BIOCHEMICAL CHARACTERIZATION OF THE RTCB FAMILY OF RNA LIGASES IN BACTERIA

A Dissertation

Presented to the Faculty of Weill Cornell Graduate School

of Medical Sciences

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

William Patrick Maughan

May 2016

BIOCHEMICAL CHARACTERIZATION OF THE RTCB FAMILY OF RNA LIGASES IN BACTERIA

William P. Maughan, Ph.D.

Cornell University 2016

All known life relies on the proper transmission of genetic information through DNA and RNA to protein. The protection of this information is critical, and rejoining of broken DNA or RNA strands by ligases is essential for maintenance of genetic integrity. The intent of this thesis is to provide a broad survey of the strategies that Nature employs to rejoin broken polynucleotide strands, with an emphasis on RNA strand ligation, finally focusing on RtcB, which accomplishes RNA ligation by a novel mechanism.

When RNA strands are broken, the fate of the phosphate at the break site dictates which enzymes must be employed to rejoin the broken ends. All ligases characterized before 2011 require a 5'-PO_4 end and a 3'-OH end for ligation. In the instance that the products of the breakage are a 3'-PO_4 or a 2',3'>PO_4 and a 5'-OH, such as in the case of enzymatic transesterification by certain endonucleases, several enzymes are needed to reconcile these ends to a 5'-PO_4 and 3'-OH before "classic" ligation may occur. This is known as the "healing and sealing" strategy. RtcB is able to bypass the need for healing by directly sealing a 2'3'>PO_4 or 3'-PO_4 end to a 5'-OH *via* a four-step mechanism in which (i) RtcB is covalently guanylylated at an active-site histidine, (ii) the guanylylated RtcB can hydrolyze a 2',3'>PO_4, producing a 3'-

PO4, (iii) RtcB transfers its GMP to the 3'-PO_4, forming $RNA_{3'}pp_{5'}G$, and (iv) RtcB catalyzes nucleophilic attack of the 5'-OH on the phosphate atom of the 3'-PO_4, sealing the RNA strand and releasing GMP.

This thesis describes the structure-guided mutagenesis of the *Escherichia coli* RtcB active site. Separation-of-function mutants parsed the contributions of specific residues in specific steps of the overall ligation reaction and prove the reaction pathway that RtcB catalyzes. This thesis also describes characterization of three RtcB paralogs from *Myxococcus xanthus*, RtcB1, RtcB2, and RtcB3. The characterization of RtcB3 shows an enzyme that, despite extremely weak ligation activity, can guanylylate either a 5'- or 3'-PO_4, and does so more efficiently on DNA than RNA.

BIOGRAPHICAL SKETCH

William Maughan received his Bachelor of Science degree in Biological Sciences from Fordham University in 2008. From 2007 to 2008 he performed research in the laboratory of Dr. Edward Dubrovsky, where he studied the tRNase Z gene in fruit flies. From 2008 to 2010 he worked as a Process Chemist at Gelest, Inc. In 2010 he joined the laboratory of Dr. Louis Levinger at York College, where he studied the tRNA-modifying enzyme Thg1. He joined Weill Cornell Graduate School of Medical Sciences in 2011 and joined the laboratory of Stewart Shuman at Sloan Kettering Institute in April of 2012.

Dedicated to my mom and dad

ACKNOWLEDGEMENTS

I would like to thank all those that made this work possible, especially my thesis mentor, Stewart Shuman, for the opportunity to grow in his lab. I was given the room to make mistakes and the guidance to learn from them. I would like to thank Minkui Luo and Christopher Lima for being on my committee, offering their sound advice, and suffering through my growing pains along with me. And I am grateful to Dirk Remus for agreeing to serve as chairman for my thesis defense.

I would like to thank all of my colleagues and mentors in the laboratory that have made this experience uniquely enjoyable. There are far too many to name each one, but those that directly mentored me deserve special recognition and gratitude. I remember being shocked by the breadth of knowledge and incisive logic that my rotation mentor, Naoko Tanaka, has. When she left the lab I inherited her bench and her project, though not her meticulous notetaking habit. I worked closely with Anupam Chakravarty for the first months after joining the Shuman lab permanently. His obvious passion for science is infectious, and he embodies one of his own favorite quotes by Linus Pauling “The best way to have a good idea is to have lots of ideas.”

I want to thank my parents for their relentless encouragement. From a young age they convinced me that education was paramount, and this thesis is the culmination of that lesson.

Finally, I'd like to thank my wife, Ella. Including any achievement listed in this thesis, without a doubt the greatest achievement of my graduate school career was convincing her to marry me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

DNA: deoxyribonucleic acid

RNA: ribonucleic acid

ATP: adenosine triphosphate

NAD^+: nicotinamide adenine dinucleotide

NTase: nucleotidyl transferase

AMP: adenosine monophosphate

PP_i: pyrophosphate

NMN: nicotinamide mononucleotide

5'-PO4: 5'-phosphate

OB: oligonucleotide binding

LigA: Ligase A

3'-OH: 3'-hydroxyl

Rnl1: RNA ligase 1

Rnl2: RNA ligase 2

Rnl3: RNA ligase 3

Rnl4: RNA ligase 4

Rnl: RNA ligase

tRNA: transfer RNA

2'-OH: 2'-hydroxyl

2',3'>PO_4: 2,'3'-cyclic phosphate

Pnkp: polynucleotide kinase/phosphatase

CPD: cyclic phosphodiesterase

*At*Rnl: *Arabidopsis thaliana* RNA ligase

Tpt1: 2'-phosphotransferase

LigT: Ligase T

ThpR: two-histidine 2',3'-cyclic phosphodiesterase acting on RNA

*Bf*Rnl: *Branchiostoma floridae* RNA ligase

Bf PNK/CPD: *Branchiostoma floridae* polynucleotide kinase/cyclic phosphodiesterase

*Hs*CNP: *Homo sapiens* cyclic nucleotide 3'-phosphodiesterase

*Hs*CLP1: *Homo sapiens* cleavage and polyadenylation factor I subunit 1

*Hs*TRPT1: *Homo sapiens* tRNA phosphotransferase 1

RtcB: RNA termincal phosphate cyclase orfB

GTP: guanosine triphosphate

HSPC117: hematopoietic stem/progenitor cell 117

RNAi: RNA interference

ExoI: *E. coli* exonuclease I

ExoIII: *E. coli* exonuclease III

PolII: Polymerase II

PolIII: Polymerase III

PolI: Polymerase I

Polβ: Polymerase β

Trl1: tRNA ligase 1

UPR: unfolded protein response

Hac1: homologous to ATF/CREB 1

mRNA: messenger RNA

ER: endoplasmic reticulum

PERK: protein kinase RNA-like ER kinase

IRE1α: inositol-requiring protein 1α

ATF6: activating transcription factor 6

ATF4: activating transcription factor 4

XBP1: X-box protein 1

PD: Parkinson's Disease

DA: dopaminergic

wt: wild type

cDNA: complementary DNA

MMP: matrix metalloprotease

TIMP2: tissue inhibitor of metalloproteinases 2

DDX1: DEAD box protein 1

ITP: inosine triphosphate

NTP: nucleotide triphosphate

ORF: open reading frame

RtcA: RNA terminal phosphate cyclase orfA

RtcR: RNA terminal phosphate cyclase orfR

GMP: guanosine monophosphate

DNAp: 3'-phosphate DNA

$_{HO}$DNA: 5'-hydroxyl DNA

pDNA: 5'-phosphate DNA

DNA$_{OH}$: 3'-hydroxyl DNA

RNAp: 3'-phosphate RNA

$_{HO}$RNA: 5'-hydroxyl RNA

pRNA: 5'-phosphate RNA

RNA$_{OH}$: 3'-hydroxyl RNA

RNA>p: RNA,2',3'-cyclic phosphate

RtcB•Mn^{2+}•GTP(αS): non-covalent complex of RtcB, Mn^{2+}, and GTP(αS)

RtcB–pG•$(Mn^{2+})_2$: covalently guanylylated RtcB in a noncovalent complex with 2 Mn^{2+}

EDTA: ethylenediaminetetraacetic acid

Tris: tris(hydroxymethyl)aminomethane

BSA: bovine serum albumin

SDS: sodium dodecyl sulfate

PAGE: polyacrylamide gel electrophoresis

DTT: dithiothreitol

CIP: calf intestinal phosphatase

NPPase: nucleotidyl pyrophosphatase

PEI: polyethyleneimine

TLC: thin layer chromatography

CMP: cytosine monophosphate

NMP: nucleotide monophosphate

CHAPTER 1

INTRODUCTION

Oligonucleotide Breakage in the Cell

The central dogma of molecular biology states that DNA encodes RNA, which encodes protein (Crick, 1970). The precise and accurate flow of information through this scheme is absolutely critical to all of life. Therefore, when a DNA or RNA strand is broken, either as part of a process that is beneficial to the organism such as recombination or replication in the case of DNA, or splicing in the case of RNA, or as the product of damage such as enzymatic restriction, it is essential that the cell is able to rejoin the broken ends in an efficient and directed manner.

Classic Ligation of Broken DNA Strands

The first discovery of a polynucleotide-joining enzyme came in 1967, when five labs simultaneously discovered DNA ligase in *Escherichia coli* (Cozzarelli et al., 1967; Gefter et al., 1967; Gellert, 1967; Olivera & Lehman, 1967; Weiss & Richardson, 1967). Following the initial discovery, DNA ligase was purified and characterized from all domains of life, with every DNA ligase discovered falling into one of two categories; ATP-dependent or NAD^+-dependent, as defined by the nucleotide cofactor that the enzyme utilizes. Members of both categories bind their cognate substrate cofactor and employ an active-site lysine of a conserved nucleotide transferase (NTase) domain to attack the adenosine monophosphate (AMP) group of the nucleotide cofactor, forming a phosphoramidate linkage between the lysine Nζ and the phosphorus atom of the AMP, with either pyrophosphate (PP_i) or nicotinamide

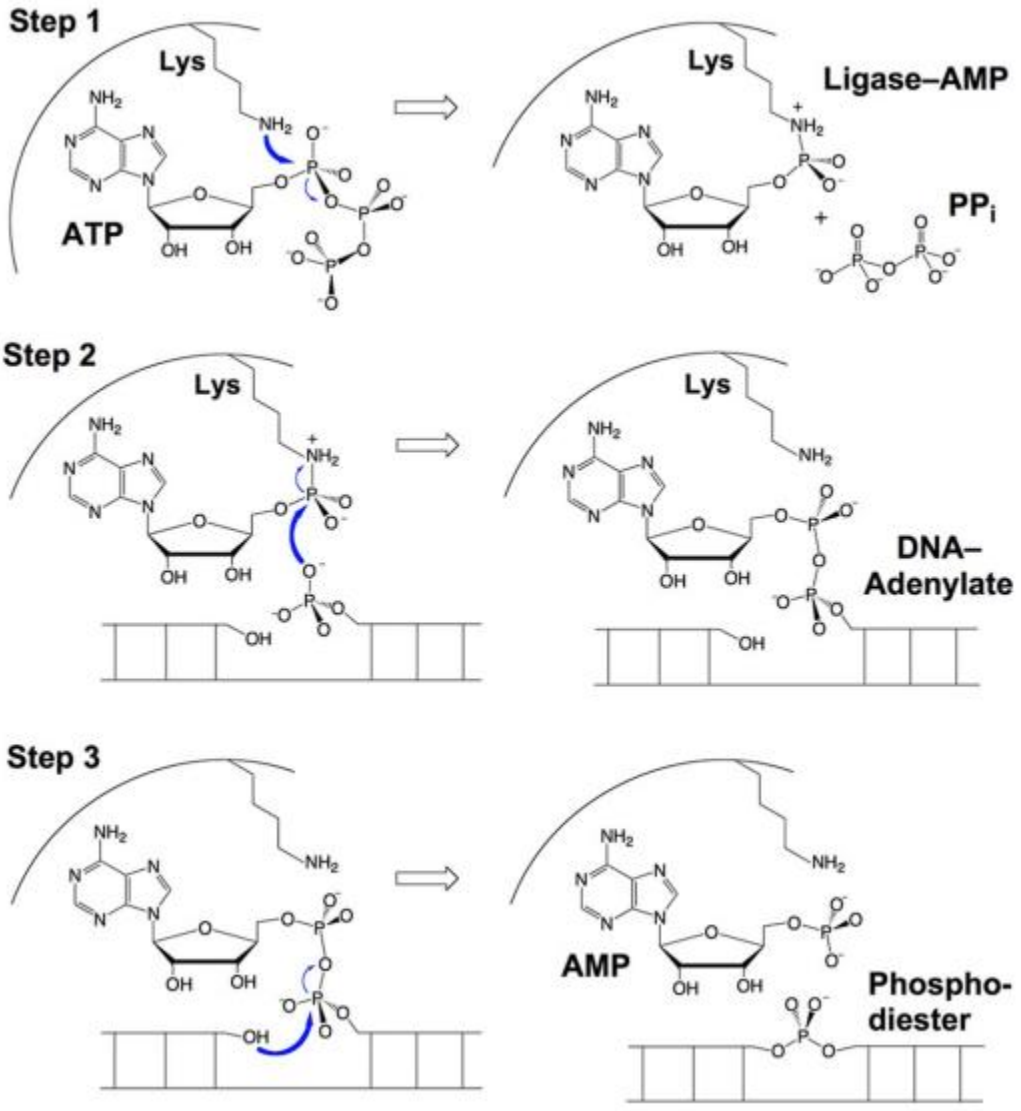


Figure 1.1 Mechanism of Nick Ligation by an ATP-Dependent DNA Ligase

In step 1, the ligase adenylylates an active-site lysine, resulting in a covalently adenylylated ligase and release of pyrophosphate (NMN is realeased in the case of NAD^+ dependent ligases). In step 2 the ligase transfers the AMP to the 5'-PO_4 on the 3'-side of the nick, creating an adenylylated DNA strand ($A_{5'}pp_{5'}DNA$). Finally, in step 3, the 3'-OH nucleophilically attacks the 5'-PO_4, sealing the DNA strand and releasing AMP (taken from Shuman, 2009).

mononucleotide (NMN) as the leaving group, respectively. The ligase then binds the DNA substrate such that the requisite 5'-PO_4 of the polynucleotide attacks the AMP phosphorus bonded to the lysine Nζ, leading to adenylylation of the DNA and release of the ligase. Finally, the 3'-hydroxyl attacks the 5'-phosphate involved in the previous adenylylation step, resulting in a 3',5'-phosphodiester linkage and release of AMP (Fig. 1.1) (reviewed in Lehman, 1974; Shuman, 2009). Different ligases adorn the NTase domain with additional domains, leading to distinct specificity or role. These are the domains that determine whether the enzyme is ATP- or NAD^+-dependent. ATP-dependent DNA ligases receive specificity from an oligonucleotide binding (OB) domain to the C-terminal side of the NTase domain (reviewed in Doherty & Suh, 2000). The OB domain folds over the NTase domain active site and provides contacts to the α-phosphorus of the ATP, aiding in enzyme adenylylation. The simplest DNA ligases (those encoded by bacteriophage T7 and *Chlorella* virus) are comprised solely of an NTase domain and an OB domain (Odell et al., 2000; Subramanya et al., 1996). NAD^+-dependent DNA ligases (aka "LigA" family ligases) have evolved a convergent mechanism that is functionally similar yet completely unrelated phylogenetically, whereby an N-terminal extension of the NTase domain closes over the active site upon NAD^+ binding, lending additional contacts to the NAD^+ ring.

Classic Ligation of Broken RNA Strands

The first RNA ligase was discovered and characterized in 1972 from T4 bacteriophage-infected *E. coli*. Analogously to DNA ligases, the RNA ligase requires ATP and Mg^{2+}, and it joins 3'-OH and 5'-PO_4 ends (Silber et al., 1972). The ligase that

they purified was later named Rnl1, one of two ligases that T4 encodes, Rnl1 and Rnl2 (Ho & Shuman, 2002), which are the eponymous members of two families of RNA ligases, the Rnl1 family and the Rnl2 family. Rnl3 was discovered in *Pyrococcus abyssi* (Brooks et al., 2008) and Rnl4 was discovered in *Clostridium thermocellum* (Smith et al., 2012). Like DNA ligases, RNA ligases employ accessory domains, though they are not entirely reliant upon accessory domains for their function. Both T4 Rnl1 and Rnl2 can perform enzyme adenylylation with the NTase domain alone, though only the Rnl1 NTase domain can perform all three ligation steps, and it does so with impaired efficiency compared to the full-length protein (Wang et al., 2007). Each Rnl family has a distinct C-terminal domain that is not found in DNA ligases or the other Rnl families. Rnl1 has a C-terminal domain that imparts specificity for tRNA with breaks in the anticodon loop (Wang, et al., 2007) while the C-terminal domain of Rnl2 provides residues essential for AMP-RNA formation (Ho et al., 2004). Rnl3 has a unique N-terminal domain and dimerizes *via* its C-terminal domain (Brooks et al., 2008). And Rnl4 has its own unique C-terminal domain as well as a module that inserts into the NTase domain (Smith et al., 2012).

The structural cause of the specificity for RNA versus DNA lies in the respective NTase domains. The binding of the RNA strand to Rnl2, specifically the terminal 2'-OH forces the ribose into an RNA-like pucker (3'-endo), whereas a 2'-H causes the 3' terminal ribose to adopt a DNA-like (2'-endo) pucker, so that upon binding of the polynucleotide to Rnl2, the 3' end is misaligned for ligation (Nandakumar et al., 2006). The facts that (i) DNA ligases require additional contacts from accessory domains to force the DNA at the nick into an RNA-like conformation

(reviewed in Pascal, 2008), and (ii) the OB domain of DNA ligases excludes polynucleotides in the RNA-like A-form on the 5' side of the break (Pascal et al., 2004; Sriskanda & Shuman, 1998), support the theory that DNA ligases evolved from an early RNA ligase, possibly solely an NTase domain (Nandakumar & Shuman, 2004).

tRNA Splicing

All domains of life encode intron-containing tRNA. Bacterial tRNA introns, while rare, are type I introns, which self-splice and therefore do not require proteins for intron excision or exon rejoining. Archaeal and eukaryal tRNAs contain introns to varying degrees, from ~6% of encoded tRNAs in humans to ~70% in some archaea (Chan & Lowe, 2015). The splicing processes for introns in these domains require endonucleolytic cleavage and rejoining performed by enzymes. The first tRNA intron discovered was in *Saccharomyces cerevisiae*, which encodes a 14-nucleotide sequence within the genomic locus that is not found in the tRNA in each of four genomic copies of $tRNA^{Tyr}$ (Goodman et al., 1977). The enzyme responsible for intron excision from tRNA is appropriately named the tRNA endonuclease, which is conserved from archaea to humans (Lykke-Andersen & Garrett, 1997), and its products are a 2',3'-cyclic phosphate ($RNA_{2',3'}$>p) and a 5'-hydroxyl ($HO_{5'}RNA$). In order to make these ends suitable for classic ligation, the 2',3'>PO_4 must be resolved to a 3'-OH, and the 5'-OH must be phosphorylated to yield a 5'-PO_4. Various taxa have evolved various mechanisms of resolving and rejoining RNA ends (aka “healing and sealing”) (Fig. 1.2). The bacteriophage T4 encodes these two activities in a single enzyme called

polynucleotide kinase/phosphatase (Pnkp) which has separate 5'-kinase and 3'-phosphatase domains to heal their respective RNA termini, leaving $RNA_{3'}OH$ and $p_{5'}RNA$. T4 Pnkp accomplishes this by sequential cyclic phosphodiesterase (CPD) and 3'-dephosphorylation reactions (Das & Shuman, 2013). Like in T4 bacteriophage, in yeast and plants the 2'3'>PO_4 is hydrolyzed but the key difference is that in this case the healing step consists solely of the CPD reaction, producing a 3'-OH and 2'-PO_4 (Greer et al., 1983; Konarska et al., 1981), and the end healing activities are present in the same polypeptide as the ligase (Englert & Beier, 2005; Sawaya et al., 2003). So instead of removing the 2'-PO_4, it remains and is necessary for ligation by the plant and yeast tRNA ligases, as illustrated by *At*RNL from *Arabidopsis thaliana*, which is deficient specifically at the step of $p_{5'}RNA$ adenylylation in the absence of the 2'-PO_4 (Remus & Shuman, 2013). The $RNA_{3'}OH$ is then linked to a $p_{5'}RNA$, yielding a 3',5'-phosphodiester bond with a monophosphate attached to the vicinal 2'-carbon (Konarska et al., 1981), which, by elimination of a 2'-OH, forms a splice junction that is resistant to repeated endonucleolytic cleavage. To restore the RNA strand to its original chemical species, the enzyme 2'-phosphotransferase (Tpt1) removes the 2'-phosphate (Culver et al., 1994; McCraith & Phizicky, 1990). Despite the fact that they do not encode tRNAs in need of enzymatically mediated splicing, *E. coli* encodes an RNA ligase named LigT that joins tRNA ends *via* an equilibrium reaction between 2',3'>p ends and 5'-OH ends, converting them to a 2',5'-phosphodiester bond and *vice versa* in a reaction that is dependent on tRNA base modification, placing the enzyme specifically in the tRNA repair pathway (Arn & Abelson, 1996). This enzyme

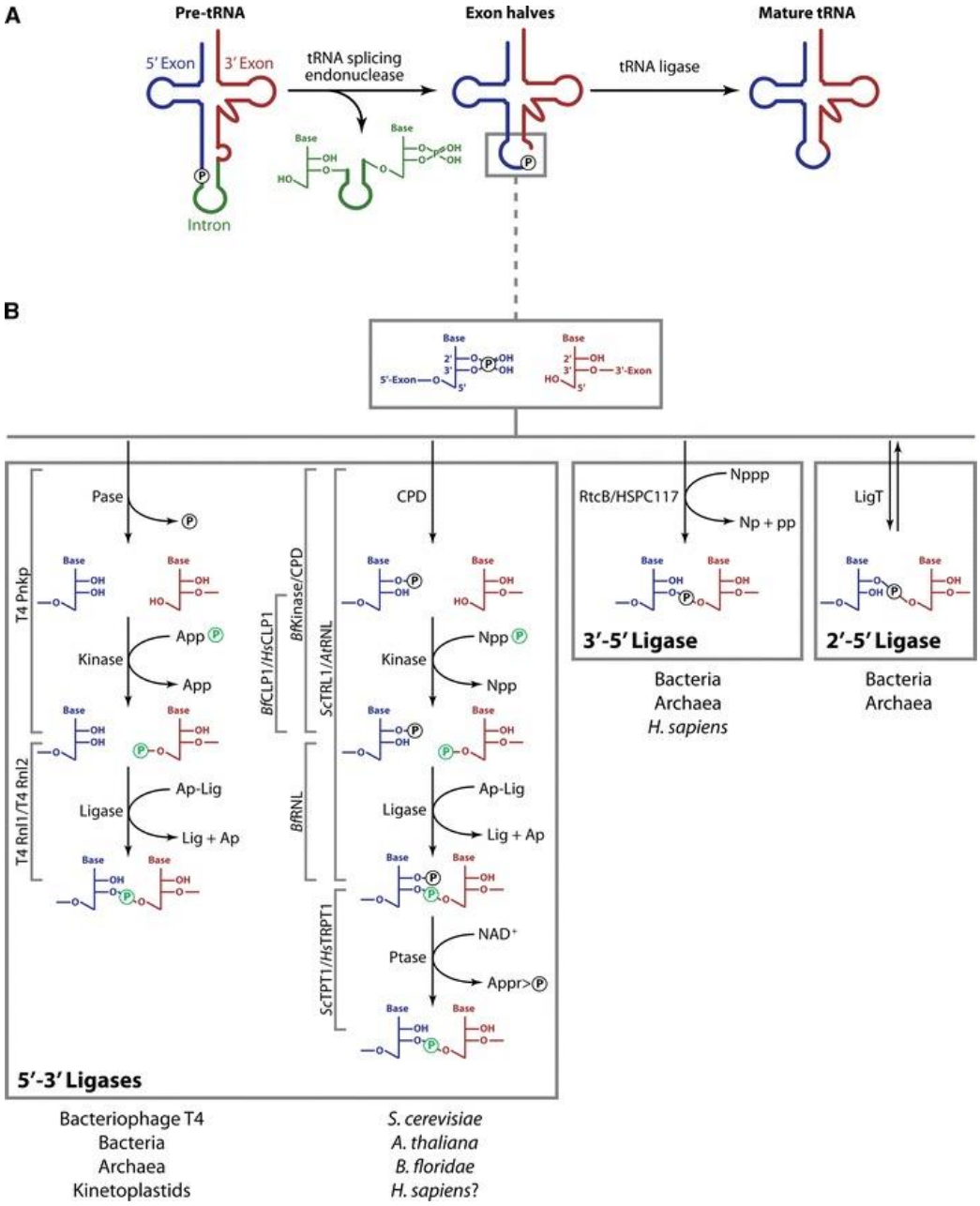


Figure 1.2 The various mechanisms of healing and sealing

A. The tRNA splicing endonuclease excises an intron from the anticodon loop of select tRNAs, leaving a 2'3'-cyclic phosphate and 5'-hydroxyl at each cleavage site. B. The various methods of resealing the product of tRNA splicing endonuclease are demonstrated, along with the enzyme responsible for each step indicated on the *left* and the organisms in which they are found indicated at the *bottom*. Reactions are grouped by mechanism of the final ligation step, where a 5'-3' ligase seals a 5'-PO_4 and a 3'-OH, a 2'-5' ligase seals a 2'PO_4 and 5'-OH, and a 3'-5' ligase seals a 5'-OH and a 2'3'>PO_4 or a 3'-PO_4. (taken from Popow et al., 2012)

was later shown to have CPD activity analogous to that of yeast and plant ligases, and based on the new activity, a rename to ThpR (two-histidine 2',3'-cyclic phosphodiesterase acting on RNA) was proposed (Remus et al., 2014). The archaeon *Pyrococcus furiosus* encodes a homologous protein (27% identity to *E. coli* ligT) that also ligates tRNA halves with 2',3'>p and 5'-OH *in vitro*, though rapid degradation of the product prevented determination of the nature of the nascent linkage (A. Kanai et al., 2009). In higher eukaryotes, current *in silico* analysis is unable to find any RNA-ligating homologs to classic RNA ligases. However, a study in the amphioxus *Branchiostoma floridae* discovered a classic RNA ligase (*Bf*Rnl) that requires a 2'-PO_4 on a 3'-OH terminus to ligate it to a 5'-PO_4. In order to heal dirty ends, *B. floridae* encodes a separate protein (*Bf* PNK/CPD) *a la* T4 Rnl1 and T4 Pnkp (Fig. 1.3). Upon discovery of *Bf*Rnl and *in silico* comparison to classic ligases, weak homology to trypanosomal Rnl was found, indicating that it is possible that current algorithms are not sensitive enough to detect metazoan homologs of Rnl family proteins (Englert et al., 2010). In humans, no functional classic ligase has been detected. However, the common accessory proteins have been purified, including CPD (*Hs*CNP) (Vogel & Thompson, 1987), kinase (*Hs*CLP1) (Weitzer & Martinez, 2007), and 2'-phosphotransferase (*Hs*TRPT1) (Spinelli et al., 1998) enzymes. Their presence combined with the demonstrated inadequacy of current algorithms to detect Rnl homologs and the detection of 5'-PO_4 ligase activity in HeLa cells and their extracts (Zillmann et al., 1991) offers encouragement that the human 5'-PO_4 ligase may exist, despite its elusiveness.

