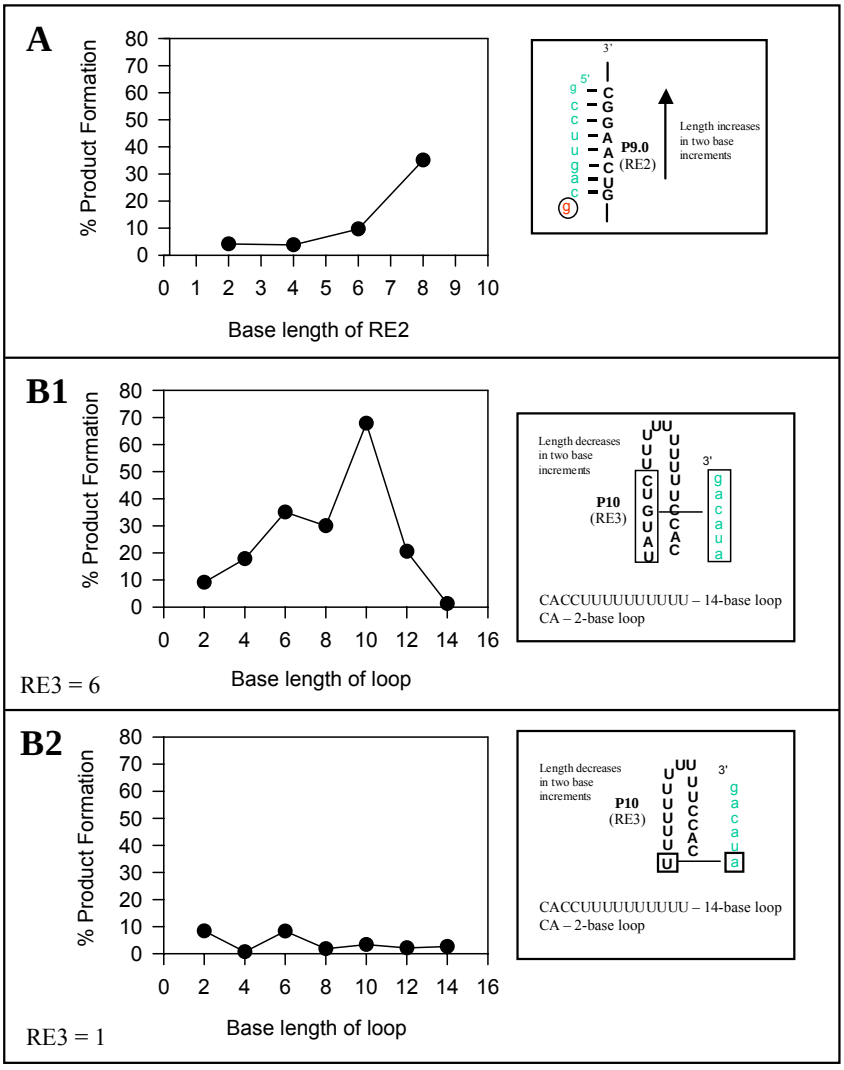

Figure 5.4. P9.0, P10, and L1 loop length dependence data. The diagram to the right of each graph shows the structures formed by the recognition elements. The reactions were run with 100 nM ribozyme, 8 nM substrate, and 50 mM $MgCl_2$ at 44°C for 20 hours. All data points are the average of 2 independent assays; the standard deviation for all points is less than 10%. (**A**) The effect of P9.0 length on product formation. The longest length tested (8 bases) is shown at the right. The RE2 region is shown in the diagram in bold lettering. Both RE3 and the L1 loop were each at 6-base lengths. (**B1**) The effect of loop length with an RE3 of 6 bases. The reactions were run as described above. The longest length tested (14 bases) is shown to the right of the graph. The RE3 and loop regions are shown on the right in bold lettering. The P9.0 was at a length of 8 bases. (**B2**) The effect of loop length with an RE3 of 1 base. The longest lengths tested (14 bases) are shown to the right of the graph. The P9.0 was at a length of 8 bases.