Discovery of 2'3'>P to 5'OH Ligation Activity

An alternative route to ligation that circumvents the need for accessory proteins to heal 2',3'>PO_4 and 5'-OH ends is the direct ligation of such ends, which was originally reported in 1983 (Witold Filipowicz & Shatkin, 1983). Pre-tRNA transcripts from yeast and *Xenopus laevis*, both of which contained introns, were incubated in HeLa cell extracts. The presence of two oligonucleotide species was monitored *via* gel electrophoresis of radiolabeled transcripts, with longer incubation times resulting in higher accumulation of the smaller pre-tRNA with introns excised, but still contained the 5' leader sequence and 3' trailer. Accumulation of the smaller species was accompanied by disappearance of the longer, intron-containing pre-tRNA. An inability of the extracts to fully ligate the substrates is attributed to phosphorylated 5'-ends. The next 28 years of ligase discovery and characterization failed to identify a

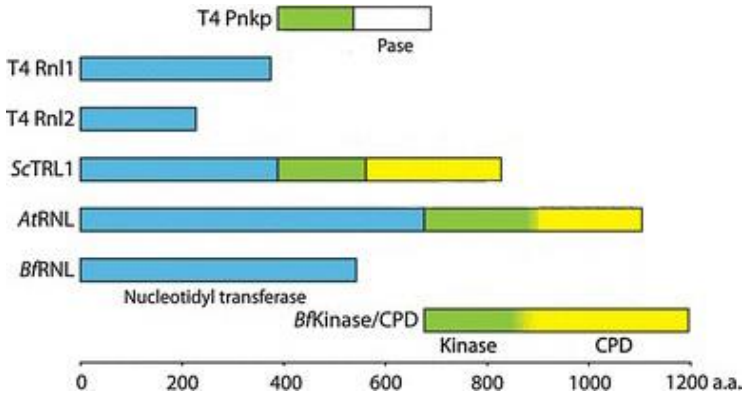


Figure 1.3 End-healing and -sealing polypeptide structure from various organisms

Polypeptides encoded by various organisms, listed at *left*, contain differing subsets of healing and sealing domains. The nucleotidyl transferase domain are colored in blue, the kinase domains in green, the cyclic phosphodiesterase (CPD) in yellow, and the phosphatase (Pase) in white. Polypeptides are depicted to scale, with the amino acid scale at *bottom*. (modified from Popow et al., 2012)

ligase that would be inhibited by 5'-PO_4, as each ligase discovered actually required such a terminus. The phosphate at the nascent phosphodiester bond was the same phosphate that was previously in the 2'3'>PO_4 terminus, indicating a "direct" ligation, as opposed to processing by a phosphodiesterase in order to produce a 3'-OH for classic ligation (Witold Filipowicz & Shatkin, 1983).

The RtcB Structure

In 2006, *Pyrococcus horikoshii* RtcB was the first member of the RtcB family to have its structure determined and published. The structure revealed no homology to known enzymes. There were, however, some clues as to its function. The structure of the apoenzyme revealed a positively charged cleft, too narrow to bind double-stranded

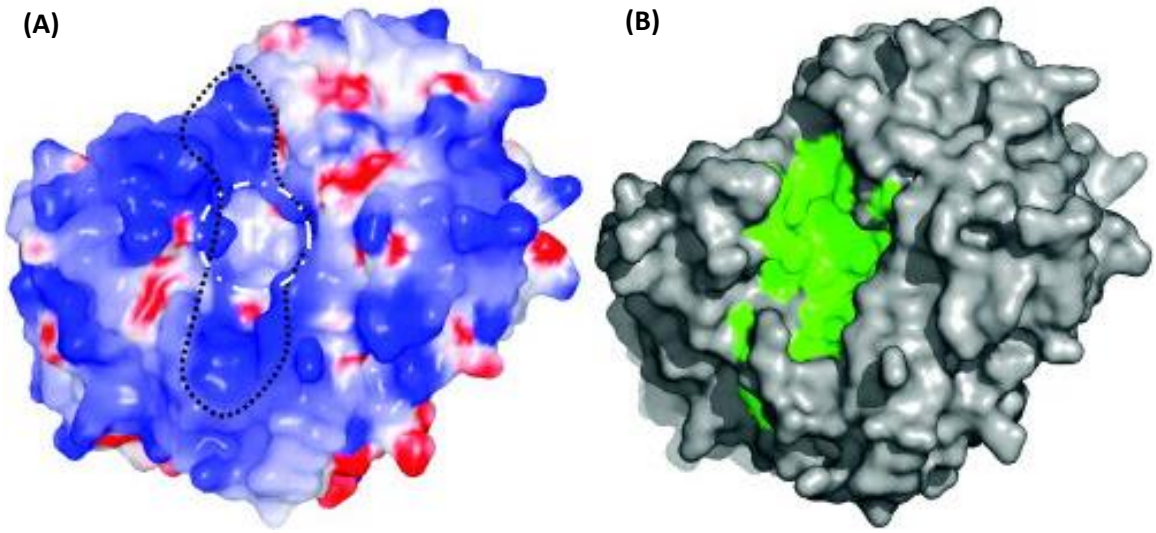


Figure 1.4 Features of the *Pyrococcus horikoshii* RtcB surface

A. The surface potential map of the *Pyrococcus horikoshii* RtcB (positive potential in blue, negative potential in red) shows a narrow cleft (highlighted by the dotted line) lined with positive potential, indicating a possible single-stranded polynucleotide binding site. At the center of the cleft is a deep pocket (circled by the white dashed line) which may indicate an active site. The green highlighted residues in B. are highly conserved that form the potential active site. (taken from Okada et al., 2006)

DNA but wide enough to accommodate a single strand, surrounding a well conserved hydrophilic pocket. A previous study (Genschik et al., 1998) suggested RtcB may be a metalloenzyme due to its conserved cysteine and histidine residues. The structure confirmed that these residues are clustered together into a conformation that could facilitate metal binding (Fig. 1.4) (Okada et al., 2006).

Identification of RtcB as the 2',3'>P, 5'OH Ligase

Five years after the structure of the apoenzyme was published, RtcB was discovered as the long-sought 2'3'>PO_4, 5'-OH ligase. Biochemical mechanistic characterization followed quickly thereafter. Three labs identified RtcB as the 2'3'>PO_4, 5'-OH direct ligase in archaea, bacteria, and human, respectively in three papers published in three consecutive months (Englert et al., 2011; Tanaka & Shuman, 2011; Popow et al., 2011).

Initial Characterization of Archaeal RtcB

In the archaeal species *Methanopyrus kandleri* the 2'3'>PO_4 ligation activity was isolated and the protein was identified by liquid chromatography and tandem mass-spectrometry of tryptic fragments. After identification of RtcB as the ligase, several archaeal RtcB's were purified and tested, with the *Pyrobaculum aerophilum* RtcB emerging as the most active of those purified. Recombinant *P. aerophilum* RtcB was produced in *E. coli*. Both isolated proteins were shown *in vitro* to ligate tRNA halves with a 2'3'>PO_4 and a 5'-OH at the splice junction. Interestingly, the archaeal ligation reaction was not stimulated by GTP or ATP. A metal was required, and Zn^{2+} was sufficient on its own to stimulate ligation (Markus Englert et al., 2011a). This is in

stark contrast to bacterial RtcB, in which not only does Zn^{2+} fail to stimulate ligation, it inhibits ligation in an otherwise catalytic mixture (Tanaka & Shuman, 2011).

Initial Characterization of Bacterial RtcB

In *E. coli*, RtcB was identified as the 2'3'>PO_4, 5'-OH ligase *via* a candidate approach. RtcB is in an operon with RtcA, which converts 3'-PO_4 ends to 2'3'>PO_4 ends. At the time of the study, it was not yet known that RtcB could ligate 3'-PO_4 ends, so it was sensible to test a protein of unknown function as a candidate to ligate the cyclic phosphate termini formed by its operonal neighbor. Like the archaeal RtcB, *E. coli* RtcB successfully ligated a substrate that mimicked a broken tRNA stem loop. In this case, the break, flanked by 2'3'>PO_4 and 5'-OH termini, was generated *via* cleavage with *Kluyveromyces lactis* γ-toxin. RtcB sealed the break in a manganese-dependent manner. *E. coli* RtcB's manganese dependence marks an interesting departure from the archaeal RtcB's metal cofactor requirements. Manganese is a soft metal, as is zinc. Zinc, however, did not sustain ligation. Interestingly, GTP stimulated the ligation but it was not strictly required. At this point it was not clear what GTP's role in ligation was. If it was not required, then perhaps it was stimulating allosterically. Or perhaps GTP was indeed strictly required, but some amount of GTP co-purified with RtcB, obscuring its necessity (Tanaka & Shuman, 2011).

Initial Characterization of the Human RtcB

The third RtcB in the initial batch of characterizations was human. Known in humans as HSPC117, RtcB was identified as a necessary component of a tRNA ligation complex. (Popow et al., 2011) set out in search of the human tRNA ligase

when they observed ligation of a broken stem-loop with 2'3'>PO_4 and 5'-OH ends upon incubation with human cell extracts. This ligation appeared to be distinct from the "healing and sealing" pathway, since removal of the 2'3'>PO_4 or phosphorylation of the 5'-OH inhibited ligation. Biochemical fractionation and tandem mass spectrometry identified 91 proteins in the active fraction. RtcB's presence in this fraction along with the previously published postulation that RtcB proteins were involved in RNA processing (Galperin & Koonin, 2004) led the authors to suspect RtcB as the ligase. RNA interference (RNAi) -induced knockdown of RtcB crippled RNA ligation, while knockdown of other proteins in the ligation complex reduced ligation to a more modest degree. Likewise, mutation of an invariant RtcB cysteine (which was later determined to bind both metals in the active site and to be strictly necessary for all catalytic RtcB functions (Tanaka et al., 2011, and below)) eliminated ligation. In contrast to archaeal and bacterial RtcBs, which were functional as monomers, human RtcB catalyzed ligation when purified along with its binding partners, but to date there are no reports of human RtcB catalyzing ligation *per se* (Popow et al., 2011).

The RtcB Chemical Mechanism

RtcB is capable of directly catalyzing the joining of a 3'-PO_4 end or a 2'3'>PO_4 end to a 5-OH end, bypassing the pathway by which a kinase and phosphoesterase convert these termini to 3'-OH and 5'-PO_4 ends ("healing" them) before sealing *via* a classic ligase. Early mechanistic characterization yielded two opposing theories regarding the reaction. Our lab made the two key discoveries that RtcB (i) readily

ligates an RNA strand with a 3'-P and (ii) will convert 2'3'>P to 3'P when a 5'-OH end is unavailable for ligation (Tanaka et al., 2011). These data suggest a mechanism whereby the cyclic phosphate precedes a 3'-PO_4 in the reaction pathway, and the 3'-PO_4 terminus is the species that is ligated. Shortly after the publication of these findings, another lab proposed the mechanism that RtcB uses GTP to activate a 3'-PO_4 by formation of a guanylylated 3'-end ($RNA_{3'}pp_{5'}G$) preceding cyclization to a 2'3'>PO_4, which is then followed by nucleophilic attack by a 5'-OH, forming a the 3',5'-phosphodiester linkage with the 2'-OH as the leaving group (Fig 1.5) (Desai & Raines, 2012). A subsequent study from our lab confirmed covalent guanylylation of the oligonucleotide strand *via* covalent guanylylation of an active site histidine, and expanded the known RtcB substrate repertoire to include DNA as a guanylate acceptor (Chakravarty et al., 2012). The role of GTP in the reaction scheme was clarified when an improved purification technique rendered RtcB completely GTP-dependent. GTP is necessary for ligation of RNA with both 2'3'>PO_4 and 3'-PO_4 ends to 5'-OH. Furthermore, the CPD activity of RtcB is dependent on covalent guanylylation of the enzyme. These two observations strongly argued in favor of the scheme in which 2'3'>PO_4 precedes 3'-PO_4 in the reaction scheme and necessarily traverses through it on the pathway to ligation (Anupam K. Chakravarty & Shuman, 2012). Finally, RtcB proceeding through a 2'3'>PO_4 as the object of attack from 5'-OH was completely disproven when it was shown that DNA, which is incapable of forming a 2'3'>PO_4, can be ligated by RtcB in the context of a broken stem loop (Das et al., 2013).

In addition to disproving the competing mechanistic model, the proof-of-concept that DNA can be ligated by RtcB introduced RtcB as a potential DNA-

modifying or even repair enzyme, the prospect of which is bolstered by the wide array of DNA damage types that generate DNAp and/or $_{HO}$DNA (Das et al., 2013; Huang & Pommier, 2013). RtcB-mediated guanylylation of a polynucleotide$_{3'}PO_4$ is independent of the preseSnce of a $HO_{5'}$polynucleotide as a ligation partner. In the event that a $DNA_{3'}p$ is guanylylated and not ligated, a stably "capped" $DNA_{3'}pp_{5'}G$ strand remains. Being a never-before observed chemical species, the fate of the capped strand was unclear with respect to enzymes that act on DNA 3' ends including exonucleases, polymerases, and phosphatases. *E. coli* exonucleases I and III resect

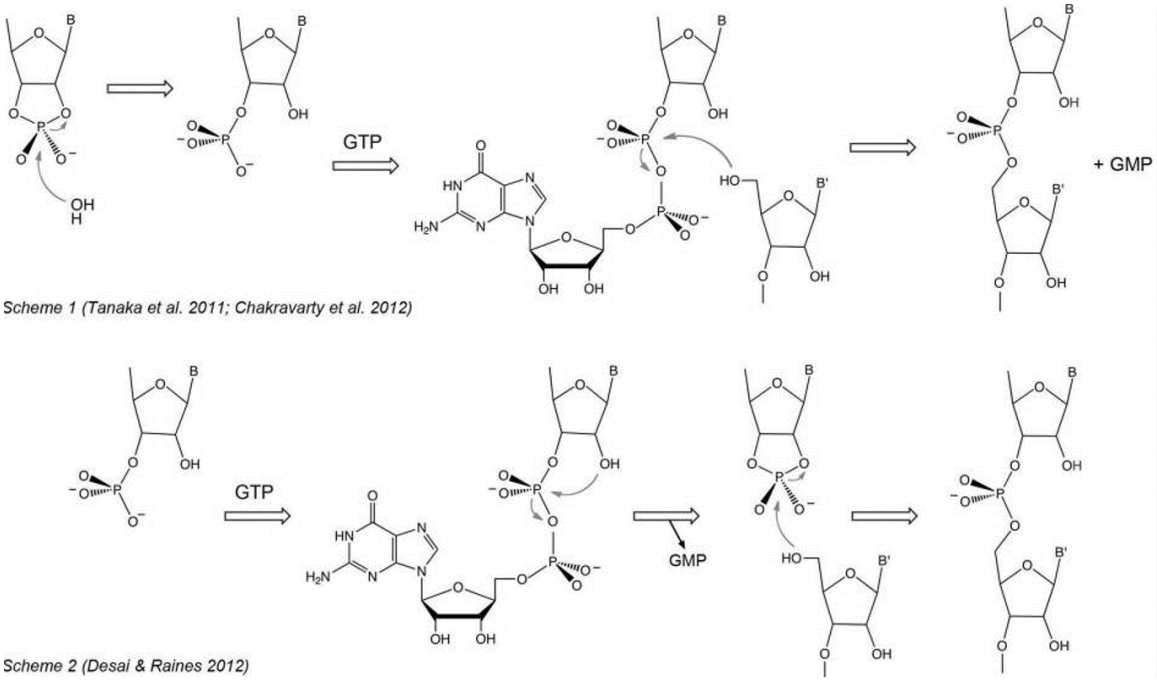


Figure 1.5 The two proposed mechanisms for RtcB-mediated ligation

Scheme 1 depicts a mechanism by which a 2'3'>PO_4 undergoes hydrolysis, producing a 3'-PO_4, which is then guanylylated, followed by nucleophilic attack by a 5'-OH, resulting in a 3',5'-phopshodiester bond. Scheme 2 depicts a proposed mechanism by which a 3'-PO_4 is guanylylated, followed by nucleophilic attack by the 2'-OH to form a 2',3'>PO_4 and sequential nucleophilic attack by a 5'-OH, forming a 3',5'-phosphodiester bond. (taken from Chakravarty & Shuman, 2012)

DNA strands from the 3' end. ExoI can trim either a $DNA_{3'}OH$ or a $DNA_{3'}p$, with preference for $DNA_{3'}OH$. ExoIII trims $DNA_{3'}OH$, and can convert $DNA_{3'}p$ to $DNA_{3'}OH$ *via* its intrinsic phosphatase activity before trimming. The $DNA_{3'}pp_{5'}G$ cap diminishes exonuclease activity by both of these *E. coli* exonucleases by ~10- to ~2500-fold, depending on the exact exonuclease and DNA 3' terminus (Das et al., 2014). The cap also prevents exonuclease activity intrinsic to *E. coli* PolII and PolIII (Chauleau et al., 2015). The cap quantitatively prevented phosphatase activity by T4 Pnkp, an exemplary 3' phosphatase. The cap structure presents a 3'OH to a DNA polymerase, which is theoretically a suitable substrate for extension, though the pyrophosphate preceding the terminal guanosine could interfere with the DNA base pairing or binding site recognition by polymerases. The *E. coli* polymerase I (PolI) Klenow fragment (a truncation that does not possess the inherent PolI 3' exonuclease activity) is indeed able to extend a $DNA_{3'}pp_{5'}G$ strand with only a modest reduction in activity, if and only if the terminal guanosine is opposite a cytosine, preserving base-pairing as if the pyrophosphate group were the familiar singular phosphate. The ability to extend the $DNA_{3'}pp_{5'}G$ cap also pertained to *Mycobacterium smegmatis* DinB1, a polymerase in the Y family (Das et al., 2014), *E. coli* Pol II (a member of the B family), *E. coli* PolIII (a member of the C family), and human Polβ (a member of the X family), though Polβ only added a single templated nucleotide (Chauleau et al., 2015). Finally, the enzyme aprataxin, which is known to remove the adenylate moiety from an $A_{5'}pp_{5'}DNA$ (the product of aborted ligation *via* classic ligase) leaving behind $p_{5'}DNA$ (Ahel et al., 2006; Tumbale et al., 2014), can also remove the guanylate moiety from $DNA_{3'}pp_{5'}G$, leaving $DNA_{3'}p$ (Das et al., 2014).

The RtcB Active Site

Having characterized the RtcB reaction mechanism, we set out to discover how RtcB accomplishes catalysis of these reactions. Several organisms express enzymes with multiple domains and active sites to carry out different reactions along the ligation pathway (Fig. 1.3), but RtcB only has one domain with a single predicted active site. So are the RtcB-mediated reactions all carried out within a single active site? If so, are the reaction steps discrete, with individual residues participating in specific subsets of reactions, or is it a concerted mechanism, with each individual step contingent on the integrity of the entire active site? In the following pages, the mutagenesis of the RtcB active site and the effects of individual residues on individual reaction steps are detailed.

RtcB Can Complement a Yeast "Classic" Ligase

Saccharomyces cerevisiae, like the preponderance of fungi, does not encode RtcB despite the presence of intron-containing tRNAs in the transcriptome. In order to rejoin tRNA halves after intron excision, yeast rely on the classic ligase Trl1 (Greer et al., 1983; Phizicky et al., 1986). For Trl1 to ligate, however, the 2'3'>PO_4 and a 5'-OH generated during intron excision must be "healed". A phosphoesterase and a kinase remove the 2'3'>PO_4 and phosphorylate the 5'-end respectively, and Trl1 joins the exons via a classic ligation reaction. To test if RtcB can truly bypass the "healing and sealing" pathway a plasmid shuffle assay was performed to test whether RtcB can rescue a *trl1*Δ knockout, which is normally lethal due to the cell's inability to splice tRNA. Indeed, the *RTCB trl1*Δ cells were viable (Tanaka et al., 2011). *RTCB trl1*Δ

cells were also protected from the otherwise lethal expression of anticodon endonuclease *Pichea acaciae* toxin (Tanaka et al., 2011), which cleaves within the anticodon of $tRNA^{Gln}_{(UUG)}$, leaving a 2',3'>PO_4 and a 5'-OH (Klassen et al., 2008), showing that RtcB is a true RNA repair enzyme. To expand the known repair substrate repertoire of RtcB beyond tRNA, the authors turned to the unfolded protein response (UPR). The yeast UPR relies on expression of Hac1, a transcriptional regulator that is responsible for the expression of several protein chaperones (Cox & Walter, 1996). The *HAC1* pre-mRNA contains an intron that is spliced in response to stress brought on by the endoplasmic reticulum (ER) being overburdened with nascent proteins (Cox & Walter, 1996). Upon knocking out TRL1, the yeast cell cannot properly splice HAC1 mRNA, and the UPR is hamstrung (Sidrauski et al., 2016). In another plasmid shuffle assay, *S. cerevisiae trl1Δ* cells that expressed RtcB were able to quantitatively splice HAC1 pre-mRNA (Tanaka et al., 2011). A homologous pathway exists in metazoa, and at the time of the *S. cerevisiae* study, the metazoan UPR ligase remained unknown. Furthermore, there was a dearth of candidates, as RNA ligases in lower organisms (LigT, *At*RNL, and TRL1) are not encoded by metazoa (Markus Englert & Beier, 2005; Greer et al., 1983; Silber et al., 1972).

RtcB in the Metazoan Unfolded Protein Response and tRNA Splicing

Three years later, RtcB's role in the UPR of higher organisms was affirmed by three separate labs (Jurkin et al., 2014; Kosmaczewski et al., 2014; Lu et al., 2014). The metazoan UPR is more complex than the fungal version, being comprised of three separate pathways. Three separate ER stress sensors; protein kinase RNA-like ER

kinase (PERK), inositol-requiring protein 1α (IRE1α), and activating transcription factor 6 (ATF6) are transmembrane proteins that reside in the ER membrane. Upon ER stress, all three sensors are triggered and an array of responses results, including degradation of specific mRNAs, enhanced protein folding quality control, ER biogenesis, inhibition of translation, and autophagy. Each sensor also activates a downstream effector (ATF4, XBP1s, and ATF6f, respectively) that travels to the nucleus and upregulates UPR target genes (reviewed in Hetz, 2012). IRE1α is an endonuclease that is inactive in non-stress conditions. Upon ER stress, IRE1α dimerizes and autotransphosphorylates, thereby becoming active (Shamu & Walter, 1996). IRE1α then cleaves the mammalian homolog of HAC1, unspliced X-box protein mRNA (XBP1u), at two neighboring stem loops, resulting in the excision of an intron, leaving behind a 2'3'>PO_4 and a 5'-OH (Gonzalez et al., 1999). RtcB is responsible for rejoining these ends to produce the spliced x-box protein (XBP1s) mRNA (Fig. 1.6), which is then translated, travels to the nucleus, and upregulates several genes in the UPR (Lee et al., 2003). In *Caenorhabditis elegans*, knocking out the lone RtcB homolog result in inviability in conditions of ER stress, strengthening the hypothesis that *E. coli* RtcB resides in a stress-induced operon. Like archaea, *C. elegans* also relies on RtcB to splice tRNA halves together, as knocking out RtcB in *C. elegans* caused accumulation of unspliced tRNA halves. The knockout strain also displayed slow growth and reduced lifespan. Exogenous expression of intact tRNAs rescue the growth and lifespan phenoytpes. Finally, RtcB-null *C. elegans* also displayed abnormalities in vulval development which were not rescued by expression of pre-spliced tRNA nor were they induced by abrogation of the UPR by knocking out

IRE1 or XBP1, suggesting additional *in vivo* roles of RtcB in addition to tRNA and XBP1 splicing. The mechanism by which RtcB protects normal vulval development remains unknown (Kosmaczewski et al., 2014).

In a separate study in *C. elegans*, RtcB was implicated as a potential protector against Parkinson's Disease (PD) *via* its XBP1 splicing activity. PD is characterized in part by accumulation of the protein α-synuclein in structures termed “Lewy bodies”. Mutation of α-synuclein or multiplication of its genomic locus can induce Lewy body formation (reviewed in Uversky, 2007). An RNAi screen was used to identify genes that protect against dopaminergic (DA) neuron degeneration induced by human α-synuclein. In a model system in which human α-synuclein is expressed under a DA neuron-specific promoter, neurodegeneration occurs in an age-dependent and α-synuclein-dose-dependent manner. Overexpression of wt *rtcb-1* cDNA rescued the age-dependent degeneration, while targeted RNAi knockdown of *rtcb-1* exacerbated the phenotype. To demonstrate that this result is due to RtcB's action on the UPR, XBP-1 was knocked down via RNAi and the neuroprotective phenotype of RtcB overexpression was rendered ineffective (Ray et al., 2014).

RtcB in the Human Unfolded Protein Response

With mounting evidence that RtcB is a key part of the UPR in metazoa, its known phylogenetic span was expanded when it was shown to be the primary XBP1 mRNA ligase in the human UPR. This discovery proved elusive despite evidence of physical association of human RtcB with XBP1 mRNA (Baltz et al., 2012). An initial attempt to test RtcB as the human XBP1 mRNA ligase by targeted RtcB RNAi

knockdown was unable to implicate it, while yeast TRL1, a classic ligase that splices HAC1 (the yeast XBP1 homolog) transcripts *in vivo*, but has no mammalian homolog, spliced human XBP1 transcripts together *in vitro* (Iwawaki & Tokuda, 2011). A subsequent study suggests that incomplete RNAi knockdown may explain the inability of the technique to implicate RtcB, or that a partially redundant ligase, which may catalyze ligation of "classic ligase" cognate 5'-PO_4 and 3'-OH ends, compensates for an RtcB knockout. In this study, the authors created a synthetic biology system in HEK293 cells in which XBP1 mRNA splicing is coupled to a pro-apoptotic phenotype upon acute ER stress. An RNAi screen was used to identify genes involved in the engineered pro-apoptotic phenotype. This approach worked, implicating RtcB in XBP1 splicing, as knockdown of RtcB produced viable cells upon ER stress, indicating that RtcB is necessary for the synthetic apoptotic pathway. The authors then produced a homozygous RtcB conditional knockout line of mouse cells. Consistent with findings from *C. elegans*, homozygous knockout slowed cell growth and increased levels of unspliced tRNA (Lu et al., 2014).

RtcB in Mammalian Embryonic Development

Further *in vivo* characterization of mammalian RtcB assigns it a crucial role in placental development. Somatic cell nuclear transfer is a cloning technique in which the nucleus of an oocyte is removed and replaced with a nucleus from a somatic cell from another organism. The oocyte is then implanted into a host mother and grown to adulthood. This process has a low success rate (<5% in mice) (Wakayama, 2007). One likely reason for the low success rate is that the gene expression of the resultant

embryos differs significantly from that of a normal embryo. Among the differentially expressed genes is RtcB, which is underexpressed in cloned monkey and mouse embryos in comparison to their *in vivo*-produced counterparts. This observation led to a study in which RtcB was knocked down *via* RNAi in mouse embryos. The resultant embryos show no defect in uterine implantation, but rapidly degenerate in late post-implantation stage, with few embryos developed to term. This result closely mimicked the fate of oocytes generated by somatic cell nuclear transfer (Wang et al., 2010). To investigate the role of RtcB in placental development, RtcB was overexpressed in JEG-3 cells, a human placental choriocarcinoma line (choriocarcinomas are derived from the fetus). RtcB overexpression reduced expression of mRNA for two matrix metalloproteases (MMPs), which improve cell motility, and increased expression of a specific MMP inhibitor called tissue inhibitor of metalloproteinases 2 (TIMP2). As one would expect from these changes in gene expression, the treated JEG-3 cells showed slower migration in response to a scratch wound assay (Ma et al., 2014).

RtcB in the Mouse Neuron

In neurons isolated from mouse brains, RtcB is present in a large protein granule (~1000S) that shuttles mRNA through the dendrites of neurons bidirectionally, with its distal movement correlated with expression levels of kinesin. The granule contained 42 proteins involved processes such as RNA transport, protein synthesis, and translational silencing. None of the mRNAs included in the granule have yet been identified as RtcB substrates, so RtcB's function in this granule remains unknown. However, the two most obvious hypotheses are that it is ligating (i) tRNA, either as

part of the splicing pathway or in response to damage, or (ii) the mRNAs present in the granules, again, either in response to programmed splicing or potentially deleterious damage (Kanai et al., 2004). Three DEAD box helicases were also present in the RNA-protein cluster, one of which (DDX1) was later shown through analysis of clusters of orthologous groups to associate with RtcB genetically. Popow *et al.* specifically identify DDX1 and archease as cofactors of RtcB. DEAD box helicases couple ATP hydrolysis to double-stranded oligonucleotide unwinding. ATP is unable to act as a nucleotide cofactor for RtcB ligation, nonetheless, addition of ATP was able to stimulate RtcB guanylylation and ligation in HeLa cell extracts. Both effects were reduced upon mutation of ATP-binding and -hydrolysis motifs of DDX1, implicating DDX1 as a stimulator of RtcB (Popow et al., 2014). To date, this study remains the sole evidence of a link between DDX1 and RtcB. Archease, however, has several further studies demonstrating its partnership with RtcB.

The Association of RtcB and Archease

Archease is a known cofactor of RtcB. It is found clustered with RtcB in the genomes of numerous organisms and it is notably absent in fungi, the lone kingdom that does not encode RtcB. The molecular mechanism of archease remains a mystery. We have, however, gained insight into its various abilities in recent publications. Its first discovered activity was increasing the specific activity of a tRNA methyltransferase (Auxilien et al., 2007). It was later shown to increase the activity of RtcB in human, *Pyrococcus horikoshii,* and *Thermus thermophilus* (Desai et al., 2015; Desai et al., 2014; Popow et al., 2014). Its most baffling activity was observed in

Pyrococcus horikoshii, wherein archease induces RtcB to relax its GTP specificity and accept ATP or inosine triphosphate (ITP) (Desai et al., 2014a). The mechanism for this remains unclear, though the most likely explanation seems to be allosteric conformational change on the NTP binding pocket upon archease binding. Further confounding these results, archease was not seen to have this nucleotide specificity liberalization effect on RtcB from another thermophilic species (Desai et al., 2015).

The E. coli Rtc Operon

The *Escherichia coli* operon in which RtcB exists encodes three proteins (Fig. 1.7). RtcR is a σ^{54}-dependent regulator, and is upstream of RtcB, transcribed in the opposite direction. Sigma factors bind DNA and recruit RNA polymerase. σ^{70}-like factors comprise a large family of diverse proteins that regulate transcription of many families of genes, including housekeeping genes. They bind the DNA at the -10 and -35 regions and trap the DNA in a melted state conducive to transcription. In contrast, the σ^{54} "family" contains a single protein. It requires an activator to bind at a site 80 to 150 basepairs upstream of the promoter to which σ^{54} is bound. The activator couples ATP hydrolysis to DNA melting and RNA polymerase transcription (Bush & Dixon, 2012). The region upstream of the RtcA/RtcB mRNA transcription start site contains sequence elements at -12 and -24 that are characteristic of promoters that are recognized by σ^{54}. RtcR contains all the elements of a σ^{54}-dependent regulator, the activator that binds upstream of the σ^{54}-promoter, hydrolyzes ATP, and interacts with σ^{54}. Indeed, deletion of the N-terminal domain, which is a sensory domain that represses ATP hydrolysis and consequent σ^{54} activation, and overexpression of RtcR

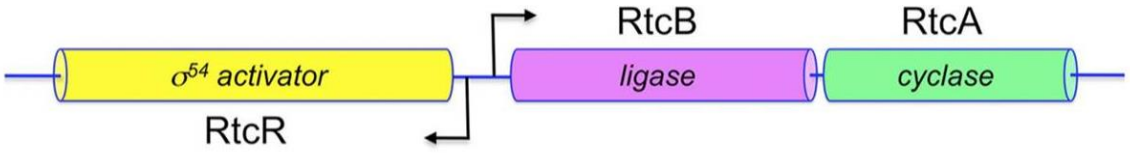


Figure 1.7 The *E. coli* Rtc operon

The arrangement and orientation of the genes of the *E. coli* Rtc operon is depicted. Arrows indicate direction of translation. (taken from Tanaka & Shuman, 2011)

were shown to increase the levels of RtcA and RtcB transcripts. This effect was not seen with expression of the wild type RtcR. Knocking out σ^{54} also eliminated this effect, demonstrating that RtcR acts in a σ^{54}-dependent manor (Genschik et al., 1998).

E. coli RtcA

RNA terminal phosphate cyclase orfA (RtcA), the protein from which the operon derives its name, is encoded downstream of RtcB and acts on an RNA strand with a 3'-terminal phosphate ($RNA_{3'}p$), converting it to a 2'3'>PO_4. RtcA accomplishes this by: first reacting with ATP, adenylylating an active-site histidine and releasing PPi; then transferring the AMP moiety to the RNA terminal phosphate; and finally allowing the vicinal 2'-hydroxyl group to nucleophilically attack the original terminal phosphate, creating a 2'3'>PO_4 and releasing AMP (Genschik et al., 1998). After initial characterization of RtcB's and RtcA's activities they appeared

to have antagonistic activities. As operons usually encode proteins that function toward the same end result, RtcB's and RtcA's co-expression would appear counterproductive. Step two of the four-step RtcB reaction mechanism is the conversion of a 2'3'>PO_4 to a 3'-PO_4. This is the exact opposite of the overall RtcA

reaction with respect to modification of the RNA strand. (Das & Shuman, 2013) suggested a possible role of RtcA and RtcB to work coopertaively when they showed that RtcA is also capable of cyclizing a 2'-PO_4. This reaction returns a chemical species that is off the RtcB chemical reaction pathway back onto it, making it a substrate for ligation.

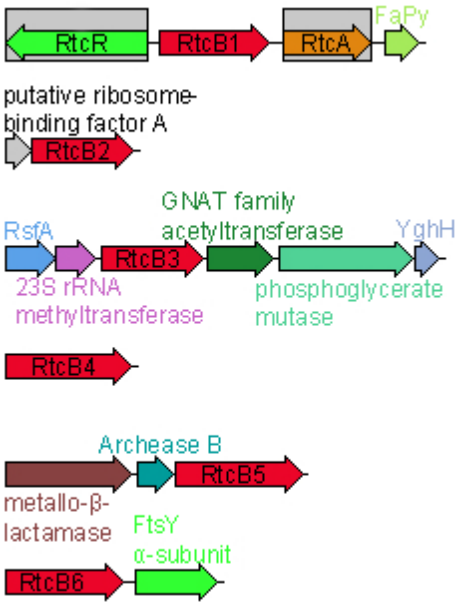


Figure 1.8 The genomic arrangement of *M. xanthus* RtcB ORFs

Each of the six RtcBs encoded by *Myxococcus xanthus* is shown in their genomic position. Putative RtcB ORFs are shown in red. The putative identity of each gene shown is above or below in its corresponding color. RtcR and RtcA, which are in the same orientation to RtcB1 as they are to *E. coli* RtcB are highlighted in gray.

Myxococcus xanthus encodes six paralogs of RtcB

In order to probe the versatility of the RtcB ligase family repertoire, I turned to the social bacterium *Myxococcus xanthus*, as detailed below. The *M. xanthus* genome shows evidence of major gene duplication. This duplication was not random, instead, genes involved in "cell-cell signaling, small molecule sensing, and multicomponent transcriptional control were amplified preferentially" (Goldman et al., 2006). RtcB is among the duplicated genes, with six ORFs showing sequence homology to the RtcB family. These RtcB ORFs appear in six separate genomic contexts, including RtcB1,

which is in the same arrangement as the *E. coli* Rtc operon (Fig. 1.8). The genomic neighbors of the individual RtcBs did not offer obvious clues as to what each RtcB's function is, so we set out to characterize them individually *in vitro*.

CHAPTER 2

DISTINCT CONTRIBUTIONS OF ENZYMIC FUNCTIONAL GROUPS TO THE 2',3'-CYCLIC PHOSPHODIESTERASE, 3'-PHOSPHATE GUANYLYLATION, AND 3'-ppG/5'-OH LIGATION STEPS OF THE ESCHERICHIA COLI RTCB NUCLEIC ACID SPLICING PATHWAY

INTRODUCTION

Escherichia coli RtcB exemplifies a novel family of RNA ligases implicated in tRNA splicing and RNA repair (Chakravarty et al., 2012; Chakravarty & Shuman, 2012; Englert et al., 2011b; Popow et al., 2011a; Tanaka et al., 2011; Tanaka et al., 2011b; Naoko Tanaka & Shuman, 2011b). Unlike classic RNA and DNA ligases, which join 3'-OH and 5'-phosphate ends, RtcB seals broken RNAs with 5'-OH and either 2',3'-cyclic phosphate (RNA>p) or 3'-phosphate (RNAp) ends. *E. coli* RtcB executes a multi-step ligation pathway (Chakravarty et al., 2012; Chakravarty & Shuman, 2012; Tanaka et al., 2011) entailing: (i) reaction of the enzyme with GTP to form a covalent RtcB-(His^{337}-N)-GMP intermediate; (ii) hydrolysis of RNA>p to RNAp; (iii) transfer of guanylate from His337 to the polynucleotide 3'-phosphate to form a polynucleotide-(3')pp(5')G intermediate; and (iv) attack of a 5'-OH on the –NppG end to form the splice junction and liberate GMP (Fig. 2.1). The *E. coli* RtcB reaction pathway requires manganese as a divalent cation cofactor. The catalytic repertoire of *E. coli* RtcB is not limited to RNA transactions. RtcB efficiently “caps” DNA 3'-phosphate (DNAp) ends

to form a stable DNAppG structure (Das et al., 2013). Moreover, RtcB can ligate DNAp and $_{HO}$DNA ends within a stem-loop structure (Das et al., 2013).

The wide phylogenetic distribution of RtcB enzymes in bacteria, bacteriophages, archaea, and metazoa, and the inauguration by RtcB of a novel enzymology based on 3' end activation, prompts us here to interrogate the enzymic requirements for several of the component steps. We aim to distinguish between models in which: (i) all pathway steps are performed by a full ensemble of active site functional groups and cofactors; or (ii) each step is driven by a subset of active site constituents. In principle, this issue can be addressed by identifying RtcB changes that cripple the end joining pathway and then testing them for separations-of-function, i.e., whether such mutations selectively affect one or more of the of chemical steps while sparing others. The utility of this approach was exemplified by an analysis of the H337A mutant of *E. coli* RtcB, which eliminates the site of covalent attachment of GMP to the enzyme, rendering the H337A protein inert in RtcB guanylylation and hence formation of the polynucleotide-ppG intermediate (Chakravarty et al., 2012). The instructive findings were that H337A was adept at joining DNAppG and $_{HO}$DNA ends in a reaction that required Mn^{2+}, but not GTP (Das et al., 2013), signifying that GTP and His337 play no essential role in the pathway downstream of the step of 3' end activation.

Although the biochemistry of RtcB is best understood for the *E. coli* enzyme, available structural insights stem from *Pyrococcus horikoshii* RtcB, which has been crystallized in various functional states: as the RtcB apoenzyme; as RtcB complexes with Mn^{2+} or Mn^{2+}•GTP(αS); and as the covalent RtcB–pG•Mn^{2+} intermediate (Desai,

et al., 2013; Englert et al., 2012; Okada et al., 2006). A key insight from the crystal structures was that RtcB binds two Mn^{2+} ions, separated by 3.6 Å, one of which (Mn1 in Fig. 2.2) coordinates the α phosphate of GTP (and the GMP phosphate in the covalent RtcB–pG complex) while the other (Mn2 in Fig. 2.2) contacts the GTP γ phosphate (11). Both metal ions also contact the β-γ bridging oxygen of GTP. For convenience, we have labeled the conserved amino acids in the archaeal RtcB structure in Fig. 2.2 according to their positions in *E. coli* RtcB. Mn1 is coordinated by a cysteine and two histidines (Cys78, His185, and His281 in *E. coli* RtcB); Mn2 is coordinated by aspartate, cysteine, and histidine side chains (Asp75, Cys78, and His168 in *E. coli* RtcB). The cysteine uniquely bridges the two metal ions (Fig. 2.2). The predominance of "soft" metal ligands (histidine nitrogens and cysteine sulfur) in the active site explains the observed metal activity and inhibition spectrum of *E. coli* RtcB (Das et al., 2013; Tanaka et al., 2011; Tanaka & Shuman, 2011b), whereby: (i) the overall ligation pathway and the components step of RtcB guanylylation and phosphodiester synthesis rely on Mn^{2+}, which is favorably coordinated by soft ligands; (ii) "hard" metals Mg^{2+} and Ca^{2+}, which favor oxygen ligands, neither support RtcB activity themselves nor inhibit RtcB function in the presence of Mn^{2+}; and (iii) alternative soft metals Zn^{2+}, Co^{2+}, Cu^{2+}, and Ni^{2+} that do not support activity are potent inhibitors of activity in the presence of Mn^{2+}, presumably because they compete with Mn^{2+} for binding to one or both metal sites, whence bound they are unable to sustain catalysis. It is an open question whether both Mn^{2+} sites in the binuclear cluster are required for each step in the RtcB pathway or if some of the steps rely on just one of the metal ions.

In the *Pyrococcus* RtcB•Mn^{2+}•GTP(αS) structure, the His404-Nε nucleophile (His337 in *E. coli* RtcB) is situated 3.3 Å from the GTP α phosphorus in a favorable apical orientation (168˚) to the bridging oxygen of the pyrophosphate leaving group, signifying that the structure plausibly mimics the Michaelis complex of the RtcB guanylylation reaction (Desai et al., 2013). The structure highlights additional active site amino acids that contact GTP or that contact nearby sulfate anions, which might mimic the phosphates at and flanking the polynucleotide ends that RtcB joins.

In a preliminary study (Tanaka et al., 2011b), antedating the elucidation of the multistep RtcB pathway, we had tested the effects of a set of alanine mutations of *E. coli* RtcB, at positions chosen based on their phylogenetic conservation and spatial proximity in the then available apoenzyme structure of *Pyrococcus* RtcB. The mutated set included the side chains depicted in Fig. 2.2 (except His281, which was mutated presently) plus two amino acids not shown in Fig. 2.2 (Arg341, a conserved residue; and His280, a non-conserved residue corresponding to Ala328 in *Pyrococcus* RtcB). Initial tests of sealing of a broken tRNA-like stem-loop with 2',3'-cyclic phosphate and 5'-OH RNA ends at a single level of input recombinant enzyme showed that: (i) H280A retained full sealing activity; (ii) R345A, N167A, H185A, R189A and K299A had low activity; and (iii) D75A, C78A, H168A, H337A and R341A had no activity under the conditions tested (Tanaka et al., 2011b). Since then, our laboratory has improved the purification of recombinant RtcB protein, such that it is fully dependent on exogenous GTP for sealing activity (Tanaka et al., 2011). Moreover, we have developed new substrates to assay the cyclic phosphodiesterase (CPDase), 3'-PO_4/5'-OH ligase, and phosphodiester synthesis reactions under single turnover conditions (Chakravarty et al.,

2012; Chakravarty & Shuman, 2012; Das et al., 2013; Tanaka et al., 2011). Here we applied these interval technical advances to re-examine the effects of active site alanine mutations on *E. coli* RtcB function *in vitro*. Our results identify separation-of-function mutations and step-specific defects. We discuss our findings in light of the available crystal structures.

MATERIALS AND METHODS

RtcB purification. Wild-type RtcB and RtcB-Ala mutants were produced in *E. coli* as His_{10}Smt3-RtcB fusions and isolated from soluble lysates by nickel-agarose chromatography as described (Tanaka et al., 2011). The tags were excised by treatment with the Smt3-specific protease Ulp1. The tag-free RtcB proteins were separated from His_{10}Smt3 by a second round of nickel-agarose chromatography, during which RtcB was recovered in the flow-through fractions, which were immediately adjusted to 20 mM EDTA and incubated for 1 h (Tanaka et al., 2011) before dialysis overnight against 10 mM Tris-HCl, pH 8.0, 350 mM NaCl, 1 mM DTT, 1 mM EDTA. The RtcB preparations were stored at –80˚C. Protein concentrations were determined by using the BioRad dye reagent with BSA as the standard.

RNA and DNA substrates. The 20-mer $_{HO}$RNAp oligonucleotide labeled with ^{32}P at the penultimate phosphate was prepared by T4 Rnl1-mediated addition of 5'-[^{32}P]pCp to a 19-mer synthetic oligoribonucleotide as described (Tanaka et al., 2011). The $_{HO}$RNAp was treated with *E. coli* RNA 3'-terminal phosphate cyclase (RtcA) and ATP to generate a 2',3'-cyclic phosphate derivative, $_{HO}$RNA>p (Tanaka et al., 2011). The $_{HO}$RNAp and $_{HO}$RNA>p substrates were gel-purified prior to use in RtcB assays. The 12-mer DNAp oligonucleotide was 5'-phosphorylated by reaction with T4 Pnkp-D167N and [$\gamma^{32}P$]ATP to form pDNAp, and then gel-purified. The pre-guanylylated pDNAppG strand was prepared by reaction with RtcB, Mn^{2+}, and GTP, then gel-purified and annealed to a complementary cold $_{HO}$DNA strand to form a DNA stem-loop as described previously (Das et al., 2013).

RtcB activity assays. Reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM radiolabeled nucleic acid substrate, and 1 µM RtcB were incubated at 37°C. The reactions were initiated by adding RtcB and quenched at the times specified by adding an equal volume of 90% formamide, 50 mM EDTA. Alternatively, in reactions with $_{HO}$RNA>p substrate, the reactions were quenched by adding 0.5 volume of 40 mM EDTA, followed by digestion for 15 min at 37˚C with 1000 U RNase T1 (Fermentas). The RNase T1 digests were then mixed with an equal volume of formamide, EDTA. The reaction products were analyzed by electrophoresis (at 50 W constant power) through a 40-cm 20% polyacrylamide gel containing 8 M urea in 45 mM Tris borate, 1.2 mM EDTA. The ^{32}P-labeled nucleic acids were visualized by autoradiography of the gel and, where specified, quantified by scanning the gel with a Fuji Film BAS-2500 imager.

RESULTS

Effects of active site mutations on HORNAp ligation. Wild-type RtcB and eleven Ala-mutants were produced in *E. coli* as His_{10}Smt3 fusions and isolated from bacterial soluble extracts by adsorption to Ni-agarose and elution with imidazole. After cleavage of the tag with Smt3 protease Ulp1, the RtcB proteins were separated from His_{10}Smt3 by a second round of Ni-agarose chromatography. SDS-PAGE of the protein preparations showed that the RtcB polypeptide was the major species in each case (Fig. 2.3).