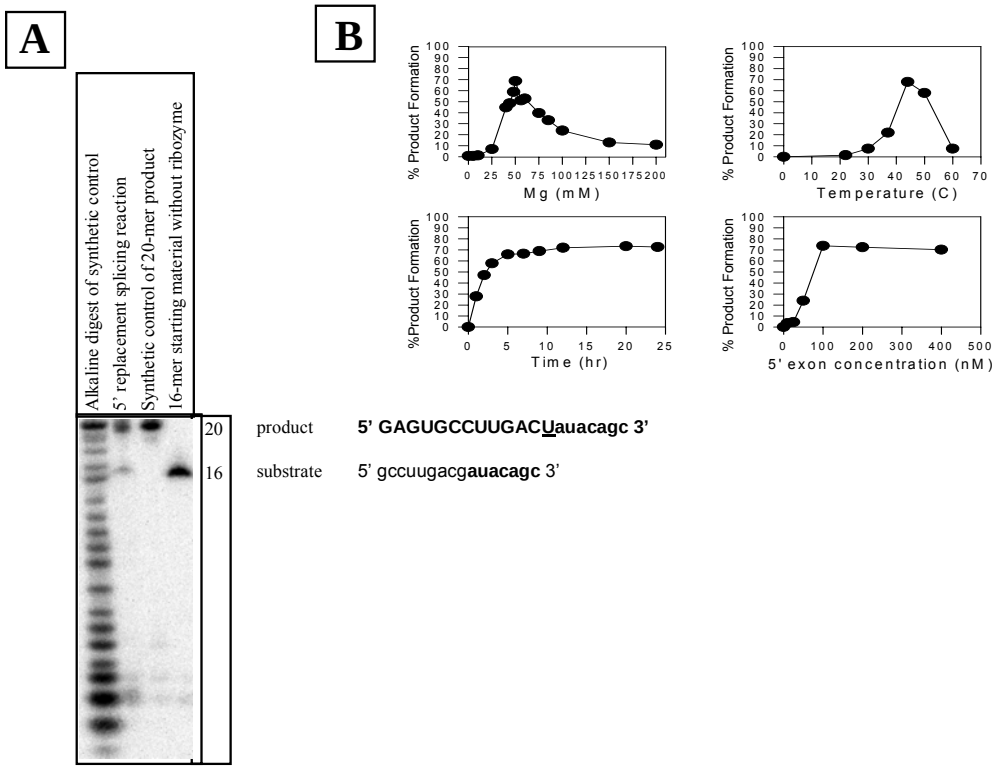

Figure 5.5. Representative gel and optimzation data using rPKA-10/8/6. (**A**) Polyacrylamide gel showing product and substrate from the 5' replacement splicing reaction with rPKA-10/8/6 using 100 nM ribozyme, 8 nM radiolabeled substrate, and 50 mM $MgCl_2$ at 44 °C for 20 hours. The first lane contains an alkaline digest of the synthetic product; the second lane contains the 5' replacement splicing reaction with the 16-mer substrate yielding the 20-mer product. The third lane shows the synthetic reaction product; and the fourth lane shows the 16-mer starting material only under reaction conditions without ribozyme. Shown to the right of the gel are the sequences and sizes of the substrate and product. Note that there is a radiolabeled cytosine on the 3' end of both the substrate and the product. (**B**) Optimization of the replacement splicing reaction with rPKA-10/8/6. The reactions were run as described above except for the changing variable. Each data point is the average of two independently run assays, with standard deviations below 10%.

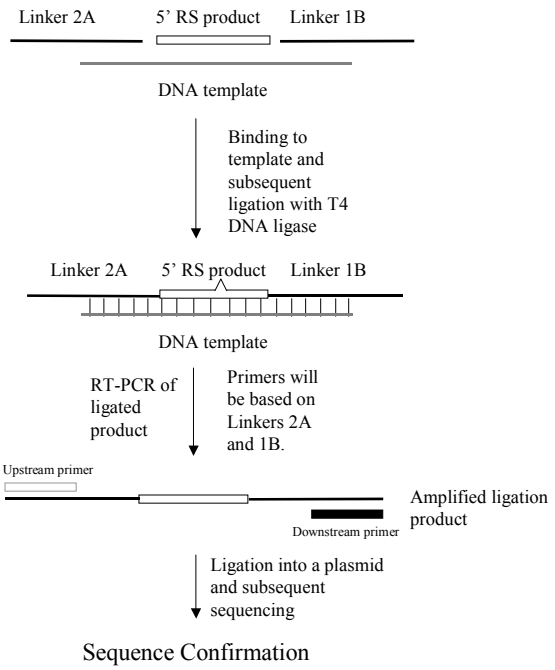

Figure 5.6. Sequencing protocol design. diagram illustrates the sequencing protocol for the small model 5' replacement splicing product. The linkers are shown in black and are DNA, the DNA template is shown in gray. The 5' replacement splicing product (from reactions using the final ribozyme construct) is shown as an open box. The system is designed so that the template does not contain a complementary base for the new uridine at the splice site; thus, there is a 1-nucleotide bulge. Once the linkers and 5' replacement splicing product have paired with the template, they are ligated using T4 DNA ligase. This ligated product is isolated and amplified using RT-PCR. The primers used correspond with the 5' end of linker 2A and the 3' end of linker 1B, respectively. The primers are shown as solid bars. Following RT-PCR amplification, the product is isolated and ligated into a plasmid. The plasmid was then sequenced for product confirmation.

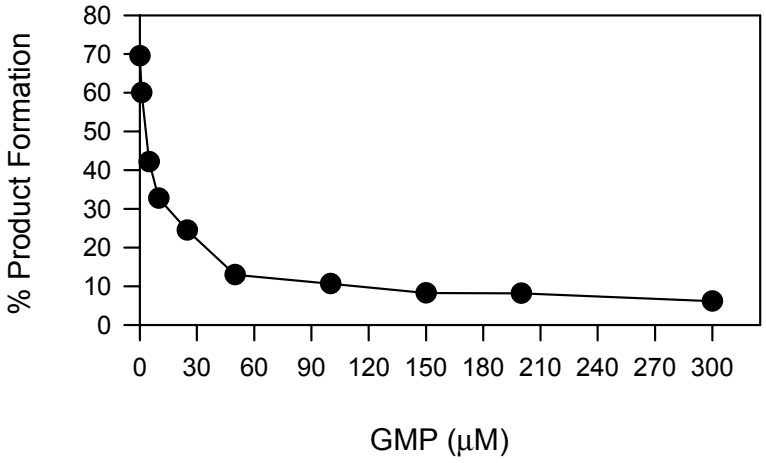

Figure 5.7. GMP study. Shown is product observed when using the optimized 5' replacement splicing construct when increasing amounts of GMP are added to the reaction. The reactions were run with 100 nM ribozyme, 8 nM substrate, and 50 mM $MgCl_2$ at 44°C for 20 hours. Each data point is the average of two independent assays with standard deviations below 10%.