To initially gauge mutational effects on RNA ligation activity, the wild-type RtcB and RtcB-Ala proteins (1 μM) were reacted for 5 min at 37˚C with 0.1 μM $_{HO}$RNAp (a 20-mer oligoribonucleotide, labeled with ^{32}P at the penultimate phosphate)

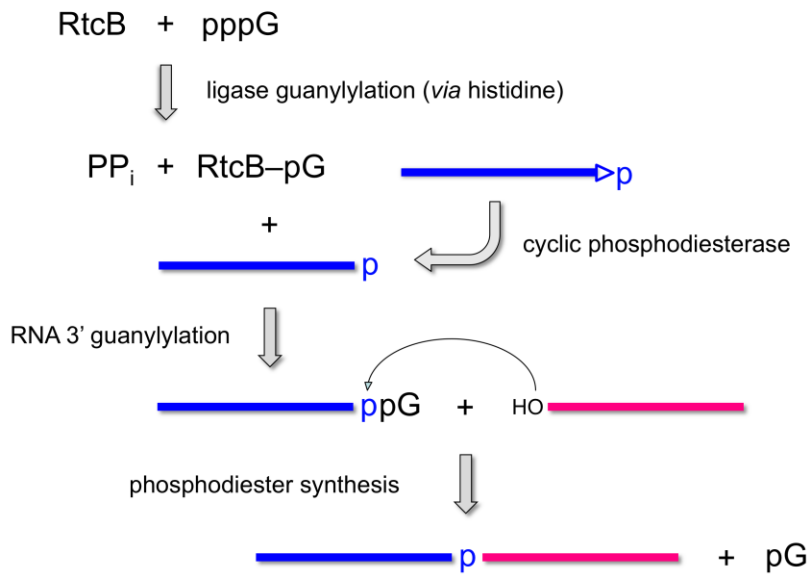


Figure 2.1. **Multi-step pathway of RtcB-catalyzed RNA ligation**. See text for details.

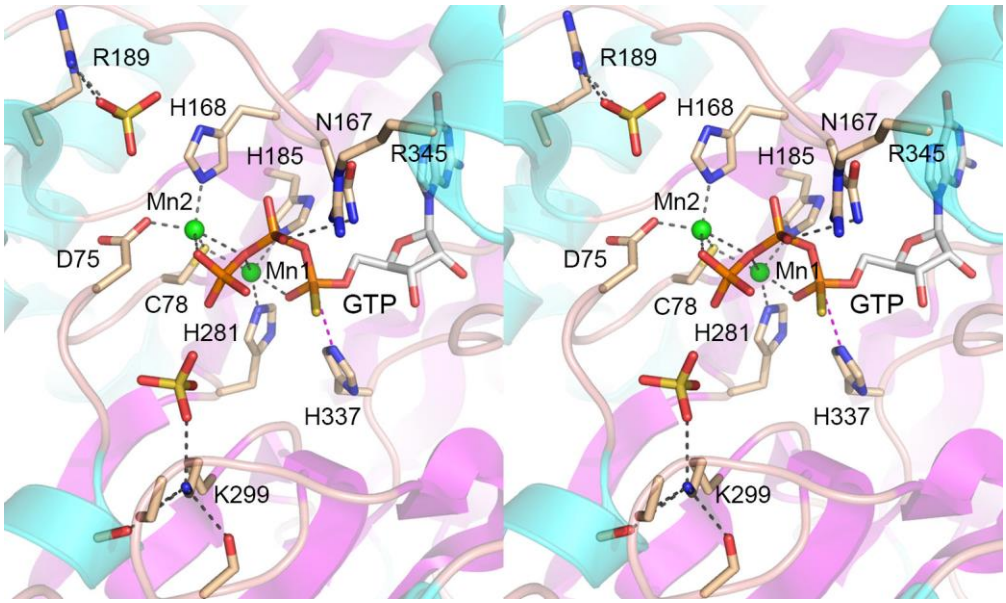


Figure 2.2. **Structure of an RtcB•GTP•$(Mn^{2+})_2$ complex**. Stereo view of the active site of *P. horikoshii* RtcB (from pdb 4ISZ) highlighting atomic interactions of two manganese ions (Mn1 and Mn2, depicted as green spheres) and GTPαS (stick model with gray carbons). The RtcB fold is shown as a cartoon model with magenta β strands and cyan α helices. Selected RtcB amino acids are depicted as stick models with beige carbons. Amino acids are numbered according to their equivalents in *E. coli* RtcB. Two sulfate anions (stick models) in the vicinity of GTPαS are suggested to mimic RNA phosphates. Atomic contacts are indicated by black dashed lines. His337-Nε is poised for nucleophilic attack on the GTP α phosphorus, as denoted by the magenta dashed line.

in the presence of 2 mM Mn^{2+} and 100 μM GTP. The products were analyzed by urea-PAGE and visualized by autoradiography (Fig. 2.4A). Wild-type RtcB converted the $_{HO}$RNAp substrate into a circular intramolecular ligation product that migrated ahead of the substrate strand. RtcB also generated more slowly migrating multimers via intermolecular end joining. H337A, which lacks the histidine nucleophile for the RtcB reaction with GTP to form RtcB–pG (Chakravarty et al., 2012), was inert in $_{HO}$RNAp ligation, as was the C78A mutant (Fig. 2.4A,B), which lacks the cysteine that coordinates both of the manganese ions in the active site (Fig. 2.2). D75A, which lacks

an aspartate that coordinates Mn2, was crippled in $_{HO}$RNAp ligation, with only 4% of the input substrate being sealed (Fig. 2.4A,B). By contrast, H185A (lacking one of the ligands for Mn1) and R345A were similar to wild-type with respect to their high extent of sealing and mixture of circular and multimer products (Fig. 2.4A,B). H168A (missing one of the Mn2 ligands) and K299A had diminished ligase activity, whereby half of the input $_{HO}$RNAp substrate was sealed. H281A was affected more severely, sealing only 13% of the $_{HO}$RNAp substrate.

The most instructive findings pertained to the N167A, R189A, and R341A mutants, which, notwithstanding their variable extents of RNA sealing (65%, 85%, and 8%, respectively), all accumulated an $_{HO}$RNAppG intermediate that migrated just above

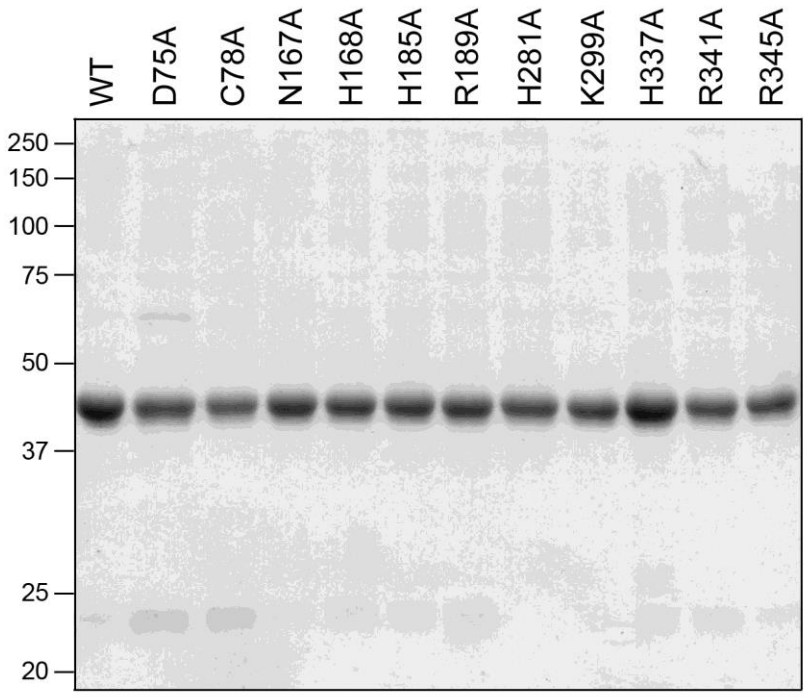
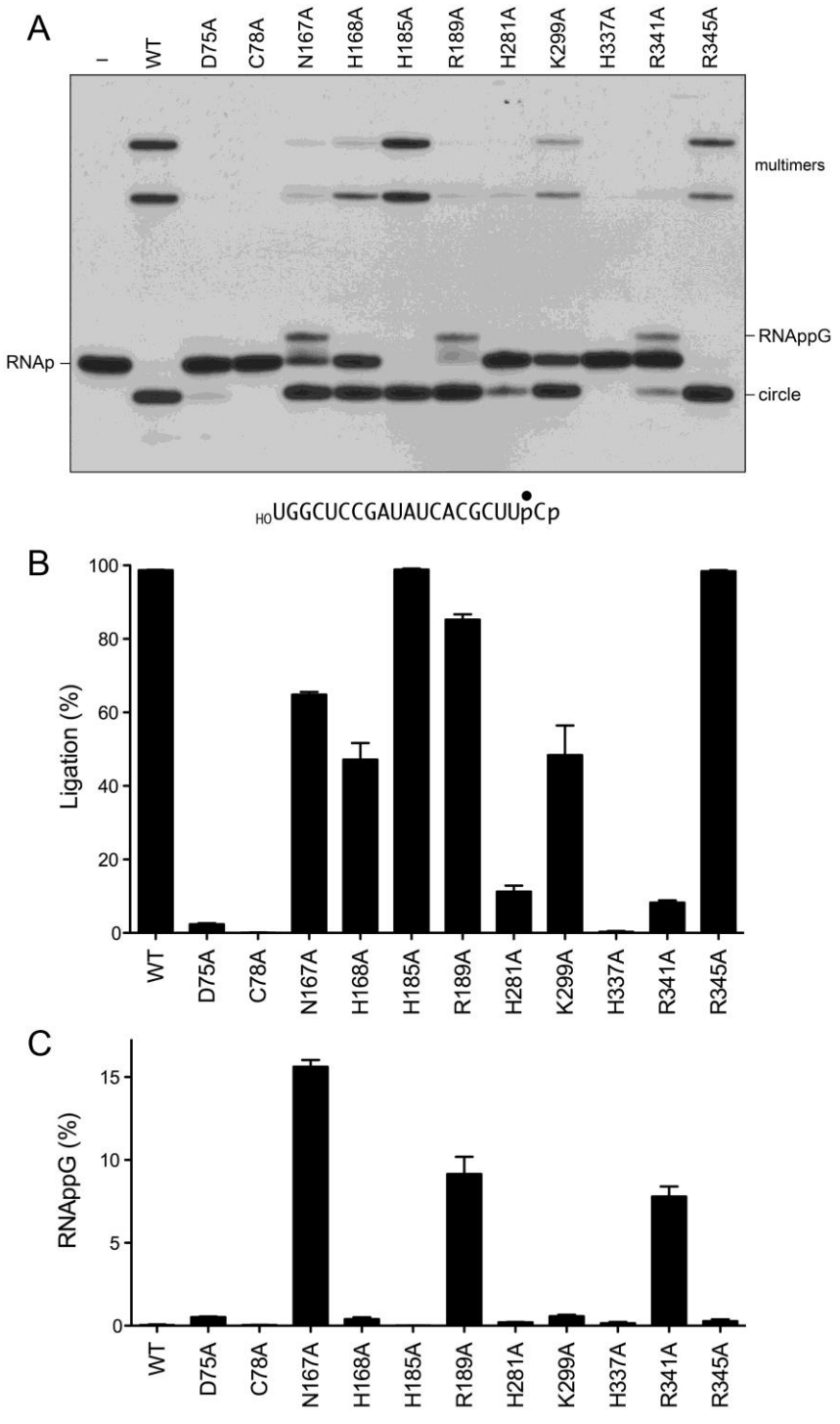


Figure 2.3. **RtcB mutants**. Aliquots (5 µg) of the indicated RtcB preparations were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (kDa) of marker polypeptides are indicated on the *left*.

Figure 2.4. **Mutational effects on $_{HO}$RNAp ligation**. (A) Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM 20-mer $_{HO}$RNAp (depicted at *bottom* with the radiolabeled phosphate denoted by ●), and either 1 µM RtcB as specified or no RtcB (lane –) were incubated at 37˚C for 5 min. The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown.) The identities of the radiolabeled RNAs are indicated at *left* and *right*. (B and C) The extents of RNA ligation (circle plus multimers; panel B) and RNAppG accumulation (panel C), expressed as a percentage of total radiolabeled RNA, are plotted in bar graph format. Each datum is the average of three experiments ±SEM.


A
–
WT
D75A
C78A
N167A
H168A
H185A
R189A
H281A
K299A
H337A
R341A
R345A
multimers
RNAppG
circle
RNAp
HOUGGCUCCGAUAUCACGCUUpCp
B
Ligation (%)
100
80
60
40
20
0
WT
D75A
C78A
N167A
H168A
H185A
R189A
H281A
K299A
H337A
R341A
R345A
C
RNAppG (%)
15
10
5
0
WT
D75A
C78A
N167A
H168A
H185A
R189A
H281A
K299A
H337A
R341A
R345A

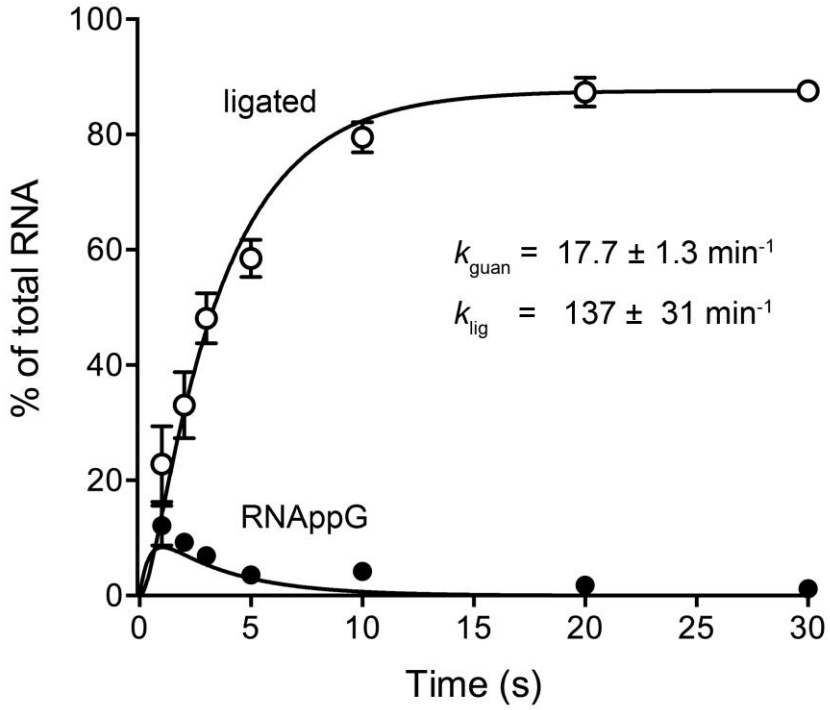


Figure 2.5. **Kinetic profile of $_{HO}$RNAp ligation by wild-type RtcB**. Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 1 µM RtcB, and 0.1 µM $_{HO}$RNAp were incubated at 37°C. The reactions were quenched with formamide, EDTA at the times specified. The products were resolved by urea-PAGE and quantified by scanning the gel. The levels of RNAppG (●) and ligated RNAs (○) are plotted as a function of time. Each datum is the average of three experiments ± SEM. The data were fit by non-linear regression in Prism to a two-step kinetic mechanism ($_{HO}$RNAp →$_{HO}$RNAppG → ligated RNA) with rate constants k_{guan} and k_{lig} as indicated.

the $_{HO}$RNAp substrate (Fig. 2.4A). The $_{HO}$RNAppG intermediate comprised between 8% and 16% of the labeled RNA in these reactions (Fig. 2.4C).

R189A selectively affects the kinetics of phosphodiester synthesis. Previous kinetic studies of the single-turnover $_{HO}$RNAp ligation reaction of wild-type RtcB detected only trace levels of $_{HO}$RNAppG (1% of the total labeled RNA) at the earliest times sampled: 15 and 30 s (Chakravarty et al., 2012). Here we revisited the kinetics of single-turnover $_{HO}$RNAp ligation by sampling the reaction products at shorter intervals. $_{HO}$RNAppG

comprised 12% and 9% of the labeled RNA at 1 and 2 s and then declined, concomitant with a steady increase in ligated RNA (Fig. 2.5). Fitting the data to a simple two-step kinetic scheme yielded apparent rate constants of 17.7 min^{-1} for formation of $_{HO}$RNAppG (k_{guan}) and 137 min^{-1} for subsequent phosphodiester synthesis (k_{lig}).

The kinetic profile of the $_{HO}$RNAp ligation reaction of the R189A mutant was markedly different, with respect to the time scale (endpoint approached in 10 min) and the transient accumulation of very high levels of the $_{HO}$RNAppG intermediate, which comprised 58% of total RNA at 15 and 30 s (Fig. 2.6, top panel). Indeed, the R189A profile verified the precursor-product relationship between $_{HO}$RNAppG and sealed RNA. Fitting the R189A data to two-step scheme yielded k_{guan} and k_{lig} values of 5.75 ± 0.55 min^{-1} and 0.66 ± 0.033 min^{-1}, respectively. Thus, whereas the R189A change slowed k_{guan} by a factor of 3 compared to wild-type RtcB, it slowed k_{lig} by a factor of 200. These results implicate Arg189 specifically in catalysis of 5'-OH attack on RNAppG to form a phosphodiester splice junction.

The N167A mutant displayed slower kinetics of RNA guanylylation (with $_{HO}$RNAppG comprising 11% of total RNA at 2, 3 and 5 min) and a lag in the accumulation of ligated product (Fig. 2.6, bottom panel). Whereas these data did not fit to a simple two-step scheme, we could derive a rate constant of 0.25 ± 0.018 min^{-1} for $_{HO}$RNAppG formation, by plotting $_{HO}$RNAppG plus ligated RNA as a function of time

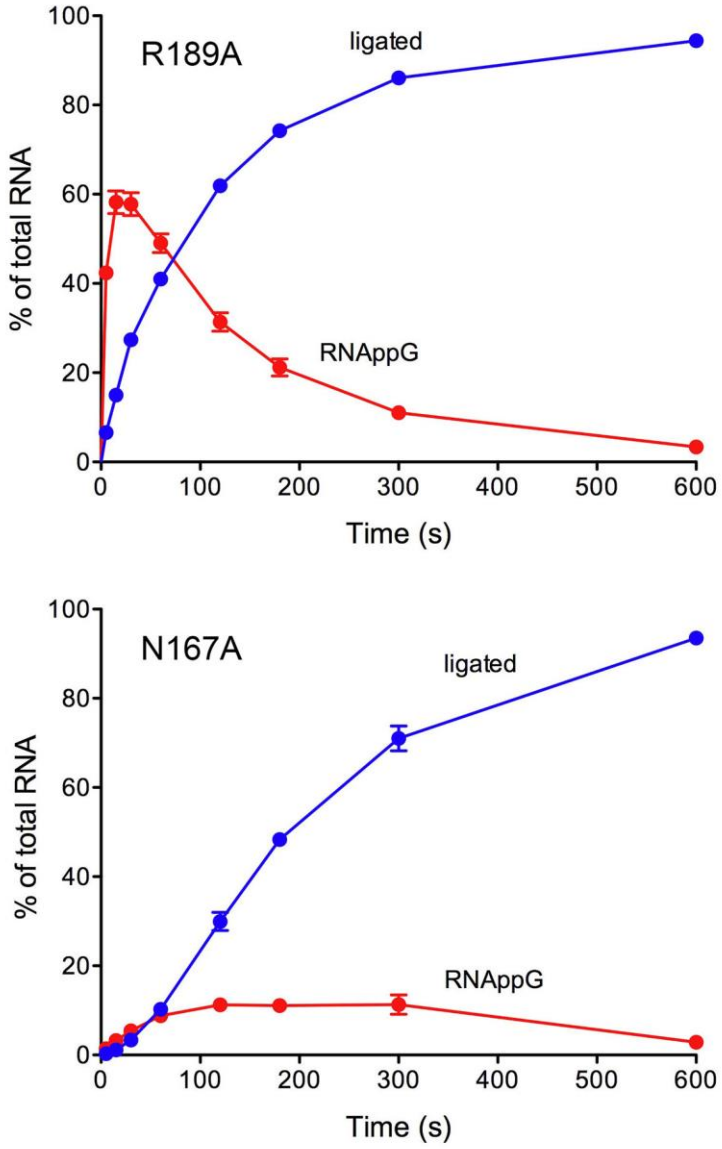
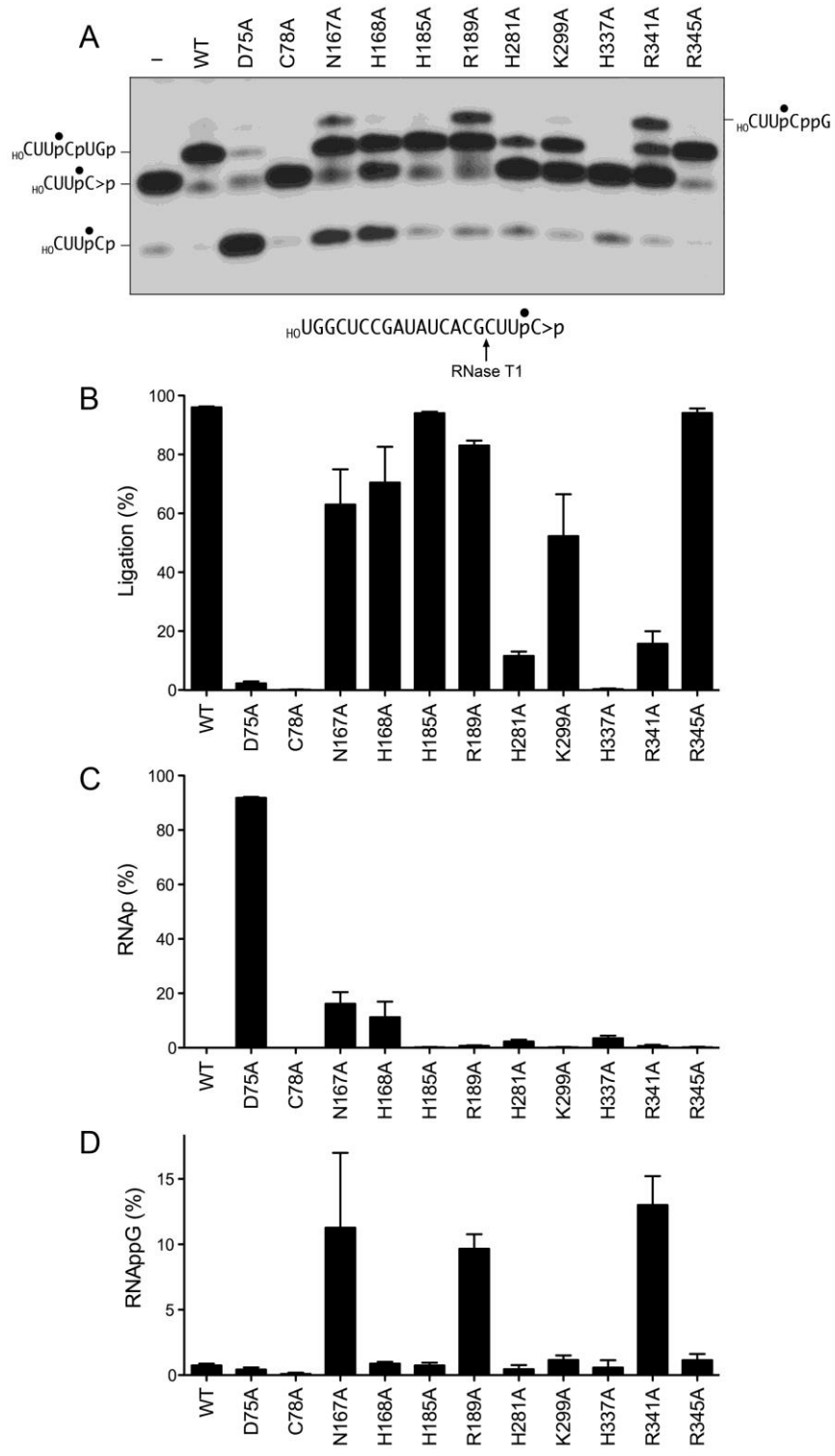


Figure 2.6. **Kinetic profiles of $_{HO}$RNAp ligation by R189A and N167A**. Reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM $_{HO}$RNAp, and 1 µM RtcB mutants R189A (top panel) or N167A (bottom panel) were incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and quenched immediately with formamide, EDTA. The products were resolved by urea-PAGE and quantified by scanning the gel. The levels of RNAppG and ligated RNAs are plotted as a function of time. Each datum is the average of three experiments ± SEM.

Figure 2.7. **Mutational effects on $_{HO}$RNA>p ligation**. (A) Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM 20-mer $_{HO}$RNA>p (depicted at *bottom* with the radiolabeled phosphate denoted by ●), and either 1 µM RtcB as specified or no RtcB (lane –) were incubated at 37˚C for 5 min. The mixtures were digested with RNase T1 and then analyzed by urea-PAGE. An autoradiograph of the gel is shown. The identities of the radiolabeled RNase T1 fragments are indicated at *left* and *right*. (B–D) The extents of RNA ligation ($_{HO}$CUU**p**CpUGp; panel B) and accumulation of RNAp ($_{HO}$CUU**p**Cp; panel C) and RNAppG ($_{HO}$CUU**p**CppG; panel D), expressed as a percentage of total radiolabeled RNA, are plotted in bar graph format. Each datum is the average of three experiments ±SEM. Either 1 µM RtcB as specified or no RtcB (lane –) were incubated at 37˚C for 5 min. The mixtures were digested with RNase T1 and then analyzed by urea-PAGE. An autoradiograph of the gel is shown. The identities of the radiolabeled RNase T1 fragments are indicated at *left* and *right*. (B–D) The extents of RNA ligation ($_{HO}$CUU**p**CpUGp; panel B) and accumulation of RNAp ($_{HO}$CUU**p**Cp; panel C) and RNAppG ($_{HO}$CUU**p**CppG; panel D), expressed as a percentage of total radiolabeled RNA, are plotted in bar graph format. Each datum is the average of three experiments ±SEM.


A
–
WT
D75A
C78A
N167A
H168A
H185A
R189A
H281A
K299A
H337A
R341A
R345A
HOCUUpCpUGp
HOCUUpC>p
HOCUUpCp
HOCUUpCppG
HOUGGCUCCGAUAUCACGCUUpC>p
RNase T1
B
Ligation (%)
C
RNAp (%)
D
RNAppG (%)

and fitting those data to a single exponential. In comparison to wild-type RtcB, the N167A change slowed $_{HO}$RNAppG formation by a factor of 70.

Mutational effects on the reaction of RtcB with HORNA>p. To illuminate mutational effects on the cyclic phosphodiesterase phase of the RNA repair/splicing pathway, wild-type RtcB and RtcB-Ala proteins (1 μM) were reacted for 5 min at 37˚C with 0.1 μM $_{HO}$RNA>p (a 20-mer RNA strand with 5'-OH and 2',3'-cyclic phosphate ends and a single radiolabel between the 3'-terminal and penultimate nucleosides) in the presence of 2 mM Mn^{2+} and 0.1 mM GTP. The reactions were quenched with EDTA. The mixtures were digested with RNase T1 and then analyzed by denaturing PAGE. RNase T1 incised the substrate 3' of the most distal guanosine to yield the ^{32}P-labeled tetranucleotide $_{HO}$CUU**p**C>p (Fig. 2.7A, lane –). Wild-type RtcB caused a depletion of the $_{HO}$CUU**p**C>p T1 fragment and the appearance of a more slowly migrating T1 fragment that corresponded to a 6-mer oligonucleotide, $_{HO}$CUU**p**CpUGp, released by T1 incision at the guanosines flanking the ligation junction (Fig. 2.7A). As reported previously (Chakravarty & Shuman, 2012), the H337A mutant did not convert RNA>p to RNAp, signifying that RtcB–pG formation precedes cyclic phosphodiester hydrolysis (Fig. 2.1). The C78A mutant was also unable to hydrolyze $_{HO}$RNA>p to $_{HO}$RNAp (Fig. 2.7A).

Mutants H185A and R354A ligated the $_{HO}$RNA>p substrate with high efficiency (94%), with little or no apparent accumulation of the $_{HO}$RNAp or $_{HO}$RNAppG intermediate species (Fig. 2.7A,B,C). These data accord with the activity of H185A and R354A in $_{HO}$RNAp ligation (Fig. 2.4) and signify that H185A and R345A are competent

for CPDase activity. H281A and K299A ligated 12% and 52% of the $_{HO}$RNA>p substrate (consistent with the $_{HO}$RNAp sealing data in Fig. 2.4) and did not accumulate $_{HO}$RNAp or $_{HO}$RNAppG intermediates (Fig. 2.7). R189A and R341A ligated 83% and 16% of the $_{HO}$RNA>p substrate and both mutants did accumulate $_{HO}$RNAppG (to an extent of 10-13% of total RNA, again in accord with the $_{HO}$RNAp sealing data in Fig. 2.4), though neither mutant accumulated $_{HO}$RNAp in the single point assay format.

The key findings from the $_{HO}$RNA>p sealing experiment were that mutants H168A, N167A, and D75A accumulated $_{HO}$RNAp, and did so in distinct fashion with respect to downstream steps in the RtcB pathway. The H168A reaction with $_{HO}$RNA>p yielded ligated RNA and $_{HO}$RNAp, without accumulating $_{HO}$RNAppG. N167A generated ligated RNA and $_{HO}$RNAp, while also accumulating $_{HO}$RNAppG (Fig. 2.7A-D). The kinetic profile of the single-turnover reaction of N167A with $_{HO}$RNA>p revealed a steady increase in $_{HO}$RNAp up to 1 min, followed by a crest of $_{HO}$RNAp at the 2 min and 3 min times, at which it comprised 28% of the total RNA (Fig. 2.8, green symbols). $_{HO}$RNAp declined thereafter, concomitant with the appearance of $_{HO}$RNAppG (Fig 2.8, red symbols), which comprised 9% of total RNA at 2, 3, and 5 min, and the steady conversion of $_{HO}$RNAppG to ligated RNA (Fig. 2.8, blue symbols) from 1 to 10 min. These data verify that $_{HO}$RNAp is an intermediate in the $_{HO}$RNA>p ligation reaction. We derived a rate constant of 0.53 ± 0.053 min^{-1} for the N167A CPDase, by plotting $_{HO}$RNAp plus $_{HO}$RNAppG plus ligated RNA as a function of time (Fig. 2.8, black open circles) and fitting those data to a single exponential.

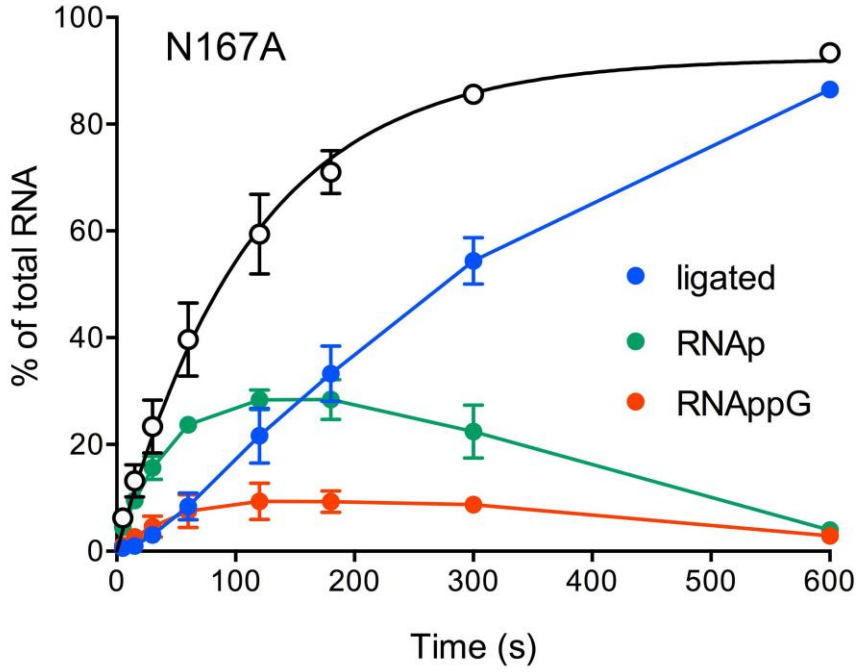


Figure 2.8. **Kinetic profile of $_{HO}$RNA>p ligation by N167A**. Reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 μM GTP, 0.1 μM 20-mer $_{HO}$RNA>p, and 1 μM RtcB mutant N167A were incubated at 37°C. Aliquots (10 μl) were withdrawn at the times specified, quenched with EDTA, digested with RNase T1, mixed with formamide, EDTA, and then analyzed by urea-PAGE. The extents of RNA ligation and accumulation of RNAp and RNAppG, expressed as a percentage of total radiolabeled RNA, are plotted as a function of time. Each datum is the average of three experiments ±SEM. The black circles denote RNAp plus RNAppG plus ligated RNA (as % of total RNA) with a fit of the data to a single exponential.

The most salient observation was that D75A converted 92% of the $_{HO}$RNA>p substrate to $_{HO}$RNAp, but achieved only 4% ligation (Fig. 2.7). These results, and those in Fig. 2.4, signify that Asp75 is not strictly essential for RtcB CPDase activity, but is essential for RNA guanylylation.

Characterization of D75A CPDase activity. Prior studies had established that the CPDase activity of wild-type RtcB is dependent on GTP (Chakravarty & Shuman, 2012). Here we queried whether this was the case for the D75A mutant, by varying the

concentration of GTP included in the single-turnover reaction of D75A with $_{HO}$RNA>p, in the range of 0 to 500 μM GTP. The conversion of $_{HO}$RNA>p to $_{HO}$RNAp required GTP and the extent of conversion increased with GTP concentration, attaining saturation at 50 μM GTP with 91% conversion to $_{HO}$RNAp and 1.8% ligation (Fig. 2.9A). The data fit to a one-site binding model with half-saturation at 6 μM GTP (Fig. 2.9B). The kinetic profile of the D75A CPDase reaction highlighted steady conversion of $_{HO}$RNA>p to $_{HO}$RNAp to attain an endpoint after 3 min (Fig. 2.9C,D). The shape of the reaction curve was notable for a slight initial lag in $_{HO}$RNAp product accumulation by D75A, suggestive of a partially rate-limiting upstream step not directly detected by the assay format (presumably, the formation of an RtcB-D75A–pG intermediate). Half of the substrate was hydrolyzed to $_{HO}$RNAp in 1.25 min, from which we can estimate a CPDase rate of ~0.55 min^{-1}.

Mutational effects on DNA 3'-phosphate capping. Bacterial RtcB enzymes transfer GMP from GTP to a single-stranded DNA 3'-phosphate end (DNAp) to form a DNAppG cap structure (Das et al., 2013). To probe the enzymic requirements for DNA capping, we reacted the *E. coli* RtcB proteins (1 μM) for 5 min at 37˚C with 0.1 μM 5' ^{32}P-labeled 12-mer pDNAp substrate (Fig. 2.10A) in the presence of 2 mM Mn^{2+} and 100 μM GTP. The products were analyzed by urea-PAGE and visualized by autoradiography. Wild-type RtcB converted 91% of the pDNAp substrate into a slower migrating pDNAppG capped product. For most of the mutants, the effects on DNA capping (Fig. 2.10) echoed what was seen for $_{HO}$RNAp ligation (Fig. 2.4). To wit: (i) H337A and C78A were inert in DNA capping; (ii) D75A, H281A, and R341A were severely compromised (2%, 0.5% and 7% pDNAp capping, respectively); (iii) N167A

(31%), H168A (25%), and K299A (43%) were modestly impaired; and (iv) R189A retained near wild-type capping activity in the single point assay (Fig. 2.10B). The exceptions were H185A and R345A, which were feeble at DNA capping (effecting 12% and 14% conversion of pDNAp to DNAppG, respectively; Fig. 2.10B) but relatively unaffected in $_{HO}$RNAp ligation (Fig. 2.4B).

Effects on DNAppG ligation and de-guanylylation. The GTP-dependent steps of the RtcB ligation pathway can be bypassed by presenting the enzyme with a "broken" stem-loop DNA substrate composed of a pre-guanylylated 5' ^{32}P-labeled pDNAppG strand annealed to an unlabeled 5'-OH DNA strand (Fig. 2.11A). As shown previously (Das et al., 2013), wild-type RtcB will catalyze manganese-dependent phosphodiester synthesis, in the absence of added GTP, via attack of the DNA 5'-OH on the 3'-phosphate of pDNAppG to yield a sealed DNA stem loop (Fig. 2.10A, ligation). RtcB will also catalyze "backward" GMP transfer from the pDNAppG cap to the enzyme to form RtcB–pG and a 5' ^{32}P-labeled pDNAp strand (Fig. 2.10A, de-guanylylation). Thus, the fate of the pDNAppG strand of the broken stem-loop reflects partitioning between ligation and de-guanylylation. For example, in the experimental set-up in Fig. 11, wild-type RtcB ligated 67% of the input substrate (Fig. 2.11B) and de-guanylylated 24% (Fig.

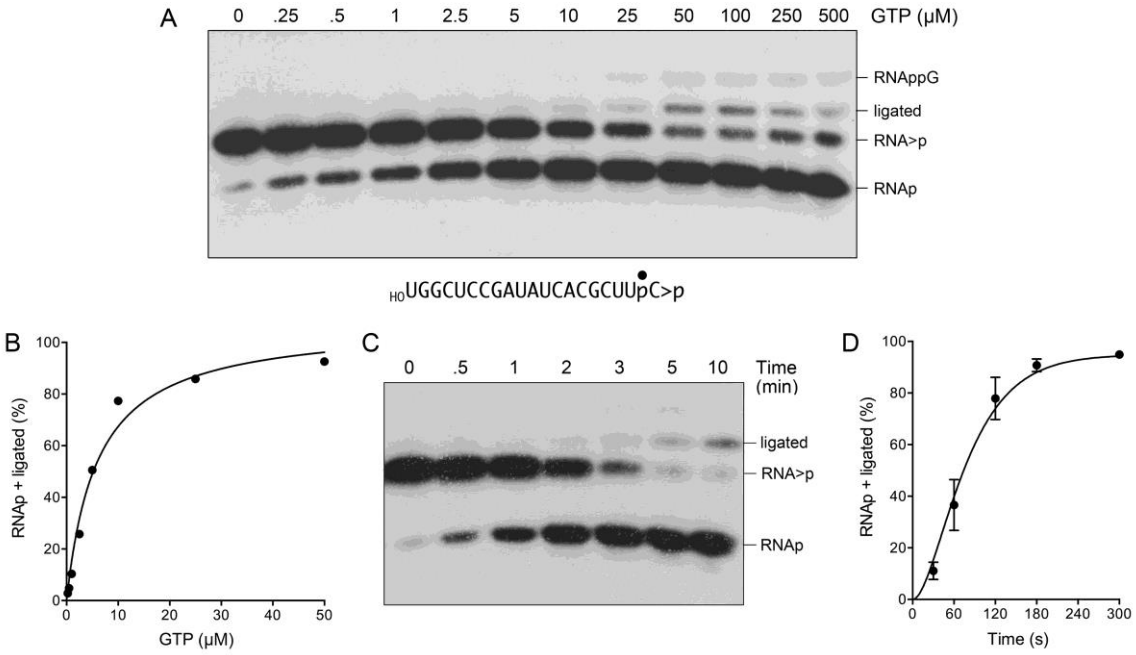


Figure 2.9. **CPDase activity of D75A**. (A and B) Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 0.1 µM 20-mer $_{HO}$RNA>p (depicted at *bottom* with the radiolabeled phosphate denoted by ●), 1 µM RtcB D75A, and GTP as specified were incubated at 37°C for 5 min. The mixtures were digested with RNase T1 and then analyzed by urea-PAGE. An autoradiograph of the gel is shown. The identities of the radiolabeled RNase T1 fragments are indicated at *right*. The extent of 2',3'-cyclic phosphodiester hydrolysis (RNAp plus ligated RNA) is plotted as a function of GTP concentration in panel B. (C) A reaction mixture containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM 20-mer $_{HO}$RNA>p, and 1 µM RtcB D75A was incubated at 37˚C. Aliquots (10 µl) were withdrawn at the times specified and quenched with EDTA. The RNAs were digested with RNase T1 and the products were analyzed by urea–PAGE. An autoradiograph of the gel is shown. (D) The extent of 2',3'-cyclic phosphodiester hydrolysis (RNAp plus ligated RNA) is plotted as a function of time. Each datum is the average of three experiments ±SEM.

2.11C). Mutational effects on the RtcB reaction with the pre-guanylylated DNA stem-loop fell into several instructive categories. C78A abolished phosphodiester synthesis and de-guanylylation, thereby attesting to the central role of the manganese-bridging cysteine at every step of the RtcB pathway. As noted previously (Das et al., 2013),

mutation of the His337 nucleophile to alanine abolished de-guanylylation (Fig. 2.11C) but preserved phosphodiester synthesis, such that 89% of the input pDNAppG substrate was sealed (Fig. 2.11B). Mutations H281A and K299A elicited similar effects on the pDNAppG reaction outcome, i.e., suppressing de-guanylylation (1% and 6%, respectively) and enhancing ligation (86% and 88%, respectively) relative to wild-type RtcB (Fig. 2.11B,C). These results establish H281A as a separation-of-function

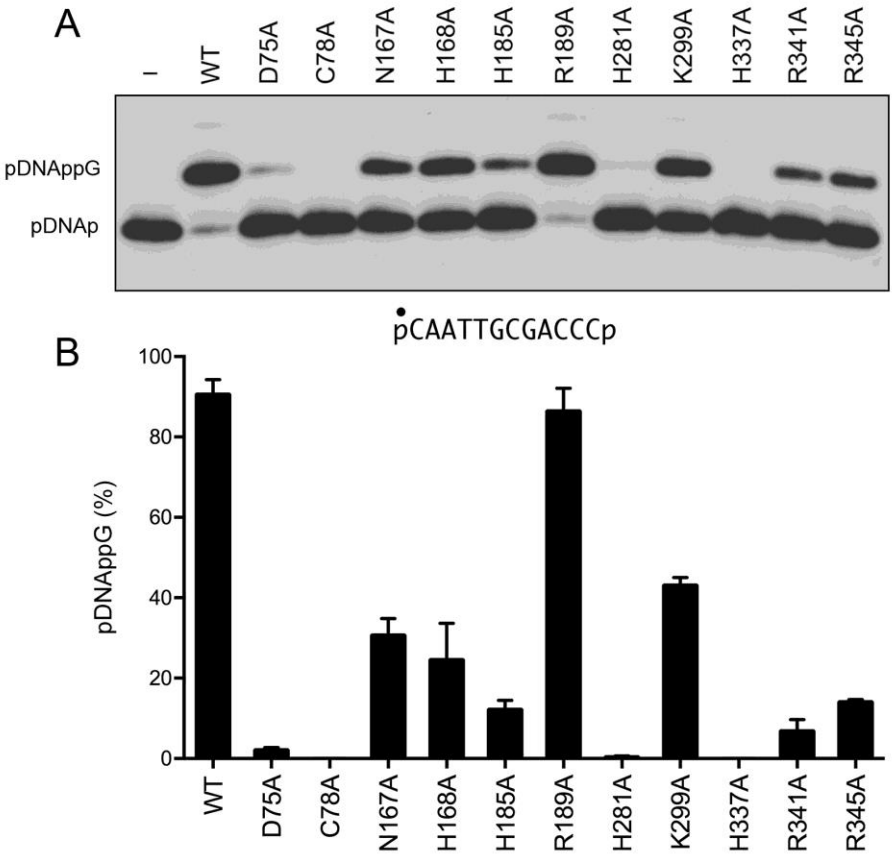


Figure 2.10. **Mutational effects on DNA 3'-phosphate capping**. Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM 12-mer pDNAp (depicted at *bottom* with the radiolabeled phosphate denoted by ●), and either 1 µM RtcB as specified or no RtcB (lane –) were incubated at 37˚C for 5 min. The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown. The pDNAp substrate and capped pDNAppG product are indicated at *left*. (B) The extents of DNAppG formation are plotted in bar graph format. Each datum is the average of three experiments ±SEM.

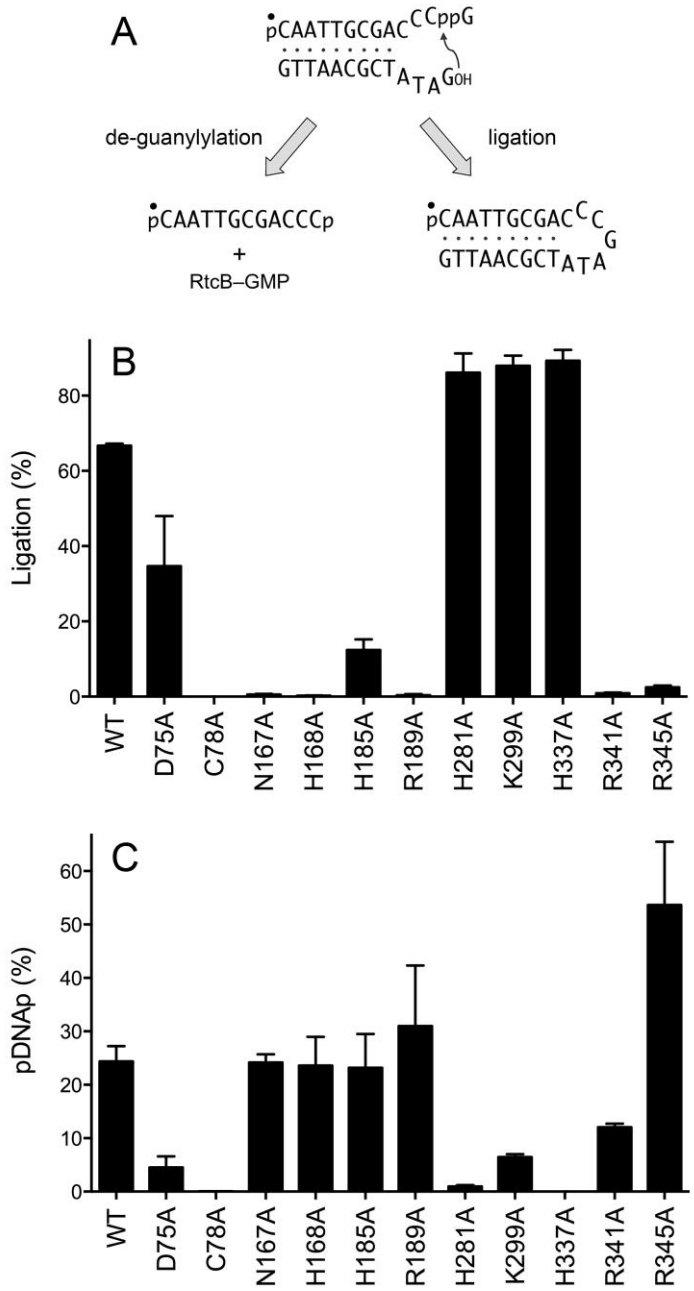


Figure 2.11. **Mutational effects of DNAppG ligation and de-guanylylation**. (A) Reactions of RtcB with a pre-guanylylated pDNAppG//$_{HO}$DNA stem-loop substrate. The 5' ^{32}P-label on the pDNAppG strand is denoted by ●. (B and C) Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 100 µM GTP, 0.1 µM 12-mer pDNAppG stem loop, and 1 µM RtcB as specified were incubated at 37˚C for 5 min. The products were resolved by urea-PAGE and quantified by scanning the gel. The extents of DNAppG ligation (panel B) and de-guanylylation to pDNAp (panel C) are plotted in bar graph format. Each datum is the average of three experiments ±SEM.

mutation that strongly impaired the GTP-dependent ligations of $_{HO}$RNAp (Fig. 2.4) and $_{HO}$RNA>p (Fig. 2.7) and GTP-dependent capping of pDNAp (Fig. 2.10), but had comparatively little effect on GTP-independent sealing of pDNAppG. We surmise that the contact of His281 with Mn1 (Fig. 2.2) is critical for the various GTP-dependent steps, but not for the step of phosphodiester synthesis at a pre-guanylylated end.

Other mutations exerted the opposite effect on the reaction of RtcB with the pDNAppG stem-loop, i.e., they suppressed ligation and favored de-guanylylation. This class included R189A, which ligated 0.4% of the pDNAppG strand, yet de-guanylylated 31% (Fig. 2.11B,C). This result accords with R189A accumulating $_{HO}$RNAppG during the $_{HO}$RNAp and $_{HO}$RNA>p sealing reactions and fortifies the conclusion that Arg189 is especially important for the step of phosphodiester synthesis. Other mutants with outcomes heavily skewed toward de-guanylylation of the pDNAppG stem-loop were: N167A (0.5% ligation, 24% de-guanylylation); H168A (0.2% ligation, 24% de-guanylylation); R341A (0.9% ligation, 12% de-guanylylation); R345A (2.5% ligation, 54% de-guanylylation) (Fig. 2.11B,C). H185A displayed a more modest preference in this direction (12% ligation, 23% de-guanylylation).

D75A, though less active than wild-type RtcB with the pDNAppG stem-loop, nonetheless retained a bias toward sealing as the reaction outcome (35% ligation, 5% de-guanylylation) (Fig. 2.11B,C). Thus, D75A is revealed as a separation-of-function mutant that is (relatively) competent in 2',3'-cyclic phosphodiester hydrolysis and phosphodiester synthesis, but severely defective in RNAp and DNAp guanylylation.

DISCUSSION

By studying the reactions of wild-type *E. coli* RtcB and RtcB-Ala mutants with 3'-phosphate, 2',3'-cyclic phosphate, and 3'-ppG terminated substrates, we provide new evidence in support of the multi-step ligation pathway in Fig. 2.1 and we illuminate the distinctive contributions of certain active site amino acids to individual steps of the ligation pathway. The mutational data, when interpreted in light of archaeal RtcB structures (Desai et al., 2013), suggest that: (i) the enzymic constituents of the two manganese coordination complexes play distinct catalytic roles; and (ii) Arg189 is a plausible candidate to position the 5'-OH polynucleotide end during the step of phosphodiester synthesis.