Chapter 6 - Summary and Future Work

Each of the investigations detailed in this dissertation lend additional knowledge to the field of nucleic acid studies. They also provide the groundwork for continued studies in the areas of DNA ligation, nucleic acid purification, and ribozyme-catalyzed reactions.

The study of canonical nucleobases as universal templates in ligation reactions revealed reaction conditions that resulted in lowered fidelity for T4 DNA ligase (118). This study also reported thymidine as the most consistently effective universal template nucleobase under these specific conditions. As previously mentioned, this work could be utilized in the future to develop new protocols for the sequencing of oligonucleotides This would require the testing of longer templates containing stretches of thymidines in tandem for the ability to direct the sequence-independent ligation (at the actual nick) of three or more oligonucleotides bound to it. In terms of the protocols that presently use DNA ligases and rely on the fidelity thereof, the findings presented here impart knowledge of what should be considered when altering reaction conditions. As the lowering of the $MgCl_2$ and ATP concentrations results in an obvious reduction in fidelity, these are definitely parameters that should be considered in depth whenever altered. As for DMSO, it is not always used in those molecular biology protocols that use ligases (44-46). Its use in any protocol, however, requiring enzyme fidelity should be considered with caution.

The study of 2-butanol extraction as a method of urea removal contributes an efficient way to eliminate urea from nucleic acid samples. This procedure could be used as a main purification method or as a preparatory step prior to ethanol precipitation or chromatography. Butanol extraction could also be tested alongside either of those methods to ascertain a direct comparison of efficiency.

The novel ribozyme-catalyzed 5’ transcript replacement reaction discussed in this dissertation supplies a foundation for future work, for therapeutic applications and to further explore group I intron-derived ribozymes. The next step with this work would be its implementation in a large transcript system. Such a system would need to be examined to compare the reaction in a large model system to the small model system. Once the large transcript system has been optimized, *in vivo* mutagenic studies would need to be done to find alterations to the ribozyme that would grant it activity under physiological conditions. This

would translate to a lower $MgCl_2$ required for activity, as well as a reduction in the inhibition seen when a guanosine cofactor is present with the reaction components. The subsequent studies would evaluate the ability of the newer ribozymes to function as a repair method in both bacterial and mammalian cells. The 5' replacement splicing reaction study also gives information as to what elements of group I intron-derived ribozymes may be exploited to increase the abilities of the catalytic constructs. Longer P9.0 and P10 helices have been shown to benefit more than one ribozyme-catalyzed reaction, as have longer L1 loop regions (32, 98, 110). Thus, these interactions could be further studied with various sequences to ascertain whether additional benefits can be obtained through modification of these molecular aspects.

References

1. Karp, G. (1999) Cell and Molecular Biology: Concepts and Experiments (Second Edition), John Wiley and Sons, Inc., New York, NY.
2. Bourgaize, D., Jewell, T.R., Buiser, R.G. (2000) Biotechnology: Demystifying the Concepts, Addison Wesley Longman, San Francisco, CA.
3. Voet, D., Voet, J.G., Pratt, C.W. (1999) Fundamentals of Biochemistry (First Edition), John Wiley and Sons, Inc., New York, NY.
4. Glick, B.R. and Pasternak, J.J. (1998) Molecular Biotechnology: Principles and Applications of Recombinant DNA (Second Edition), ASM Press, Washington, D.C.
5. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982) Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of *Tetrahymena*. *Cell*, **31**, 147-157.
6. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell*, **35**, 849-857.
7. Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004) A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science*, **306**, 275-279.
8. Weiss, B. and Richardson, C.C. (1967) Enzymatic Breakage and Joining of Deoxyribonucleic Acid, I. Repair of Single-Strand Breaks in DNA by an Enzyme System From *Escherichia Coli* Infected With T4 Bacteriophage. *Proc. Natl. Acad. Sci. USA*, **57**, 1021-1028.
9. Cozzarelli, N.R., Melechen, N.E., Jovin, T.M., and Kornberg, A. (1967) Polynucleotide Cellulose As a Substrate For a Polynucleotide Ligase Induced By Phage T4. *Biochem. Biophys. Res. Commun.*, **28**, 578-586.
10. Gupta, N.K., Ohtsuka, E., Weber, H., Chang, S.H., and Khorana, H.G. (1968) Studies on Polynucleotides, LXXXVII. The Joining of Short Deoxyribopolynucleotides by DNA-Joining Enzymes. *Proc. Natl. Acad. Sci. USA*, **60**, 285-292.
11. Weiss, B., Thompson, A., and Richardson, C.C. (1968) Enzymatic Breakage and Joining of Deoxyribonucleic Acid, VII. Properties of the Enzyme-Adenylate Intermediate in the Polynucleotide Ligase Reaction. *J. Biol. Chem.*, **243**, 4556-4563.

12. Lehman, I.R. (1974) DNA Ligase: Structure, Mechanism, and Function. *Science*, **260**, 790-797.
13. Wu, D. and Wallace, R.B. (1989) Specificity of the Nick-Closing Activity of Bacteriophage T4 DNA Ligases. *Gene*, **76**, 245-254.
14. Cherepanove, A., Yildirim, E., and de Vries, S. (2001) Joining of Short DNA Oligonucleotides with Base Pair Mismatches by T4 DNA Ligase. *J. Biochem. (Tokyo, Jpn),* **129**, 61-68.
15. Cherepanov, A.V. and de Vries, S. (2002) Scanning Mutagenesis Using T4 DNA Ligase and Short Degenerate DNA Oligonucleotides Containing Tri-Nucleotide Mismatches. J. Biochem (Tokyo, Jpn), **132**, 143-147.
16. Tsiapalis, C.M. and Narang, S.A. (1970) On the Fidelity of Phage T4-Induced Polynucleotide Ligase in the Joining of Chemically Synthesized Deoxyribooligonucleotides. *Biochem. Biophys. Res. Commun.*, **39**, 631-636.
17. Sgaramella, V. and Khorana, H.G. (1972) Total Synthesis of the Structural Gene for an Alanine Transfer RNA From Yeast. Enzymic Joining of the Chemically Synthesized Polydeoxynucleotides to Form the DNA Duplex Representing Nucleotide Sequence 1 to 20. *J. Mol. Biol.*, **72**, 427-444.
18. Wiaderkiewicz, R. and Ruiz-Carillo, A. (1987) Mismatch and Blunt to Protruding-End Joining by DNA ligases. *Nucleic Acids Res.*, **15**, 7831-7848.
19. Goffin, C., Bailly, V., and Verly, W.G. (1987) Nicks 3' or 5' to AP sites or to Mispaired Bases, and One-Nucleotide Gaps Can Be Sealed by T4 DNA Ligase. *Nucleic Acids Res.*, **15**, 8755-8771.
20. Tomkinson, A.E., Tappe, N.J, and Friedberg,E .C. (1992) DNA Ligase I from Saccharomyces Cerevisiae: Physical and Biochemical Characterization of the CDC9 Gene Product. *Biochemistry*, **31**, 11762-11771.
21. Harada, K. and Orgei, L.E. (1993) Unexpected Substrate Specificity of T4 DNA Ligase Revealed by *In Vitro* Selection. *Nucleic Acids Res.*, **21**, 2287-2291.
22. Shuman, S. (1995) Vaccinia Virus DNA Ligase: Specificity, Fidelity, and Inhibition. *Biochemistry*, **34**, 16138-16147.

23. Husain, I., Tomkinson, A.E., Burkhart, W.A., Moyer, M.B., Ramos, W., Mackey, A.B., Besterman, J.M., and Chen, J. (1995) Purification and Characterization of DNA Ligase III from Bovine Testes. *J. Biol. Chem.*, **270**, 9683-9690.
24. Luo, J., Bergstrom, D.E., and Barany, F. (1996) Improving the Fidelity of *Thermus Thermophilus* DNA Ligase. *Nucleic Acids Res.*, **24**, 3071-3078.
25. Pritchard, C.E. and Southern, E.M. (1997) Effects of Base Mismatches on Joining of Short Oligodeoxynucleotides by DNA Ligases. *Nucleic Acids Res.*, **25**, 3403-3407.
26. Sriskanda, V. and Shuman, S. (1998) Specificity and Fidelity of Strand Joining by Chlorella Virus DNA Ligase. *Nucleic Acids Res.*, **26**, 3536-3541.
27. Pan, J., and Woodson, S.A. (1998) Folding Intermediates of a Self-Splicing RNA: Mispairing of the Catalytic Core, *J Mol Biol*., **280,** 597-609.
28. Rook, M.S., Treiber, D.K., and Williamson, J.R. (1998) Fast Folding Mutants of the *Tetrahymena* Group I Ribozyme Reveal a Rugged Folding Energy Landscape. *J Mol Biol.*, **281,** 609-620.
29. Sullenger, B.A., and Cech, T.R. (1994) Ribozyme-Mediated Repair of Defective mRNA by Targeted, Trans-Splicing. *Nature*, **371**, 619-622.
30. Scherer, L.J. and Rossi, J.J. (2003) Approaches for the Sequence-Specific Knockdown of mRNA. *Nat. Biotech.*, **21**(17), 1457-65.
31. Testa, S.M., Haidaris, C.G., Gigliotti, F., and Turner, D.H. (1997) A *Pneumocystis carinii* Group I Intron Ribozyme That Does Not Require 2' OH Groups on Its 5' Exon Mimic for Binding to the Catalytic Core. *Biochemistry*, 36, 15303-15314.
32. Kohler, U., Ayre, B.G., Goodman, H.M., and Haseloff, J. 1999. Trans-splicing Ribozymes for Targeted Gene Delivery. *J. Mol. Biol.*, 285, 1935-1950.
33. Bell, M.A., Johnson, A.K., and Testa, S.M. (2002) Ribozyme-Catalyzed Excision of Targeted Sequences from within RNAs. *Biochemistry*, **41**, 15327-15333.
34. Sargueil, B. and Tanner, N.K. (1993) A Shortened Form of the *Tetrahymena thermophila* Group I Intron Can Catalyze the Complete Splicing Reaction in trans. *J. Mol. Biol*., **233**, 629-643.
35. de Bolster, M.W.G (1997) Glossary of Terms Used in Bioinorganic Chemistry. *Pure Appl. Chem*. **69**, 1251-1303.