Initial characterization of *E. coli* RtcB established a requirement for manganese as the cofactor for ligation (Tanaka & Shuman, 2011b) by virtue of its essentiality for both RtcB–pG formation (Tanaka et al., 2011) and sealing of a pre-guanylylated stem-loop (Das et al., 2013). *Pyrococcus* RtcB structures highlighted RtcB as a binuclear metalloenzyme with closely spaced Mn^{2+} ions bridged by an invariant cysteine (Englert et al., 2012; Desai et al., 2013). Mutations of this cysteine abolished the overall ligase activity of *E. coli*, *Pyrococcus*, and mammalian RtcBs (Englert et al., 2012; Popow et al., 2011a; Tanaka et al., 2011b). In the present study, we showed that Cys78 of *E. coli* RtcB was essential for every reaction step tested. However, subtractions of other enzymic side chain metal ligands elicited defects of varying magnitude on different component pathway steps.

His281 and His185 are the unique amino acid components of the Mn1 coordination complex, but their mutations had disparate effects. Loss of His281 was quite deleterious to overall $_{HO}$RNAp and $_{HO}$RNA>p ligation and to pDNAp capping, but was well tolerated with respect to sealing of a pre-guanylylated stem-loop, implying that His281 and its contact with Mn1 are critical for RtcB's reaction with GTP. By contrast, loss of His185 was relatively well tolerated with respect to $_{HO}$RNAp and $_{HO}$RNA>p ligation, suggesting that the H185A mutant's reactivity with GTP was preserved. In this vein, it is worth noting that the reaction of *Pyrococcus* RtcB with GTP was preserved when the equivalent His234 Mn1 ligand was mutated to alanine (Englert et al., 2013). (H185A did have a detrimental effect on pDNAp capping and pDNAppG ligation by *E. coli* RtcB. Selective impact on DNA substrates is a feature shared with R345A. The basis for this effect is unclear, absent a structure of RtcB in complex with nucleic acid.)

His168 and Asp75 are the unique constituents of the Mn2 coordination complex that includes the GTP γ phosphate, and their mutations also had disparate effects. Subtracting Asp75 virtually effaced overall $_{HO}$RNAp and $_{HO}$RNA>p ligation and DNA 3'-phosphate capping, while permitting hydrolysis, in high yield, of an RNA 2',3'-cyclic phosphate in a manner that requires GTP, just as the CPDase activity of wild-type-RtcB is GTP-dependent (Chakravarty & Shuman, 2012). No other mutants in our collection displayed this biochemical phenotype. One early model for the RtcB mechanism had invoked a concerted single-step reaction of RtcB–pG with RNA>p to form RNAppG (Englert et al., 2013), but the concerted model does not stand up to the findings that: (i) wild-type RtcB is full capable of joining RNAp ends via an RNAppG intermediate; and (ii) the D75A mutation separates the CPDase activity from downstream steps. Our

findings implicate Mn2 ligand Asp75 as critical for the step of polynucleotide-3'-phosphate guanylylation, and they are in accord with the observation that the equivalent D95A change in *Pyrococcus* RtcB preserved enzyme guanylylation but abolished RNA ligation (Englert et al., 2013).

By contrast, loss of Mn2 ligand His168 had a modest effect on $_{HO}$RNAp and $_{HO}$RNA>p ligation and DNAp capping (e.g. compared to that of D75A). The detection of RNAp (but not RNAppG) among the products of the reaction of H168A with $_{HO}$RNA>p (Fig. 2.7) suggests that RNA-3'-phosphate guanylylation is most sensitive to this mutation.

The counterpart of *E. coli* RtcB Arg189 (Arg238 in *Pyrococcus* RtcB) coordinates a sulfate anion in the RtcB•GTP•$(Mn^{2+})_2$ structure (Fig. 2.2) and the RtcB–pG•$(Mn^{2+})_2$ structure (Desai et al., 2013) and may be construed to mimic the position of a phosphate in one of the RNA substrate strands. The sulfate sulfur atom is located 11 Å away from the GMP phosphorus atom of the RtcB–pG•$(Mn^{2+})_2$ covalent intermediate (Desai et al., 2013). Based on our findings that mutating Arg189 to alanine selectively affected phosphodiester synthesis, while sparing RNAp guanylylation, we speculate that Arg189 promotes catalysis of phosphodiester synthesis by coordinating the 5'-terminal phosphodiester of the $_{HO}$RNA strand ($_{HO}N^1\mathbf{p}N^2pN^3$–) and thereby positioning the 5'-OH for its attack on the 3'-phosphate of RNAppG. This is analogous to the established role of an essential arginine in polynucleotide kinase enzymes, whereby the arginine coordinates the 5'-terminal phosphodiester of the $_{HO}$RNA or $_{HO}$DNA strand to poise the 5'-OH for its attack on the γ-phosphate of the NTP phosphate donor (Das et al., 2014;

Eastberg et al., 2004). Needless to say, the proof of this speculation for the RtcB arginine will hinge on capturing crystal structures of RtcB–pG in complex with RNAp and/or $_{HO}$RNA strands.

The N167A mutation affected the CPDase, RNA guanylylation, and phosphodiester synthesis steps of the RtcB pathway. The involvement of Asn167 is multiple reaction steps is in keeping with the atomic interactions of the equivalent Asn202 side chain in *Pyrococcus* RtcB. The Asn contacts a GTP β phosphate non-bridging oxygen in the RtcB•GTP•$(Mn^{2+})_2$ structure (Fig. 2.2); the Asn undergoes a rotamer shift in the RtcB–pG•$(Mn^{2+})_2$ structure so that it contacts one non-bridging oxygen and the 5'-bridging oxygen of the covalently bound GMP (Desai et al., 2013).

Finally, we suspect that the functional consequences of mutations K299A, R341A, and R345A reflect the "structural" roles of these functional groups. The *Pyrococcus* RtcB equivalent of Lys299 (Lys351) donates hydrogen bonds to three main-chain carbonyls that tether multiple secondary structure elements of the RtcB fold (Fig. 2.2). The counterpart of Arg345 (Arg412) makes a salt bridge to a conserved glutamate. The equivalent of Arg341 (Arg408) makes a salt bridge to the same glutamate, and also donates hydrogen bonds to a main-chain carbonyl (Desai et al., 2013).

CHAPTER 3

CHARACTERIZATION OF 3'-PHOSPHATE RNA LIGASE PARALOGS RTCB1, RTCB2, AND RTCB3 FROM MYXOCOCCUS XANTHUS HIGHLIGHTS A DNA AND RNA 5'-PHOSPHATE CAPPING ACTIVITY OF RTCB3

INTRODUCTION

Escherichia coli RtcB is a founding member of a recently discovered family of RNA repair/splicing enzymes that join RNA 2',3'-cyclic-PO_4 or 3'-PO_4 ends to RNA 5'-OH ends (Englert et al., 2011; Popow et al., 2011b; Tanaka et al., 2011b; Tanaka & Shuman, 2011b). RtcB executes a four-step pathway that requires GTP as an energy source and Mn^{2+} as a cofactor (Chakravarty et al., 2012; Chakravarty & Shuman, 2012; Tanaka et al., 2011). RtcB first reacts with GTP to form a covalent RtcB-(histidinyl-N)-GMP intermediate. It then hydrolyzes the RNA 2',3'-cyclic-PO_4 end to a 3'-PO_4 and transfers guanylate to the RNA 3'-PO_4 to form an $RNA_{3'}pp_{5'}G$ intermediate. Finally, RtcB catalyzes the attack of an RNA 5'-OH on the $RNA_{3'}pp_{5'}G$ end to form the 3'-5' phosphodiester splice junction and liberate GMP. The unique chemical mechanism of RtcB overturned a longstanding canon of nucleic acid enzymology, which held that synthesis of polynucleotide 3'-5' phosphodiesters proceeds via the attack of a 3'-OH on a high-energy 5' phosphoanhydride: either a nucleoside 5'-triphosphate in the case of RNA/DNA polymerases or an adenylylated intermediate $A_{5'}pp_{5'}N$– in the case of classic RNA/DNA ligases.

The wide distribution of RtcB proteins in bacteria, archaea and metazoa – and the fact that crystal structures of thermophilic archaeal (Desai et al., 2013; Englert et al., 2012; Okada et al., 2006) and bacterial (Sekine et al., 2007) RtcBs unveiled RtcB to have a novel fold and active site, with no similarity whatsoever to classic RNA/DNA ligases – highlighted RtcB as a harbinger of an alternative enzymology based on covalently activated 3'-PO_4 ends (Chakravarty et al., 2012). After speculating that the chemistry of 3'-PO_4/5'-OH end joining by RtcB might be portable to DNA transactions, (Das et al., 2013) showed that *Escherichia coli* RtcB can indeed modify DNA 3'-PO_4 ends. Specifically, RtcB transfers GMP from a covalent RtcB–GMP intermediate to a DNA 3'-PO_4 to form a chemically stable "capped" 3' end structure, $DNA_{3'}pp_{5'}G$.

The implications of RtcB-mediated DNA capping for nucleic acid break repair are potentially significant, given the many biological settings in which damage generates 3'-PO_4 ends that are refractory to the action of DNA polymerases and classic ligases. Biochemical studies of the impact of DNA 3' capping on repair reactions highlighted two key points: (i) the cap protects DNA ends from 3' exonucleases; and (ii) the cap guanosine 3'-OH can serve as a primer for templated DNA synthesis by exemplary members of five different DNA polymerase families (Chauleau et al., 2015; Das et al., 2014).

RtcB homologs are present in the proteomes of scores of bacterial taxa, even though bacteria have no evident need for protein-catalyzed tRNA splicing or mRNA splicing, which are the functions ascribed to metazoan RtcB homologs during the maturation of intron-containing tRNAs and during non-spliceosomal splicing of XBP1

mRNA during the unfolded protein response to ER stress (Jurkin et al., 2014; Kosmaczewski et al., 2014; Lu et al., 2014; Ray et al., 2014; Tanaka et al., 2011b). In *E. coli*, RtcB is encoded in an operon with the RNA 3'-phosphate cyclase enzyme RtcA (Genschik et al., 1998). The fact that the *rtcBA* operon in *E. coli* is regulated by the σ^{54} co-activator RtcR (encoded by the *rtcR* gene, located immediately upstream of *rtcBA* and transcribed in the opposite orientation) suggested that the RNA repair functions are induced in response to cellular stress (Genschik et al., 1998), conceivably as a means to recover from RNA damage inflicted by bacterial ribotoxins that are turned on in stress situations. This scenario remains speculative because the stress signal that activates RtcR is unknown and deletion of *rtcBA* elicits no overt phenotype in *E. coli* (Genschik et al., 1998).

Based on the biochemical activities of *E. coli* RtcB on RNA and DNA substrates, there is no basis to restrict our ideation about RtcB function to RNA *versus* DNA transactions. Indeed, the fact that some bacteria encode multiple RtcB paralogs raises the prospect that bacterial RtcB-type ligases might acquire functional specialization for particular repair pathways or for the sealing of specific nucleic acid substrates. By analogy, many bacteria that code multiple DNA ligase enzymes have evolved a division of labor whereby one ligase serves an essential replicative function (e.g., sealing of Okazaki fragments) while a different ligase is dedicated to double-strand break repair via non-homologous end joining (Stewart Shuman & Glickman, 2007). The limitation to this line of thinking anent RtcB is that the *E. coli* enzyme is the only bacterial RtcB that has been characterized to date.

The goal of the present study was to gauge the activities of paralogous RtcB proteins from a single bacterial species. We chose to focus on *Myxococcus xanthus* by virtue of the presence of six predicted RtcBs in its proteome. Here we report the production, purification, and characterization of recombinant *M. xanthus* RtcB1, RtcB2, and RtcB3.

MATERIALS AND METHODS

Recombinant M. xanthus RtcB proteins. The *rtcB* ORFs encoding RtcBs 1–6 were amplified by PCR from *M. xanthus* genomic DNA with primers designed to introduce a BamHI site flanking the predicted ATG start codon and an NotI site immediately downstream of the stop codon. The PCR products were digested with BamHI and NotI and then inserted between the BamHI and NotI sites of pET28b-His_{10}Smt3. The pET28b-His_{10}Smt3-RtcB plasmids were transformed into *E. coli* BL21-CodonPlus(DE3). 1-liter cultures derived from single transformants were grown at 37°C in LB medium containing 50 μg/ml kanamycin until the A_{600} reached 0.6 to 0.8. The cultures were adjusted to 2% (v/v) ethanol and chilled on ice for 30 min, then adjusted to 0.1 mM isopropyl-β-D-thiogalactoside. Incubation was continued at 17°C for 16 h with constant shaking. Cells were harvested by centrifugation and stored at −80°C. All subsequent procedures were performed at 4°C. The cell pellets were suspended in 50 ml buffer A (50 mM Tris-HCl, pH 8.0, 350 mM NaCl, 10% sucrose, 10% glycerol). After adding lysozyme to 0.2 mg/ml and a protease inhibitor tablet (Roche), the suspensions were nutated for 1 h. The lysates were sonicated to reduce viscosity and insoluble material was removed by centrifugation at 20,000 g for 30 min. The soluble lysates were mixed for 1 h with 4 ml His60 Ni Superflow resin (Clontech) that had been equilibrated in buffer A. The resins were recovered by centrifugation and resuspended in 25 ml buffer B (50 mM Tris-HCl, pH 8.0, 350 mM NaCl, 10% glycerol) containing 25 mM imidazole. The cycle of centrifugation and resuspension of the resin in buffer B was repeated three times, after which the washed resins were poured into columns. The bound material was eluted stepwise with 8-ml aliquots of 50, 100, 200, 300, 400, and

500 mM imidazole in buffer B. The elution profiles were monitored by SDS-PAGE. The eluate fractions containing each His_{10}Smt3-RtcB were pooled and digested with 90 µg of the Smt3-specifc protease Ulp1 during overnight dialysis against buffer B. The dialysates were mixed for 1 h with 1.5 ml His60 Ni Superflow resin that had been equilibrated in buffer B and then poured into columns. After washing the columns with buffer B, the flow-through and wash fractions containing each tag-free RtcB were pooled, adjusted to 25 mM EDTA, and then concentrated by centrifugal ultrafiltration (Amicon Ultra-15, 10 kDa cut-off; Millipore). The RtcB proteins were then gel-filtered through a 120-ml 16/60 HiLoad Superdex 200 pg column (GE Healthcare) equilibrated with buffer C (10 mM Tris-HCl, pH 8.0, 350 mM NaCl, 1 mM DTT, 1mM EDTA, 10% glycerol). The peak RtcB-containing fractions were pooled and concentrated by centrifugal ultrafiltration. Protein concentration was determined by using the Biorad dye reagent with bovine serum albumin as the standard. The yields of purified RtcB1, RtcB2 and RtcB3 proteins from 1-liter cultures were 0.2 mg, 0.6 mg, and 0.6 mg, respectively.

RNA and DNA substrates. A 20-mer $_{HO}RNA_{3'}p$ strand labeled with ^{32}P at the penultimate phosphate (Fig. 3.2) was prepared by T4 Rnl1-mediated addition of [5'-^{32}P]pCp to a 19-mer synthetic oligoribonucleotide as described (Tanaka et al., 2011). 5′ ^{32}P-labeled pDNAp, $pDNA_{OH}$, and $pRNA_{OH}$ strands were prepared by enzymatic phosphorylation of synthetic oligonucleotides using [$\gamma^{32}P$]ATP and a phosphatase-dead mutant of T4 polynucleotide kinase (Pnkp-D167N). The radiolabeled strands were gel-purified prior to use in enzyme assays.

RESULTS

Myxococcus xanthus encodes six RtcB paralogs. M. xanthus is a Deltaproteobacterium of the order *Myxococcales* that is distinguished by its "social" behavior and starvation-induced differentiation into a multicellular fruiting body. The 9.14 Mbp single-chromosome genome of *M. xanthus* strain DK1622 encodes 7388 predicted proteins, 3532 of which comprise 872 paralogous sets (defined by having at least two members) (Goldman et al., 2006). It is suggested that selective gene duplications and divergence of paralogs enabled evolution of the multicellular lifestyle of *Myxococcus* (Goldman et al., 2006). In this vein, it is notable that the *M. xanthus* proteome includes six RtcB-like paralogs, which we name as follows: RtcB1 (411-aa, Genbank accession ABF90081); RtcB2 (384-aa, ABF90287); RtcB3 (385-aa, ABF86333); RtcB4 (474-aa, ABF92494); RtcB5 (497-aa, ABF87339); and RtcB6 (450-aa, ABF85892). Another social *Myxococcales* species, *Sorangium cellulosum*, has a 13 Mbp genome that encodes 9367 predicted proteins (Schneiker et al., 2007), including five RtcB paralogs. By contrast, other Deltaproteobacteria such as *Geobacter metallireducens*, *Bdellovibrio bacteriovorus*, and *Desulfovirvio vulgaris* encode a single RtcB homolog.

A COBALT-based alignment of the amino acid sequences of the *M. xanthus* RtcBs highlights 77 positions of side chain identity/similarity in all six proteins (Fig. 3.1). The conserved amino acids embrace virtually all of the components imputed to the RtcB active site based on the crystal structure of a *Pyrococcus horikoshii* RtcB•GTPαS•$(Mn^{2+})_2$ complex and mutational analyses of *E. coli* and *P. horikoshii* RtcBs (Chakravarty et al., 2012; Desai et al., 2013; M. Englert et al., 2012; Tanaka et al., 2011b). These include: (i) the histidine nucleophile that forms the covalent RtcB-

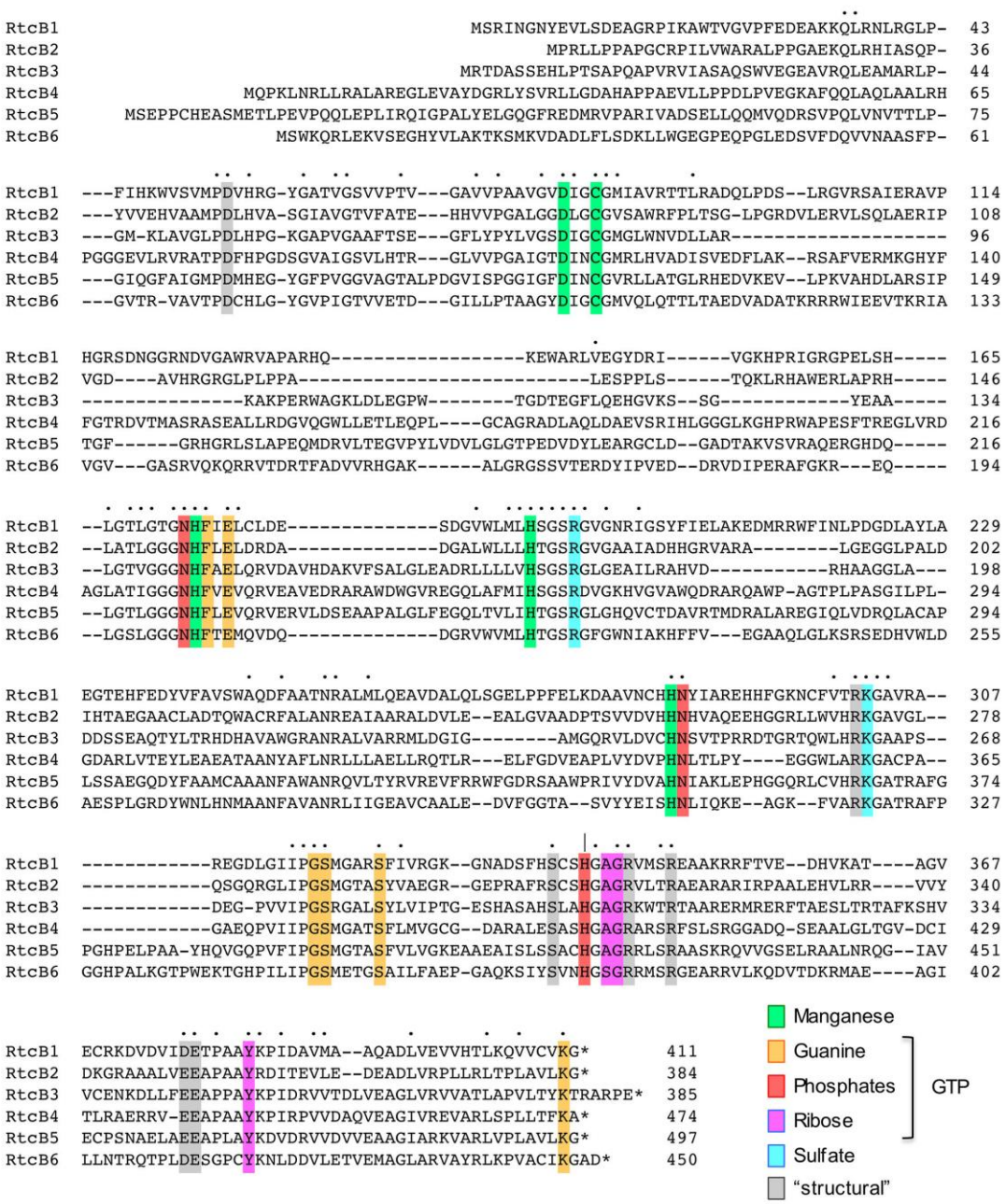


Figure 3.1. ***Myxococcus xanthus*** **encodes six RtcB paralogs**. (A) A COBALT-based alignment of the amino acid sequences of the six *M. xanthus* RtcB paralogs is shown. Positions of amino acid side chain identity/similarity in all six polypeptides are denoted by dots. The histidine nucleophile that becomes covalently linked to GMP is denoted by |. The conserved residues that, in the crystal structure of the *Pyrococcus* RtcB•GTPαS•$(Mn^{2+})_2$ complex contact the two manganese ions; the guanine nucleobase, ribose and phosphate moieties of GTP; or sulfate ions (putative mimetics of RNA phosphate groups) are highlighted in color-coded boxes indicated at bottom right. Conserved "structural" residues in gray shading are those that either coordinate active site residues or tether secondary structure elements of the RtcB fold, e.g., via salt bridges.

(histidinyl)–GMP intermediate (denoted by | in Fig. 3.1); (ii) the Asp-Cys-His-His-His pentad that coordinates two manganese ions; (iii) six residues that contact the guanine nucleobase of GTP; (iv) three residues that contact the GTP ribose; and (v) two asparagines that coordinate the GTP phosphates (Fig. 3.1). A conserved arginine and a lysine correspond to residues in *P. horikoshii* RtcB that coordinate sulfate anions proposed to mimic RNA phosphates. Also conserved are multiple amino acids that play a "structural" role, either by coordinating active site residues (e.g., an aspartate near the N-terminus that orients the histidine nucleophile) or by tethering secondary structure elements of the RtcB fold (e.g., via salt bridges).

The six RtcB paralogs are dispersed in the *M. xanthus* chromosome in six entirely distinct genetic neighborhoods. *M. xanthus rtcB1*, like *E. coli rtcB* (Genschik et al., 1998), is located in a ⇦*rtcR*•*rtcB*–*rtcA*⇨ RNA repair operon in which: (i) the ORF encoding the RNA ligase RtcB is oriented tail-to head with the ORF encoding the RNA 3'-phosphate cyclase enzyme RtcA, and (ii) a σ^{54} co-activator RtcR is encoded by the *rtcR* gene, located immediately upstream of *rtcBA* and transcribed in the opposite orientation. *M. xanthus rtcB5* is adjacent and distal to a co-oriented ORF encoding a protein called archease. Genetic clustering of RtcB or RtcA with archease is common in archea and is also seen sporadically in bacteria. *Pyrococcus horikoshii* and mammalian archeases have been shown to stimulate the ligase activities of *P. horikoshii* and mammalian RtcBs, respectively (Desai et al., 2014b; Popow et al., 2014). The *M. xanthus rtcB2*, *rtcB3*, *rtcB4*, and *rtcB6* ORFs have distinctive flanking gene sets.