36. Doudna, J.A. and Cech, T.R. (2002) The Chemical Repertoire of Natural Ribozymes. *Nature*, **418**, 222-228.
37. Sokolova, N.I., Ashirbekova, D.T., Dolinnaya, N.G., and Shabarova, Z.A. (1988) Chemical Reactions Within DNA Duplexes: Cyanogen Bromide as an Effective Oligodeoxyribonucleotide Coupling Agent. *FEBS Letters*, **232**, 153-155.
38. Dolinnaya, N.G., Sokolova, N.I., Gryaznova, O.I., and Shabarova, Z.A. (1988) Site-Directed Modification of DNA Duplexes by Chemical Ligation. *Nucleic Acids Res.*, **16**, 3721-3738.
39. Gryaznov, S.M., Schultz, R., Chaturvedi, S.K., and Letsinger, R.L. (1994) Enhancement of Selectivity in Recognition of Nucleic Acids via Chemical Autoligation. *Nucleic Acids Res.*, **22**, 2366-2369.
40. James, K.D. and Ellington, A.D. (1997) Surprising Fidelity of Template-Directed Chemical Ligation of Oligonucleotides. *Chemistry & Biology*, **4**, 595-605.
41. Harada, K. and Orgel, L.E. (1994) *In Vitro* Selection of Optimal DNA Substrates for Ligation by a Water-Soluble Carbodiimide. *J. Mol. Evol.*, **38**, 558-560.
42. James, K.D. and Ellington, A.D. (1999) The Fidelity of Template-Directed Oligonucleotide Ligation and the Inevitability of Polymerase Function. *Origins of Life and the Evolution of the Biosphere*, **29**, 375-390.
43. Rossi, R., Montecucco, A., Ciarrocchi, G. and Biamonti, G. (1997) Functional Characterization of the T4 DNA Ligase: A New Insight into the Mechanism of Action. *Nucleic Acids Res.*, **25**, 2106-2113.
44. Barany, F. (1991) Genetic Disease Detection and DNA Amplification Using Cloned Thermostable Ligase. *Proc. Natl. Acad. Sci. USA*, **88**, 189-193.
45. Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988) A Ligase-Mediated Gene Detection Technique. *Science*, **241**, 1077-1080.
46. Alves, A.M. and Carr, F.K. (1988) Dot Blot Detection of Point Mutation with Adjacently Hybridising Synthetic Oligonucleotide Probes. *Nucleic Acids Res.,* **18**, 8723.
47. Loakes, D. (2001) The Applications of Universal DNA Base Analogues. *Nucleic Acids Res.*, **29**, 2437-2447.

48. Loakes, D., Aerschot, A.V., Brown, D.M., and Hill, D. (1996) Enzymatic Recognition of Acyclic Universal Base Analogues in Oligonucleotides. *Nucleosides & Nucleotides*, **15**, 1891-1904.
49. Loakes, D. and Brown, D.M. (1994) 5-Nitroindole as a Universal Base Analogue. *Nucleic Acids Res.*, **22**, 4039-4043.
50. Oliver, J.S., Parker, K.A., and Suggs, J.W. (2001) Effect of the Universal Base 3-Nitropyrrole on the Selectivity of Neighboring Natural Bases. *Organic Letters.* **3**(13), 1977-1980.
51. Nasri, M. and Thomas, D. (1986) Relaxation of Recognition Sequence of Specific Endonuclease HindIII. *Nucleic Acids. Res*., **14**, 811-821.
52. Pingoud, A. and Jeltsch, A. (1997) Recognition and Cleavage of DNA by Type-II Restriction Endonucleases. *Eur. J. Biochem.*, **246**, 1-22.
53. Zimmerman, S.B. and Pheiffer, B.H. (1983) Macromolecular Crowding Allows Blunt-End Ligation by DNA Ligases from Rat Liver of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. **80**, 5852.
54. Kaczorowski, T. and Syzbalski, W. (1994) Assembly of 18-Nucleotide Primers by Ligation of Three Hexamers: Sequencing of Large Genomes by Primer Walking. *Anal. Biochem*., **221**, 127-135.
55. Kaczorowski, T. and Syzbalski, W. (1996) Co-Operativity of Hexamer Ligation. *Gene*, **179**, 189-193.
56. Sambrook, J., and Russell, D.W. (2001) Molecular Cloning, A Laboratory Manual (Third Edition), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
57. Hsieh, J., Andrews, A.J., and Fierke, C.A. (2004) Roles of Protein Subunits in RNA-Protein Complexes: Lessons from Ribonuclease P. *Biopolymers*. **73**(1), 79-89.
58. Uhlenbeck, O.C. (1987) A Small Catalytic Oligoribonucleotide. *Nature*, **328**, 596-600.
59. Symons, R.H. (1992) Small Catalytic RNAs. *Annu. Rev. Biochem*., **61**, 641-671.
60. Campbell, T.B. and Sullenger, B.A. (1995) Alternative Approaches for the Application of Ribozymes as Gene Therapies for Retroviral Infections. *Adv. Pharm*., **33**, 143-178.
61. van der Veen, R., Amberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F., and Grivell, L.A. (1986) Excised Group II Introns in Yeast Mitochondria Are Lariats and Can Be Formed by Self-Splicing *In Vitro*. *Cell*, **44** (2), 225-34.

62. Sarver, N., Cantin, E.M., Chang, P.S., Zaia, J.A., Ladne, P.A., Stephens, D.A., and Rossi, J.J. (1990) Ribozymes as Potential Anti-HIV-1 Therapeutic Agents. *Science*, **247**(4947), 1222-1225.
63. Ojwang, J.O., Hampel, A., Looney, D.J., Wong-Staal, F., and Rappaport, J. (1992) Inhibition of Human Immunodeficiency Virus Type 1 Expression by a Hairpin Ribozyme. *Proc. Natl. Acad. Sci. USA*, **89**, 10802-10806.
64. Sullenger, B.A. and Gilboa, E. (2002) Emerging Clinical Applications of RNA. *Nature*, **418**, 252-258.
65. Gesteland, R.F. and Atkins, J.F. (1993) *The RNA World*. Cold Spring Harbor Laboratory Press. Plainview, NY.
66. Lambowitz, A.M. and Perlman, P.S. (1990) Involvement of Aminoacyl-tRNA Synthetases and Other Proteins in Group I and Group II Intron Splicing. *Trends Biochem. Sci*., **15**, 440-444.
67. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004) Crystal Structure of a Self-Splicing Group I Intron with Both Exons. *Nature*, **2642**, 1-6.
68. Cate, J.H. et al. (1996) Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science*, **273**, 1678-1685.
69. Strobel, S.A., Ortoleva-Connelly, L., Ryder, S.P., Cate, J.H., and Moncoeur, E. (1998) Complementary Sets of Noncanonical Base Pairs Mediate RNA Helix packing in Group I Intron Active Site. *Nature Struct. Biol.*, **5**, 60-66.
70. Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. (1998) A Preorganized Active Site in the Crystal Structure of the *Tetrahymena* Ribozyme. *Science*, **282**, 259-264.
71. Michel, F. and Westhof, E. (1990) Modelling of the Three-Dimensional Architecture of Group I Catalytic Introns Based on Comparative Sequence Analysis. *J. Mol. Biol*., **216**, 585-610.
72. Johnson, A.K., Baum, D.A., Tye, J., Bell, M.A., and Testa, S.M. (2003) Molecular Recognition Properties of IGS-Mediated Reactions Catalyzed by a *Pneumocystis carinii* Group I Intron. *Nucleic Acids Res.*, **31**(7), 1921-1934.
73. Szostak, J.W. (1986) Enzymatic Activity of the Conserved Core of a Group I Self-Splicing Intron. *Nature*, **322**(6074), 83-86.

74. Watanabe,T. and Sullenger,B.A. (2000) Induction of Wild-Type p53 Activity in Human Cancer Cells by Ribozymes That Repair Mutant p53 Transcripts. *Proc. Natl. Acad. Sci. USA*, **97**, 8490-8495.

75. Lan, N., Howrey, R.P., Lee, S.W., Smith, C.A, and Sullenger, B.A. (1998) Ribozyme-Mediated Repair of Sickle Beta-Globin mRNAs in Erythrocyte Precursors. *Science*, **280**: 1593-1596.

76. Sogin, M.L. and Edman, J.C. (1989) A Self-Splicing Intron in the Small Subunit rRNA Gene of *Pneumocystis carinii*. *Nucleic Acids Res*., **17**, 5349-5359.

77. Reinhold-Hurek, B. and Shub, D.A. (1992) Self-Splicing Introns in tRNA Genes of Widely Divergent Bacteria. *Nature*, **357**, 173-176.

78. Haugen, P., De Jonckheere, J.F., and Johansen, S. (2002) Characterization of the Self-Splicing Products of Two Complex *Naegleria* LSU rDNA Group I Introns Containing Homing Endonuclease Genes. *Eur. J.Biochem*., **269**, 1641-1649.

79. Barbacid, M. (1987) ras Genes. *Ann. Rev. Biochem.*, **56**, 779-827.

80. Chipperfield, R.G., Jones, S.S., Lo, K.M., and Weinberg, R.A. (1985) Activation of Ha-ras p21 Substitution, Deletion, and Insertion Mutations. *Mol. Cell. Biol.*, **5,** 1809-1813.

81. Ellis, C.A. and Clark, G. (2000) The Importance of Being K-Ras. *Cell Signaling*, **12**, 425-434.

82. Umanoff, H., Edelmann, W., Pellicer, A., and Kucherlapati, R. (1995) The Murine N-ras Gene Is Not Essential for Growth and Development. *Proc. Natl. Acad. Sci. USA*, **92**, 1709-1713.