Recombinant M. xanthus RtcB proteins. The primary structures of the six *M. xanthus* RtcBs suggest that they might all be enzymatically active. The question is: what

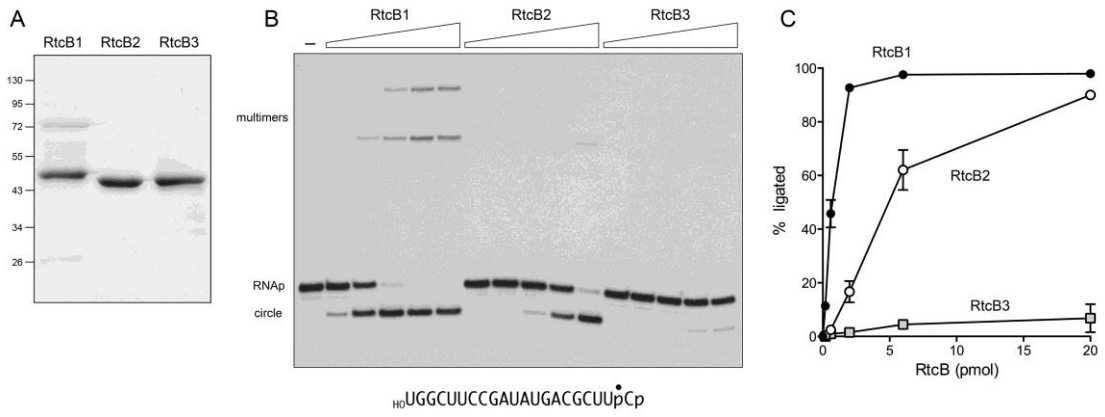


Figure 3.2. **RNA 3'-PO$_4$/5'-OH ligase activity.** (A) Recombinant RtcB1, RtcB2 and RtcB3 proteins. Aliquots (5 µg) of the indicated protein preparations were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (kDa) of marker polypeptides are indicated at *left.* (B) RNA ligase reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 0.1 mM GTP, 1 pmol (0.1 µM) 3' ^{32}P-labeled 20-mer $_{HO}$RNAp substrate (depicted at the bottom of the panel with the ^{32}P-label denoted by ●), and 0.2, 0.6, 2, 6, or 20 pmol RtcB protein (0.02, 0.06, 0.2, 0.6, or 2 µM RtcB; proceeding from left to right in each titration series) were incubated for 10 min at 37°C. RtcB was omitted from a control reaction in lane –. The reactions were quenched with an equal volume of 90% formamide/50 mM EDTA. The products were analyzed by electrophoresis through a 40-cm 20% polyacrylamide gel containing 7.5 M urea in 90 mM Tris·borate, 2 mM EDTA. An autoradiograph of the gel is shown. The position and identities of the radiolabeled RNAp substrate and the various sealed products are indicated on the *left*. (C) The substrate and ligated products were quantified by scanning the gel with a Fujix BAS2500 imager. The extents (%) of ligation, calculated as [(circle + multimers)/(circle + multimers + RNAp)] x 100, are plotted as a function of input RtcB. Each datum is the average of four separate titration experiments ±SEM.

activities inhere to each paralog and how might they differ? To address this issue, we expressed the six RtcB proteins in *E. coli* as His_{10}Smt3•RtcB fusions. The RtcB1, RtcB2, and RtcB3 proteins were isolated from soluble bacterial extracts by Ni-affinity chromatography. The His_{10}Smt3 tag was removed with the Smt3-specific protease Ulp1 and the native RtcB1, RtcB2, and RtcB3 proteins were separated from the tag by a second round of Ni-affinity chromatography. The tag-free RtcBs were further purified by gel filtration. SDS-PAGE of the purified RtcB2 and RtcB3 proteins revealed

predominant ~44 kDa polypeptides, consistent with their similar lengths: 384-aa and 385-aa, respectively (Fig. 3.2A). The RtcB1 preparations contained a major ~48 kDa polypeptide, as expected for the 411-aa RtcB protein, as well as a ~73 kDa contaminant (Fig. 3.2A). The RtcB4, RtcB5, and RtcB6 proteins were produced in *E. coli* but were recovered exclusively in the insoluble pellet fraction, thereby impeding their purification and characterization.

Test of RNA ligase activity. We reacted the recombinant *M. xanthus* RtcBs with a 20-mer $_{HO}$RNAp substrate that was 3' ^{32}P-labeled at the penultimate phosphate (Fig. 3.2B), the same substrate used previously to characterize the RNA sealing reaction of *E. coli* RtcB (Chakravarty et al., 2012). When the products were analyzed by denaturing PAGE, we found that *M. xanthus* RtcB1 efficiently converted the linear substrate to a more rapidly migrating circular RNA species as a consequence of intramolecular ligation; more slowly migrating multimers (products of inter-molecular ligation) were also generated (Fig. 3.2B). The extent of ligation by RtcB1 was proportional to input enzyme, attaining 98% at saturation (Fig. 3.2C). RtcB2 catalyzed intra- and inter-molecular RNA sealing (Fig. 3.2B) to an extent of 90% of input substrate, albeit 5-fold less effectively than RtcB1 on a per enzyme basis in the linear range of the titration (Fig. 3.2C). By contrast, RtcB3 was a notably feeble RNA ligase (Fig. 3.2B), sealing only 7% of the substrate at the highest level of input enzyme, and displaying a specific activity 130-fold lower than that of RtcB1 (Fig. 3.2C). Thus, the three RtcB paralogs are not equivalent with respect to their 3'-PO_4/5'-OH RNA sealing activities.

Tests of DNA ligase and DNA 3' capping activities. A prior study (Das et al., 2013) had shown that *E. coli* RtcB could splice DNA 3'-PO_4 and 5'-OH ends in the context of the

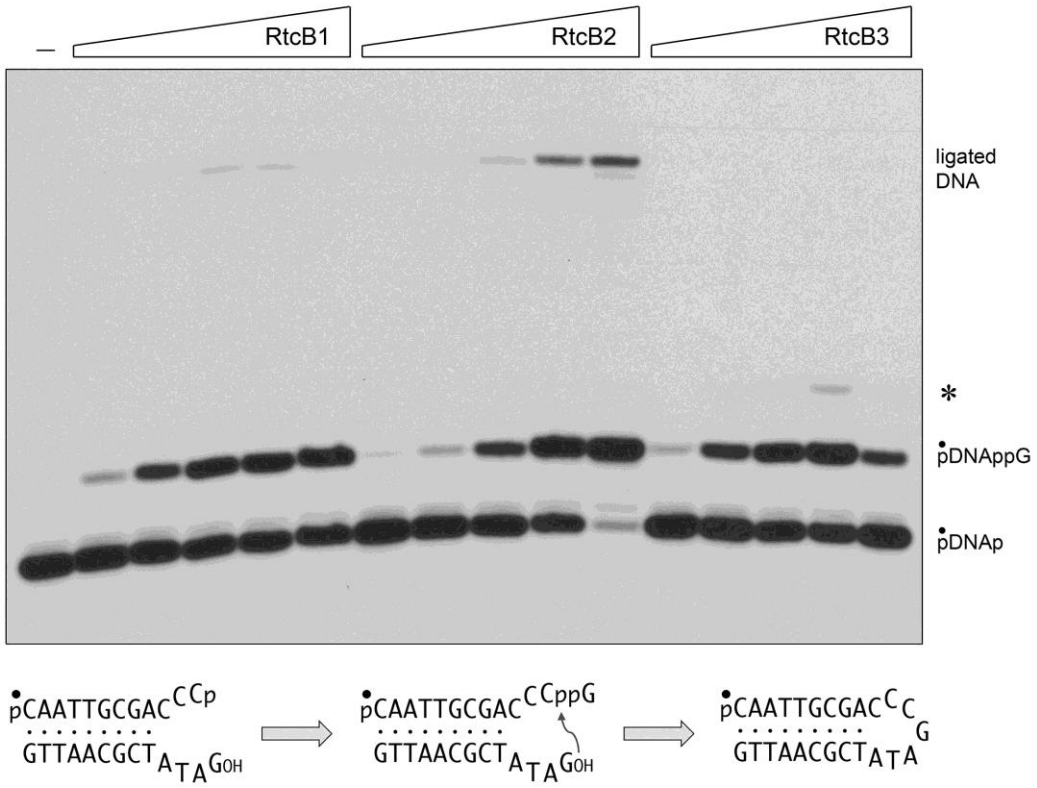


Figure 3.3. **Test of DNA 3'-PO$_4$/5'-OH ligase activity.** Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 0.1 mM GTP, 1 pmol (0.1 µM) 5' ^{32}P-labeled broken DNA stem-loop substrate (depicted at *bottom* with the ^{32}P-label denoted by ●), and 0.2, 0.6, 2, 6, or 20 pmol RtcB protein (0.02, 0.06, 0.2, 0.6, or 2 µM RtcB; proceeding from left to right in each titration series) were incubated for 10 min at 37°C. RtcB was omitted from a control reaction in lane –. The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown. The position and identities of the radiolabeled pDNAp substrate strand, the 3' capped pDNAppG strand, and the ligated DNA stem loop are indicated on the *right*. An RtcB3 product migrating above pDNAppG is indicated by [.

broken DNA stem-loop structure depicted in Fig. 3.3. DNA splicing proceeds via a chemically stable DNAppG intermediate formed by transfer of GMP from RtcB–GMP to the DNAp substrate strand (Chakravarty et al., 2012; Das et al., 2013). The DNAppG species migrates more slowly than DNAp when analyzed by denaturing PAGE (Das et al., 2013). Here, when we tested the *M. xanthus* RtcB proteins for DNA sealing, we noted that whereas they were all capable of guanylylating the 5' ^{32}P-labeled 12-mer

pDNAp strand of the stem-loop to form a more slowly migrating capped pDNAppG strand, only RtcB1 and RtcB2 generated detectable amounts of a ligated DNA product (Fig. 3.3), which comprised 8% and 18% of the radiolabeled DNA at saturating levels of RtcB1 and RtcB2, respectively. RtcB1, RtcB2 and RtcB3 were also adept at capping the 3'-PO_4 of the pDNAp single strand substrate (Fig. 3.4). The RtcB3 reactions with the DNA stem-loop and pDNAp single strand substrates were distinguished by the

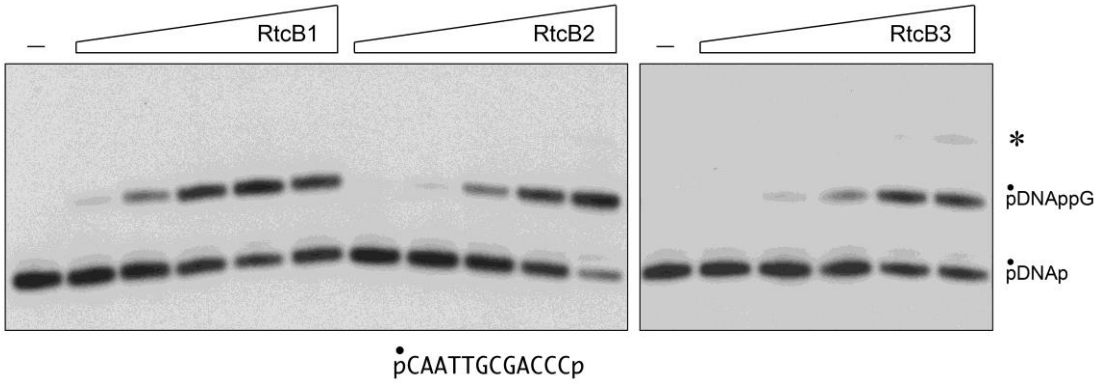


Figure 3.4. **DNA 3'-PO_4 capping activity.** Reaction mixtures (10 µl) containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 0.1 mM GTP, 1 pmol (0.1 µM) 5' ^{32}P-labeled 12-mer pDNAp substrate (depicted at *bottom* with the ^{32}P-label denoted by ●), and 0.2, 0.6, 2, 6, or 20 pmol RtcB protein (0.02, 0.06, 0.2, 0.6, or 2 µM RtcB; proceeding from left to right in each titration series) were incubated for 10 min at 37°C. RtcB was omitted from a control reaction in lane –. The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown. The position and identities of the radiolabeled pDNAp substrate strand and the 3' capped pDNAppG strand are indicated on the *right*. An RtcB3 product migrating above pDNAppG is indicated by [.

formation of a minor radiolabeled species ([in Figs. 3.3 and 3.4) that migrated above the pDNAppG strand.

RtcB3 can cap DNA 3'-PO4 and 5'-PO4 ends. In the experiment shown in Fig. 3.5A, 1 µM RtcB was reacted with 0.1 µM 5' ^{32}P-labeled 12-mer pDNAp strand; aliquots were withdrawn at 0.5, 1, 2, 5, 10, and 20 min and split into two halves that were either mock-

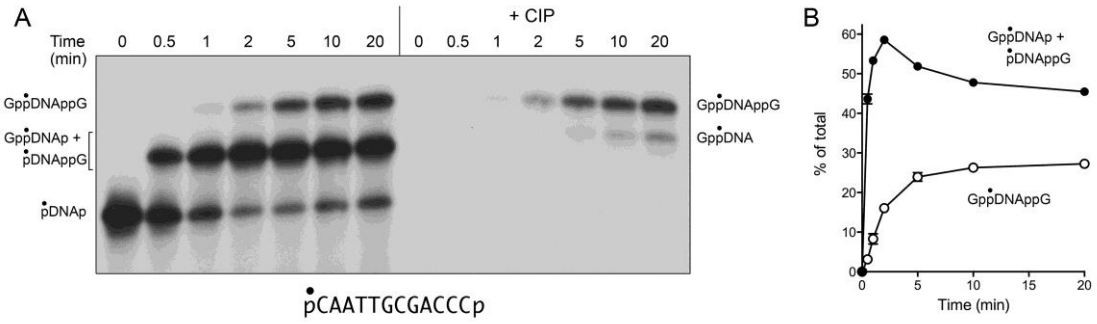


Figure 3.5. **5'-PO_4 capping activity of RtcB3.** (A) Reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 2 mM $MnCl_2$, 0.1 mM GTP, 0.1 µM 5' ^{32}P-labeled 12-mer pDNAp substrate (depicted at *bottom* with the ^{32}P-label denoted by ●), and 1 µM RtcB3 were incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and mixed with 8 µl of 2.5x CIP buffer (125 mM NaCl, 25 mM $MgCl_2$, 5 mM $ZnCl_2$). The samples were split into equal aliquots, which were then incubated at 37°C for 30 min either with no further additions (series at *left*), or after supplementation with 1 µl (10 units) calf intestine phosphatase (+CIP series at *right*). The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown. The positions and identities of the radiolabeled pDNAp substrate strand and the singly and doubly capped strands are indicated on the *left* and *right*. (B) The product distributions from no-CIP control reactions are plotted as a function of time. Each datum is the average of three separate experiments ±SEM.

treated or digested with calf intestine alkaline phosphate (CIP) before analyzing the samples by urea-PAGE. In the control reaction, the pDNAp strand was converted into two electrophoretically resolved products (Fig. 3.5A, left). As expected, CIP treatment eliminated the radiolabeled pDNAp substrate strand, by hydrolyzing the ^{32}P phosphomonoester (Fig. 3.5A, right). CIP treatment also eliminated the majority of the radiolabel in the "lower" product (Fig. 3.5A, right), signifying that the predominant species was 3' capped pDNAppG. The instructive finding was that the "upper" radiolabeled product was refractory to CIP, i.e., the labeled 5'-PO_4 was guanylylated by RtcB3 to form a doubly capped strand, GppDNAppG. A second CIP-resistant species – that migrated faster than GppDNAppG, and slightly slower than pDNAppG – corresponded to a $GppDNA_{OH}$ strand generated after CIP hydrolyzed the 3'-PO_4 of a 5'

capped GppDNAp RtcB3 reaction product. Thus, the lower product in the untreated control reaction actually comprised a mixture of co-migrating pDNAppG and GppDNAp strands, with the former predominating. [Note that CIP conversion of 3'-PO_4 to 3'-OH is expected to slow the electrophoretic mobility of an otherwise identical DNA strand.] The kinetic profile of the RtcB3 reaction suggests that the ends are capped sequentially, with most events entailing 3' capping prior to 5' capping (Fig. 3.5A and B).

5' capping of pDNAOH and pRNAOH ends by RtcB3. To focus exclusively on 5' capping, and to query whether 5' capping is contingent on prior 3' capping, we reacted RtcB3 with a 5' ^{32}P-labeled 12-mer $pDNA_{OH}$ strand and observed the conversion of pDNA to a single slower migrating GppDNA product, the yield of which was optimal in Tris buffer at pH 7.0 to 7.5 (Fig. 3.6A). No 5' capped product was formed when either GTP or manganese was omitted from the reaction mixture (not shown). In the experiments in Fig. 3.6B, we reacted RtcB3 with a series of 5' ^{32}P-labeled 24-mer, 12-mer, and 6-mer $pDNA_{OH}$ strands and tracked the kinetic profile of the conversion of pDNA to GppDNA product. The rate and yield of 5' capping were virtually identical for the 24-mer and 12-mer pDNA substrates; the extent of capping of the 6-mer pDNA was reduced slightly (Fig. 3.6B). A 12-mer $pRNA_{OH}$ substrate was capped half as well as a 12-mer pDNA of identical nucleotide sequence (Fig. 3.6B). We conclude that RtcB performs 5'-PO_4 polynucleotide capping, without requirement for a 3'-PO_4 or its guanylylation.

The capped 12-mer DNA product formed by RtcB3 was gel-purified to radiochemical homogeneity (Fig. 3.7A, compare lanes M and 0) and subjected to

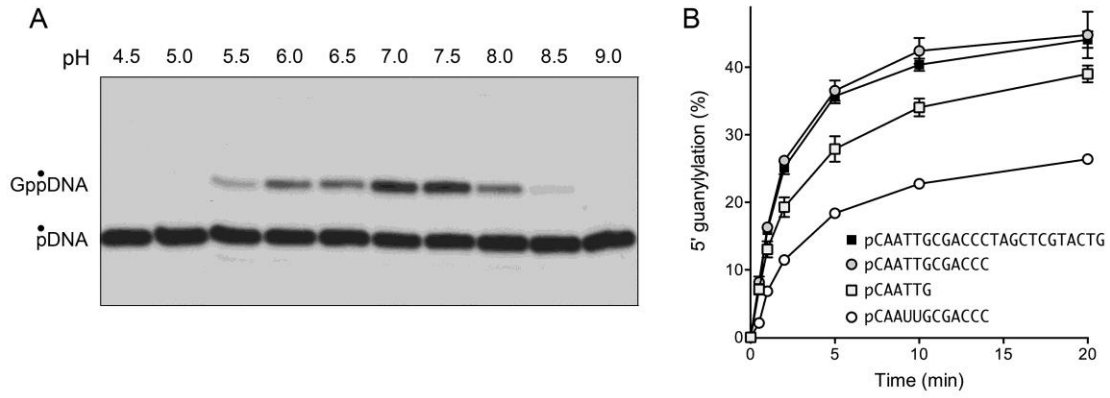
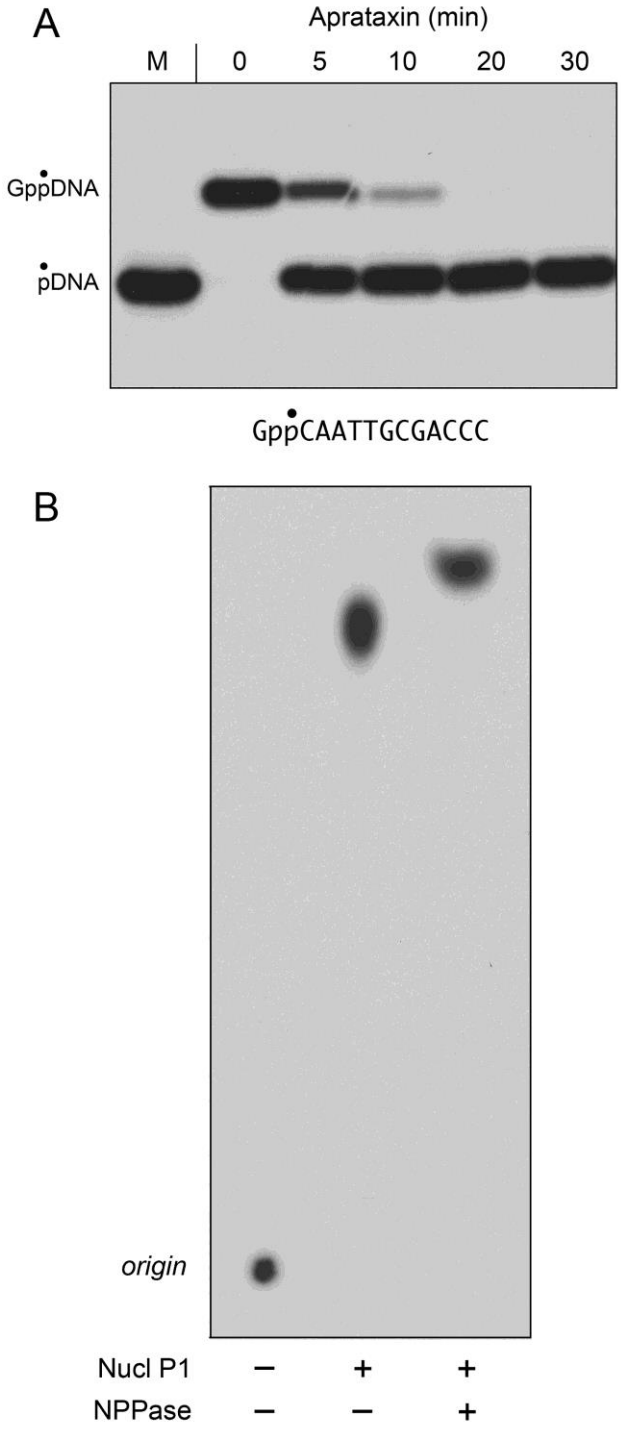


Figure 3.6. **DNA and RNA 5'-PO_4 capping by RtcB3**. (A) Reaction mixtures (10 µl) containing either 50 mM Tris-acetate (pH 4.5, 5.0, 5.5, 6.0, 6.5) or Tris-HCl (pH 7.0, 7.5, 8.0, 8.5, 9.0) as specified, 2 mM $MnCl_2$, 0.1 mM GTP, 1 pmol (0.1 µM) 5' ^{32}P-labeled 12-mer $pDNA_{OH}$ substrate (shown in panel C), and 10 pmol (1 µM) RtcB3 were incubated at 37°C for 20 min. The products were analyzed by urea-PAGE. An autoradiograph of the gel is shown. The position and identities of the radiolabeled pDNA substrate strand and the 5' capped strand are indicated on the *left*. (B) Reaction mixtures containing 50 mM Tris-HCl, pH 7.5, 2 mM $MnCl_2$, 0.1 mM GTP, either 0.1 µM 5' ^{32}P-labeled 24-mer, 12-mer or 6-mer $pDNA_{OH}$ substrates or 0.1 µM 5' ^{32}P-labeled 12-mer $pRNA_{OH}$ substrate (with nucleotide sequences as shown), and 1 µM RtcB3 were incubated at 30°C. Aliquots (10 µl) were withdrawn at the times specified and quenched with formamide/EDTA. The products were analyzed by urea-PAGE and quantified by scanning the gel. The extents of 5' capping are plotted as a function of time for each substrate. Each datum is the average of three separate experiments ±SEM.

digestion with either nuclease P1 alone or sequential digestions with nuclease P1 and nucleotidyl pyrophosphatase (NPPase). The intact capped 12-mer and the digestion products were analyzed by PEI-cellulose TLC (Fig. 3.7B). Whereas the intact 12-mer DNA remained at the chromatographic origin, nuclease P1 treatment liberated a radiolabeled cap dinucleotide that, upon further treatment with NPPase, was converted to a more rapidly migrating ^{32}P-labeled species that co-migrated with a cold CMP standard. These results fortify our designation of the RtcB reaction product as 5' capped GppDNA.

Figure 3.7. **De-capping of GppDNA by aprataxin or nucleotidyl pyrophosphatase**. (A) A ^{32}P-labeled 12-mer GppCAATTGCGACCC strand generated by RtcB3 was gel-purified and used as a substrate for purified recombinant fission yeast aprataxin. Reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 40 mM NaCl, 3 nM GppDNA, and 0.6 μM aprataxin were incubated at 37˚C. Aliquots (10 μl) were withdrawn at the times specified and quenched with formamide/EDTA. The products were analyzed by urea-PAGE, in parallel with an aliquot of the 5' ^{32}P-labled pCAATTGCGACCC strand (lane M). An autoradiograph of the gel is shown. (B) ^{32}P-labeled 12-mer GppCAATTGCGACCC strand (60 fmol) was incubated for 20 min at 50˚C with 0.002 units nuclease P1 (Sigma) in a 10 μl reaction mixture containing 50 mM NaOAc, pH 5.2, and 1 mM $ZnCl_2$. A duplicate sample was digested with nuclease P1, then adjusted to 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM NaCl, supplemented with 0.001 unit nucleotidyl pyrophosphatase (NPPase; Sigma), and incubated for an additional for 30 min at 37°C. The digestions were quenched by adding 2 μl of 5 M formic acid. Aliquots of the nuclease P1 and nuclease P1 plus NPPase digests, and of an undigested GppDNA control, were spotted on a polyethyleneimine (PEI)-cellulose TLC plate (Merck) alongside unlabeled marker nucleotide CMP. Ascending TLC was performed with 0.75 M LiCl as the mobile phase. An autoradiograph of the TLC plate is shown, with the position of the origin indicated at *left*


A
Aprataxin (min)
M
0
5
10
20
30
GppDNA
pDNA
GppCAATTGCGACCC
B
origin
Nucl P1 – + +
NPPase – – +

De-capping of GppDNA by aprataxin. The DNA repair enzyme aprataxin has a DNA 3' de-capping activity, whereby it converts DNAppG (synthesized by *E. coli* RtcB) to DNAp and GMP (Chauleau et al., 2015; Das et al., 2014). Aprataxin is a member of the histidine triad family of nucleotidyltransferases that act via a covalent enzyme-(histidinyl)–NMP intermediate. Mutations of aprataxin are the cause of the human neurological disorder ataxia oculomotor apraxia-1 (Date et al., 2001; Moreira et al., 2001). Aprataxin was initially shown to de-adenylate abortive AppDNA intermediates that can accumulate when classic DNA ligases attempt to seal DNA 5'-PO_4 ends at sites of damage or RNA 5'-PO_4 ends embedded in DNA (Ahel et al., 2006; Tumbale et al., 2014). Its dual capacity to de-adenylate a 5' capped AppDNA and de-guanylate a 3' capped DNAppG indicate that aprataxin is eclectic in its substrate specificity. Crystal structures of *Schizosaccharomyces pombe* aprataxin with AMP or GMP in the active site revealed that AMP and GMP bind at the same position and in the same *anti* nucleoside conformation, and that aprataxin makes more extensive nucleobase contacts with guanine than with adenine (Chauleau et al., 2015; Gong et al., 2011; Tumbale et al., 2011). Because RtcB3 is the first instance of an enzyme that caps a DNA 5'-PO_4 end with GMP, we were interested in testing whether aprataxin could de-guanylylate the 12-mer GppDNA strand produced by RtcB3. As shown in Fig. 3.7A, recombinant *S. pombe* aprataxin quantitatively converted ^{32}P-labeled GppDNA to pDNA in a time-dependent fashion.