83. Koera, K., Nakamura, K., Nakao, K., Miyoshi, J., Toyoshima, K., Hatta, T., Otani, H., Aiba, A., and Katsuki, M. (1997) K-Ras Is Essential for the Development of the Mouse Embryo. *Oncogene*, **15**, 1151-9.

84. Jackson, E.L., Willis, N., Mercer, K., Bronson, R.T., Crowley, D., Montoya, R., Jachs, T., Tuveson, D.A. (2001) Analysis of Lung Tumor Initiation and Progression Using Conditional Expression of Oncogenic k-ras. *Genes & Development,* **15**, 3243-3248.

85. Ahrendt, S.A, et al. (2001) Cigarette Smoking is Strongly Associated with Mutation of the K-ras Gene in Patients with Primary Adenocarcinoma of the Lung. *Cancer*, **92**(6), 1525-1530.

86. Huntington's Disease Collaborative Research Group. (1993) A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes. *Cell*, **72**, 971-983, 1993.
87. Bell, M.A., Johnson, A.K., and Testa, S.M. (2002) Ribozyme-Catalyzed Excision of Targeted Sequences from Within RNAs. *Biochemistry*, **41**, 15327-15333.
88. Johnson, A.K., Baum, D.A., Tye, J., Bell, M., and Testa, S.M. (2003) Molecular Recognition Properties of IGS-Mediated Reactions Catalyzed by a *Pneumocystis carinii* Group I Intron. *Nucleic Acids Res.*, **31**, 1921-1934.
89. Xia, T., SantaLucia, J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., and Turner, D.H. (1998) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs. *Biochemistry*, **42**, 14719-14735.
90. Allawi, H.T. and SantaLucia, J. (1998) Nearest-Neighbor Thermodynamic Parameters for Internal G•A Mismatches in DNA. *Biochemistry*, **37**, 2170-2179.
91. Eun, H.M. (1996) Enzymology Primer for Recombinant DNA Technology. Academic Press, San Diego, CA.
92. Herschlag, D. (1991) Implications of Ribozyme Kinetics for Targeting the Cleavage of Specific RNA Molecules *In Vivo*: More Isn't Always Better. *Proc. Natl. Acad. Sci. USA*, **88**, 6921-6925.
93. Peyret, N., Seneviratne,P .A., Allawi, H.T., and SantaLucia, J. (1999) Nearest-Neighbor Thermodynamics and NMR of DNA Sequences With Internal A•A, C•C, G•G, and T•T Mismatches. *Biochemistry*, **38**, 3468-3477.
94. Aboul-ela, F., Koh, D., Tinoco, I., and Martin, F.H. (1985) Base-Base Mismatches. Thermodynamics Double Helix Formation for dCA3XA3G + dCT3YT3G (X,Y = A,C,G,T). *Nucleic Acids Res.*, **13**, 4811-4824.
95. Allawi, H.T. and SantaLucia, J. (1997) Thermodynamics and NMR of Internal G•T Mismatches in DNA. *Biochemistry*, **36**, 10581-10594.
96. Allawi, H.T. and SantaLucia, J. (1998) Nearest-Neighbor Thermodynamics of Internal A•C Mismatches in DNA: Sequence Dependence and pH Effects. *Biochemistry*, **37**, 9435-9444.

97. Allawi, H.T. and SantaLucia, J. (1998) Thermodynamics of Internal C•T Mismatches in DNA. *Nucleic Acids Res.*, **26**, 2694-2701.

98. England, T.G. and Uhlenbeck, O.C. (1978) 3'-Terminal Labeling of RNA with T4 RNA Ligase. *Nature.* **275**, 560-561.

99. Bell, M.A, Sinha, J., Johnson, A.K, and Testa, S.M. (2004) Enhancing the Second Step of the Trans Excision-Splicing Reaction of a Group I Ribozyme by Exploiting P9.0 and P10 for Intermolecular Recognition. Biochemistry, 43, 4323-4331.

100. Djikeng, A., Shi, H., Tschudi, C., and Ullu, E. (2001) RNA Interference in *Trypanosoma brucei*: Cloning of Small Interfering RNAs Provides Evidence for Retroposon-Derived 24-26-Nucleotide RNAs. *RNA*, **7**, 1522-1530.

101. Virlon, B, Cheval, L., Buhler, J., Billon, E., Doucet, A., and Elalouf, J. (1999) Serial Microanalysis of Renal Transcriptomes. *Proc. Natl. Acad. Sci.*, **96**(26), 15286-15291.

102. Traut, T.W. (1994) Physiological Concentrations of Purines and Pyrimidines. *Mol. Cell Biochem*. **140**(1), 1-22.

103. Samarsky, D., Ferbeyre, G., and Bertrand, E. (2000) Expressing Active Ribozymes in Cells. *Curr. Issues Mol. Biol.*, **2**, 87-93.

104. Ayre, B.G, Kohler, U., Goodman, H.M., and Haseloff, J. (1999) Design of Highly Specific Cytotoxins by Using Trans-Splicing Ribozymes. *Proc. Natl. Acad. Sci. USA*, **96**, 3507-3512.

105. Ayre, B.G., Kohler, U., Turgeon, R., and Haseloff, J. (2002) Optimization of Trans-Splicing Ribozyme Efficiency and Specificity by *In Vivo* Genetic Selection. *Nucleic Acids Res.*, **30**(24), e141.

106. Byun, J., Lan, N., Long, M., and Sullenger, B.A. (2003) Efficient and Specific Repair of Sickle β-Globin RNA by Trans-Splicing Ribozymes. *RNA*, **9**, 1254-1263.

107. Shin, K., Sullenger, B.A., and Lee, S. (2004) Riboyzme-Mediated Induction of Apoptosis in Human Cancer Cells by Targeted Repair of Mutant p53 RNA. *Molec. Ther.*, **10**(2), 365-372.

108. Kastanos, E., Hjiantoniou, E., and Phylactou, L.A. (2004) Restoration of Protein Synthesis in Pancreatic Cancer Cells by Trans-Splicing Ribozymes. *Biochem. and Biophys. Res. Com.* **322**: 930-934.

109. Zarrinkar, P.P. and Sullenger, B.A. (1999) Optimizing the Substrate Specificity of a Group I Intron Ribozyme. *Biochemistry*, **38**, 3426-32.

110. Hasegawa, S., Jackson, W.C., Tsien, R.Y., and Rao, J. (2003) Imaging *Tetrahymena* Ribozyme Splicing Activity in Single Live Mammalian Cells. *Proc. Natl. Acad. Sci. USA.*, **100**, 14892-14896.

111. Chao, H., Mansfield, S.G., Bartel, R.C., Hiriyanna, S., Mitchell, L.G, Garcia-Blanco, M.A. and Walsh, C.E. (2003) Phenotype Correction of Hemophilia A Mice by Spliceosome-Mediated RNA Trans-Splicing. *Nat. Med.*, **9**(6), 1015-1016.

112. Garcia-Blanco, M. (2003). Mending the Message. *Nat. Biotech.*, **21**, 1448-1449.

113. Mansfield, S.G., Clark, R.H., Puttaraju, M., Kole, J., Cohn, J.A., Mitchell, L.G., and Garcia-Blanco, M.A. (2003) 5' Exon Replacement and Repair by Spliceosome-Mediated RNA Trans-Splicing. *RNA*, **9**, 1290-1297.

114. Puttanjaru, M., Jamison, S.F, Mansfield, S.G., Garcia-Blanco, M.A., and Mitchell, L.G. (1999) Spliceosome-Mediated RNA Trans-Splicing as a Tool for Gene Therapy. *Nat. Biotech.*, **17**, 246-252.

115. Testa, S.M., Gryaznov, S.M, and Turner, D.H. (1999) *In Vitro* Suicide Inhibition of Self-Splicing of a Group I Intron from *Pneumocystis carinii* by an N3'→P5' Phosphoramidate Hexanucleotide. *Proc. Natl. Acad. Sci. USA*, **96**, 2734-2739.

116. Guo, F. and Cech, T.R. (2002) *In Vivo* Selection of Better Self-Splicing Introns in *Escherichia coli*: The Role of the P1 Extension Helix of the *Tetrahymena* Intron. *RNA*, **8**, 647-658.

117. Young, B., Herschlag, D., and Cech, T.R. (1991) Mutations in a Nonconserved Sequence of the *Tetrahymena* Ribozyme Increase Activity and Specificity. *Cell,* **67,** 1007-19.

118. Alexander, R.C., Johnson, A.K., Thorpe, J.A., Gevedon, T., and Testa, S.M. (2003) Canonical Nucleosides Can Be Utilized by T4 DNA Ligase as Universal Template Bases at Ligation Junctions. *Nucl. Acids Res.*, **31**(12), 3208-3216.

Vita

Rashada Corine Alexander was born on August 26, 1977 in Birmingham, Alabama. She graduated from Coffee High School in Florence, Alabama, in May of 1995. She received a Bachelor of Science degree from Youngstown State University in Youngstown, Ohio, in June of 1999. While at YSU, she was worked as both a student tutor and a research assistant and served as the Secretary-Treasurer of Alpha Kappa Mu and the Vice-President of Omicron Lambda. In addition, she did undergraduate research in the Pipeline for Excellence and the REU programs at the University of Kentucky. She began her doctoral studies in chemistry at the University of Kentucky in August of 1999. During this time, she was awarded the Research Challenge and Trust Fund, Lyman T. Johnson, and Murrill Fellowships, as well as being chosen as a Gates Millenium Scholar. She was also awarded the 100% Plus Award, as well as the Layperson of the Year Award for the 13th Episcopal A.M.E. District and the Lyman T. Johnson Torchbearer Award. She also served as the Vice-President and President of the Chemistry Graduate Student Association, the President and Treasurer of the Black Graduate and Professional Student Association, and a student representative on the UK Graduate Council.

Publications

1. Alexander, R.C., Johnson, A., Thorpe, J., Gevedon, T., and Testa, S. M. (2003) Canonical Nucleosides Can Be Utilized by T4 DNA Ligase as Universal Bases at Ligation Junctions. *Nucleic Acids Res.,* **31***,* 3208-3216.

2. Alexander, R.C., Baum, D.A., and Testa, S.M. 5' Transcript Replacement *in vitro* Catalyzed by a Group I Intron-Derived Ribozyme (Manuscript in preparation).