DISCUSSION

The present characterization of three RtcB paralogs from *M. xanthus* extends our knowledge of the bacterial branch of this enzyme family beyond the single case of *E. coli* RtcB that was studied previously. The instructive findings are that all RtcBs are not created equal anent their RNA splicing activity *in vitro*. We find that *M. xanthus* RtcB1 resembles *E. coli* RtcB in its ability to perform intra- and inter-molecular sealing of a $_{HO}$RNAp substrate and capping of a DNA 3'-PO_4 end. The resemblance extends to the genetic organization of *M. xanthus rtcB1* in a ⇦*rtcR•rtcB–rtcA*⇨ operon and the high degree of structural conservation of *M. xanthus* RtcB1 and *E. coli* RtcB (243 positions of amino acid identity plus 35 positions of side chain similarity). *M. xanthus* RtcB2 can splice RNA but has five-fold lower RNA ligase specific activity compared to RtcB1 under the assay conditions employed. *M. xanthus* RtcB2 is less similar to *E. coli* RtcB (119 positions of amino acid identity plus 54 positions of side chain similarity).

By contrast, *M. xanthus* RtcB3 is distinctively feeble at ligating the $_{HO}$RNAp substrate, though it caps a DNA 3'-PO_4 end. It is conceivable that RtcB3 is uniquely fastidious in its RNA substrate requirement (e.g., for a specific RNA sequence, secondary structure, or three-dimensional fold). Alternatively, the ligase function of *M. xanthus* RtcB3 might rely on a specific partner protein. *M. xanthus* RtcB3 is most closely related to a separate branch of bacterial RtcBs, referred to as release factor H-coupled RtcBs because they are encoded by genes that are physically clustered with a bacterial genes encoding PrfH, a homolog of class-I protein release factors that terminate translation (Baranov et al., 2006). PrfH-coupled RtcBs are found in the

proteomes of *Salmonella*, *Pseudomonas*, and *Ralstonia* and have 50-55% amino acid identity to *M. xanthus* RtcB3. Certain strains of *E. coli* also encode a PrfH-coupled RtcB paralog, in addition to the RtcB encoded by the *rtcBA* operon. Indeed, *M. xanthus* RtcB3 is more distantly related to the *E. coli* RtcB encoded by the *rtcBA* operon (with which it shares 117 positions of amino acid identity and 50 positions of side chain similarity).

The novelty of *M. xanthus* RtcB3 is its capacity to cap DNA or RNA 5'-PO_4 ends to form GppDNA and GppRNA products. This reaction differs from that of canonical mRNA capping enzymes, which transfer GMP from GTP, via a covalent enzyme-(lysyl-Nζ)-GMP intermediate, to a 5'-diphosphate terminus to form a GpppRNA product (Håkansson et al., 1997; Shuman & Hurwitz, 1981). Although the prevalence of GppDNA and GppRNA caps *in vivo* is uncharted territory, one can imagine that 5' capping of nucleic acids by RtcB3-like enzymes might serve useful functions, e.g., protecting DNA/RNA ends from 5' exonucleases and/or installing a modification mark that directs downstream transactions. Indeed, recent studies have shown that a subset of bacterial RNAs are capped at their 5' ends with NAD^+ (reviewed in Luciano & Belasco, 2015), although the biochemical mechanism of NAD^+ capping has not yet been determined.

RtcB3 is revealed by the present study to be versatile in capping either 3'-PO_4 or 5'-PO_4 ends. As such RtcB3 joins a growing list of enzymes that can cap both ends of a polynucleotide substrate. Covalent adenylylation of an RNA 3'-PO_4 end to form RNAppA was first described as a fleeting intermediate in the synthesis of an RNA 2',3'-cyclic phosphate terminus by RtcA, which transfers AMP from ATP via a covalent RtcA-(histidinyl)–AMP intermediate (Chakravarty et al., 2011; Filipowicz et al., 1985).

Subsequent studies showed that RtcA catalyzes highly efficient adenylylation of RNA and DNA 5'-PO_4 ends to form stable AppRNA and AppDNA end products (Chakravarty & Shuman, 2011). Thus, there is considerable plasticity in the direction of the RtcA pathway (2',3' cyclization *versus* 5' A capping), with resulting uncertainty as to what the real substrates of RtcA are. This theme was amplified by the report that certain classic ATP-dependent thermophilic RNA ligases, which join 3'-OH/5'-PO_4 ends via an AppRNA intermediate, are also capable of adenylylating a DNA 3'-PO_4 end to form a stable DNAppA product (Zhelkovsky & McReynolds, 2014). In effect, the thermophilic RNA ligases double as DNA 3' capping enzymes, albeit with an A cap rather than a G cap.

The theme of action at either end extends to de-capping of nucleic acids by aprataxin. Initially described as an adenylate de-capping enzyme acting on abortive AppDNA intermediates formed by ATP-dependent DNA ligases, aprataxin was subsequently shown to have 3' de-capping activity on DNAppG end formed by *E. coli* RtcB (Chauleau et al., 2015; Das et al., 2014) and DNAppA ends generated by an ATP-dependent thermophilic RNA ligases RNA ligase (Zhelkovsky & McReynolds, 2014). Here we extend the aprataxin repertoire by showing that it can remove a 5' guanylate cap from GppDNA formed by *M. xanthus* RtcB3.

CONCLUSIONS

In this thesis, I have described my efforts to broaden our understanding of RtcB, its mechanism, and its capabilities. I mutated the *E. coli* RtcB active site in order to parse the four-step RtcB ligation mechanism and determine which residues are responsible for catalysis of each step. I was able to determine the function of several residues in the *E. coli* RtcB active site. The clearest phenotypes resulted from the mutation of Asp75, Cys78, Arg189, and His281. Asp75, which contacts one manganese ion in the RtcB active site, is implicated in guanylylation transactions. Mutation of Asp75 to Ala resulted in severe impairment of RNA and DNA guanylylation as well as ablation of DNA de-guanylylation activity. Meanwhile, CPD and phosphodiester synthesis reactions were far less impaired. Mutation of Cys78 to Ala completely destroyed all RtcB-mediated catalysis, consistent with the distinction of Cys78 being the only residue to contact both manganese ions. Arg189 likely coordinates the phosphate immediately preceding the terminal 5'-sugar. Existing crystal structures show Arg189 bound to a sulfate anion, likely mimicking said phosphate. Additionally, mutation of Arg189 to Ala specifically slows the phosphodiester synthesis step, the only step that requires coordination of the 5'-end of the polynucleotide. Finally, His281, which is involved in coordination of the manganese which Asp75 does not contact, is implicated in CPD and polynucleotide 3'-end guanylylation, as these steps are slowed upon His281 mutation to alanine, while phosphodiester synthesis remains intact.

These findings have illuminated how RtcB achieves RNA and DNA ligation. However, no structure yet exists of RtcB bound to RNA or DNA. When obtained, this structure will dovetail with the findings contained herein and will provide its own further key insights as to the RtcB mechanism.

Equipped with the knowledge of the RtcB active site, I turned to *Myxococcus xanthus*, which encodes six paralogs of RtcB, each with the known key residues conserved. I wanted to know the diversity that RtcB is capable of in terms of either the reaction itself or the substrate specificity. Three of six RtcB paralogs were characterized. RtcB1 and RtcB2 were both capable of ligating RNA. RtcB3 is severely meek with regards to phosphodiester synthesis, however it displayed a unique phenotype in that it is capable of guanylylating both a 5'- and 3'-PO_4 of either an RNA or DNA strand, with a modest preference for guanylylation of DNA. The guanylylation of a 5'-PO_4 of a DNA strand has not been described before and its implications are yet to be detailed.

The results described above have illuminated a new field that opened with the 2011 discovery of RtcB as the 3'-PO_4/5'-OH ligase. Activation of a polynucleotide 3'-end is a novel concept in nucleic acid enzymology, as the only enzymes known to utilize 3'-end activation are RtcB and its operonal partner RtcA. I have shown the biochemical mechanism by which RtcB achieves ligation *via* 3'-end activation biochemically. I have also shown that the RtcB reaction can be extended to DNA as a preferred substrate for guanylylation and that RtcB is also capable of guanylylating the 5'-end of a polynucleotide. The latter discoveries came from a single enzyme in

Myxococcus xanthus. Three more paralogs remain to be characterized, and further insight into the capabilities of RtcB likely remain to be discovered.

REFERENCES

Ahel, I., Rass, U., El-Khamisy, S. F., Katyal, S., Clements, P. M., McKinnon, P. J., … West, S. C. (2006). The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*, *443*(7112), 713–716. http://doi.org/10.1038/nature05164

Arn, E. A., & Abelson, J. N. (1996). The 2'-5' RNA ligase of Escherichia coli. Purification, cloning, and genomic disruption. *Journal of Biological Chemistry*, *271*(49), 31145–31153. http://doi.org/10.1074/jbc.271.49.31145

Auxilien, S., El Khadali, F., Rasmussen, A., Douthwaite, S., & Grosjean, H. (2007). Archease from Pyrococcus abyssi improves substrate specificity and solubility of a tRNA m5C methyltransferase. *Journal of Biological Chemistry*, *282*(26), 18711–18721. http://doi.org/10.1074/jbc.M607459200

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., … Landthaler, M. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell*, *46*(5), 674–690. http://doi.org/10.1016/j.molcel.2012.05.021

Baranov, P. V, Vestergaard, B., Hamelryck, T., Gesteland, R. F., Nyborg, J., & Atkins, J. F. (2006). Diverse bacterial genomes encode an operon of two genes, one of which is an unusual class-I release factor that potentially recognizes atypical mRNA signals other than normal stop codons. *Biology Direct*, *1*, 28. http://doi.org/10.1186/1745-6150-1-28

Brooks, M. A., Meslet-Cladiére, L., Graille, M., Kuhn, J., Blondeau, K., Myllykallio, H., & van Tilbeurgh, H. (2008). The structure of an archaeal homodimeric ligase which has RNA circularization activity. *Protein Science : A Publication of the Protein Society*, *17*(8), 1336–1345. http://doi.org/10.1110/ps.035493.108

Bush, M., & Dixon, R. (2012). The role of bacterial enhancer binding proteins as specialized activators of σ54-dependent transcription. *Microbiology and Molecular Biology Reviews : MMBR*, *76*(3), 497–529. http://doi.org/10.1128/MMBR.00006-12

Chakravarty, A. K., & Shuman, S. (2011). RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like adenylylation of DNA and RNA 5'-monophosphate ends. *Journal of Biological Chemistry*, *286*(6), 4117–4122. http://doi.org/10.1074/jbc.M110.196766

Chakravarty, A. K., & Shuman, S. (2012). The sequential 2',3'-cyclic phosphodiesterase and 3'-phosphate/5'-OH ligation steps of the RtcB RNA splicing pathway are GTP-dependent. *Nucleic Acids Research*, *40*(17), 8558–8567. http://doi.org/10.1093/nar/gks558

Chakravarty, A. K., Smith, P., & Shuman, S. (2011). Structures of RNA 3'-phosphate cyclase bound to ATP reveal the mechanism of nucleotidyl transfer and metal-

assisted catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(52), 21034–21039. http://doi.org/10.1073/pnas.1115560108

Chakravarty, A. K., Subbotin, R., Chait, B. T., & Shuman, S. (2012). RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6072–7. http://doi.org/10.1073/pnas.1201207109

Chan, P. P., & Lowe, T. M. (2015). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, *44*(D1), D184–189. http://doi.org/10.1093/nar/gkv1309

Chauleau, M., Das, U., & Shuman, S. (2015). Effects of DNA3′pp5′G capping on 3′ end repair reactions and of an embedded pyrophosphate-linked guanylate on ribonucleotide surveillance. *Nucleic Acids Research* , *43* (6), 3197–3207. http://doi.org/10.1093/nar/gkv179

Chauleau, M., Jacewicz, A., & Shuman, S. (2015). DNA(3′)pp(5′)G de-capping activity of aprataxin: effect of cap nucleoside analogs and structural basis for guanosine recognition. *Nucleic Acids Research*, *43*(12), 6075–6083. http://doi.org/10.1093/nar/gkv501

Cox, J. S., & Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. *Cell*, *87*(3), 391–404. http://doi.org/S0092-8674(00)81360-4 [pii]

Cozzarelli, N. R., Melechen, N. E., Jovin, T. M., & Kornberg, A. (1967). Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. *Biochemical and Biophysical Research Communications*, *28*(4), 578–586. http://doi.org/10.1016/0006-291X(67)90353-1

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563. http://doi.org/10.1038/227561a0

Culver, G. M., Consaul, S. A., Tycowski, K. T., Filipowicz, W., & Phizicky, E. M. (1994). tRNA splicing in yeast and wheat germ. A cyclic phosphodiesterase implicated in the metabolism of ADP-ribose 1',2'-cyclic phosphate. *Journal of Biological Chemistry*, *269*(40), 24928–24934.

Das, U., Chakravarty, A. K., Remus, B. S., & Shuman, S. (2013). Rewriting the rules for end joining via enzymatic splicing of DNA 3'-PO4 and 5'-OH ends. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20437–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24218597

Das, U., Chauleau, M., Ordonez, H., & Shuman, S. (2014). Impact of DNA3'pp5'G capping on repair reactions at DNA 3' ends. *Proceedings of the National Academy of Sciences*, *111*(31), 11317–11322.

http://doi.org/10.1073/pnas.1409203111

Das, U., & Shuman, S. (2013). 2'-Phosphate cyclase activity of RtcA: a potential rationale for the operon organization of RtcA with an RNA repair ligase RtcB in Escherichia coli and other bacterial taxa. *Rna*, *19*, 1355–1362. http://doi.org/10.1261/rna.039917.113

Das, U., & Shuman, S. (2013). Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Research*, *41*(1), 355–365. http://doi.org/10.1093/nar/gks977

Das, U., Wang, L. K., Smith, P., Jacewicz, A., & Shuman, S. (2014). Structures of bacterial polynucleotide kinase in a Michaelis complex with GTP•Mg(2+) and 5′-OH oligonucleotide and a product complex with GDP•Mg(2+) and 5′-PO(4) oligonucleotide reveal a mechanism of general acid-base catalysis and the determinants of phosphoacceptor recognition. *Nucleic Acids Research*, *42*(2), 1152–1161. http://doi.org/10.1093/nar/gkt936

Date, H., Onodera, O., Tanaka, H., Iwabuchi, K., Uekawa, K., Igarashi, S., … Tsuji, S. (2001). Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nature Genetics*, *29*(2), 184–188. http://doi.org/10.1038/ng1001-184

Desai, K. K., Beltrame, A. L., & Raines, R. T. (2015). Coevolution of RtcB and Archease created a multiple-turnover RNA ligase. *RNA (New York, N.Y.)*, *21*(11), 1866–72. http://doi.org/10.1261/rna.052639.115

Desai, K. K., Bingman, C. a, Phillips, G. N., & Raines, R. T. (2013). Structures of the noncanonical RNA ligase RtcB reveal the mechanism of histidine guanylylation. *Biochemistry*, *52*(15), 2518–25.

Desai, K. K., Cheng, C. L., Bingman, C. A., Phillips, G. N., & Raines, R. T. (2014). A tRNA splicing operon: Archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation. *Nucleic Acids Research*, *42*(6), 3931–3942. http://doi.org/10.1093/nar/gkt1375

Desai, K. K., & Raines, R. T. (2012). TRNA ligase catalyzes the GTP-dependent ligation of RNA with 3'-phosphate and 5'-hydroxyl termini. *Biochemistry*, *51*(7), 1333–1335. http://doi.org/10.1021/bi201921a

Doherty, A. J., & Suh, S. W. (2000). Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research , 28* (21), 4051–4058. http://doi.org/10.1093/nar/28.21.4051

Eastberg, J. H., Pelletier, J., & Stoddard, B. L. (2004). Recognition of DNA substrates by T4 bacteriophage polynucleotide kinase. *Nucleic Acids Research*, *32*(2), 653–660. http://doi.org/10.1093/nar/gkh212

Englert, M., & Beier, H. (2005). Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other

phylogenetic origins. *Nucleic Acids Research*, *33*(1), 388–399. http://doi.org/10.1093/nar/gki174

Englert, M., Sheppard, K., Aslanian, A., Yates, J. R., & Söll, D. (2011). Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1290–1295.

Englert, M., Sheppard, K., Gundllapalli, S., Beier, H., & Söll, D. (2010). Branchiostoma floridae has separate healing and sealing enzymes for 5'-phosphate RNA ligation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 16834–16839. http://doi.org/10.1073/pnas.1011703107

Englert, M., Xia, S., Okada, C., Nakamura, A., Tanavde, V., Yao, M., … Wang, J. (2012). Structural and mechanistic insights into guanylylation of RNA-splicing ligase RtcB joining RNA between 3'-terminal phosphate and 5'-OH. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1213795109

Filipowicz, W., & Shatkin, A. J. (1983). Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. *Cell*, *32*(2), 547–557. http://doi.org/10.1016/0092-8674(83)90474-9

Filipowicz, W., Strugala, K., Konarska, M., & Shatkin, A. J. (1985). Cyclization of RNA 3'-terminal phosphate by cyclase from HeLa cells proceeds via formation of N(3')pp(5')A activated intermediate. *Proc Natl Acad Sci U S A*, *82*(5), 1316–1320. http://doi.org/10.1073/pnas.82.5.1316

Galperin, M. Y., & Koonin, E. V. (2004). "Conserved hypothetical"proteins: prioritization of targets for experimental study. *Nucleic Acids Research*, *32*(18), 5452–5463.

Gefter, M. L., Becker, A., & Hurwitz, J. (1967). The enzymatic repair of DNA. I. Formation of circular lambda-DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(1), 240–247. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC335624/

Gellert, M. (1967). Formation of covalent circles of lambda DNA by E. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(1), 148–155. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC335477/

Genschik, P., Drabikowski, K., & Filipowicz, W. (1998). Characterization of the Escherichia coli RNA 3'-terminal phosphate cyclase and its sigma54-regulated operon. *J Biol Chem*, *273*(39), 25516–25526. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9738023

Goldman, B. S., Nierman, W. C., Kaiser, D., Slater, S. C., Durkin, a S., Eisen, J. a, … Kaplan, H. B. (2006). Evolution of sensory complexity recorded in a

myxobacterial genome. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(41), 15200–15205. http://doi.org/10.1073/pnas.0607335103

Gong, Y., Zhu, D., Ding, J., Dou, C.-N., Ren, X., Gu, L., … Wang, D.-C. (2011). Crystal structures of aprataxin ortholog Hnt3 reveal the mechanism for reversal of 5'-adenylated DNA. *Nature Structural & Molecular Biology*, *18*(11), 1297–9. http://doi.org/10.1038/nsmb.2145

Gonzalez, T. N., Sidrauski, C., Dörfler, S., & Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. *The EMBO Journal*, *18*(11), 3119–3132. http://doi.org/10.1093/emboj/18.11.3119

Goodman, H. M., Olson, M. V, & Hall, B. D. (1977). Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5453–7. http://doi.org/10.1073/pnas.74.12.5453

Greer, C. L., Peebles, C. L., Gegenheimer, P., & Abelson, J. (1983). Mechanism of action of a yeast RNA ligase in tRNA splicing. *Cell*, *32*(2), 537–546. http://doi.org/10.1016/0092-8674(83)90473-7

Håkansson, K., Doherty, A. J., Shuman, S., & Wigley, D. B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell*, *89*(4), 545–53. http://doi.org/10.1016/S0092-8674(00)80236-6

Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nature Reviews Molecular Cell Biology*, *13*(2), 89–102. http://doi.org/10.1038/nrm3270

Ho, C. K., & Shuman, S. (2002). Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12709–12714. http://doi.org/10.1073/pnas.192184699

Ho, C. K., Wang, L. K., Lima, C. D., & Shuman, S. (2004). Structure and Mechanism of RNA Ligase. *Structure*, *12*(2), 327–339. http://doi.org/10.1016/S0969-2126(04)00023-1

Huang, S. N., & Pommier, Y. (2013). Cross-over of RNA 3′-phosphate ligase into the DNA world. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20354–20355. http://doi.org/10.1073/pnas.1320033110

Iwawaki, T., & Tokuda, M. (2011). Function of yeast and amphioxus tRNA ligase in IRE1alpha-dependent XBP1 mRNA splicing. *Biochemical and Biophysical Research Communications*, *413*(4), 527–531. http://doi.org/10.1016/j.bbrc.2011.08.129

Jurkin, J., Henkel, T., Nielsen, A. F., Minnich, M., Popow, J., Kaufmann, T., … Martinez, J. (2014). The mammalian tRNA ligase complex mediates splicing of XBP1 mRNA and controls antibody secretion in plasma cells. *The EMBO Journal*, *33*(24), 2922–36. http://doi.org/10.15252/embj.201490332

Kanai, A., Sato, A., Fukuda, Y., Okada, K., Matsuda, T., Sakamoto, T., … Tomita, M. (2009). Characterization of a heat-stable enzyme possessing GTP-dependent RNA ligase activity from a hyperthermophilic archaeon, Pyrococcus furiosus. *RNA (New York, N.Y.)*, *15*(3), 420–431. http://doi.org/10.1261/rna.1122109

Kanai, Y., Dohmae, N., & Hirokawa, N. (2004). Kinesin transports RNA: Isolation and characterization of an RNA-transporting granule. *Neuron*, *43*(4), 513–525. http://doi.org/10.1016/j.neuron.2004.07.022

Klassen, R., Paluszynski, J. P., Wemhoff, S., Pfeiffer, A., Fricke, J., & Meinhardt, F. (2008). The primary target of the killer toxin from Pichia acaciae is tRNA Gln. *Molecular Microbiology*, *69*(3), 681–697. http://doi.org/10.1111/j.1365-2958.2008.06319.x

Konarska, M., Filipowicz, W., Domdey, H., & Gross, H. J. (1981). Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature*, *293*(5828), 112–116. http://doi.org/10.1038/293112a0

Kosmaczewski, S. G., Edwards, T. J., Han, S. M., Eckwahl, M. J., Meyer, B. I., Peach, S., … Hammarlund, M. (2014). The RtcB RNA ligase is an essential component of the metazoan unfolded protein response. *EMBO Reports*, *15*(12), 1278–1285. http://doi.org/10.15252/embr.201439531

Lee, A.-H., Iwakoshi, N. N., & Glimcher, L. H. (2003). XBP-1 Regulates a Subset of Endoplasmic Reticulum Resident Chaperone Genes in the Unfolded Protein Response. *Molecular and Cellular Biology*, *23*(21), 7448–7459. http://doi.org/10.1128/MCB.23.21.7448-7459.2003

Lehman, I. R. (1974). DNA ligase: structure, mechanism, and function. *Science (New York, N.Y.)*, *186*(4166), 790–7. http://doi.org/10.1126/science.186.4166.790

Lu, Y., Liang, F. X., & Wang, X. (2014). A Synthetic Biology Approach Identifies the Mammalian UPR RNA Ligase RtcB. *Molecular Cell*, *55*(5), 758–770. http://doi.org/10.1016/j.molcel.2014.06.032

Luciano, D. J., & Belasco, J. G. (2015). NAD in RNA: Unconventional headgear. *Trends in Biochemical Sciences*. http://doi.org/10.1016/j.tibs.2015.03.004

Lykke-Andersen, J., & Garrett, R. A. (1997). RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO Journal*, *16*(20), 6290–6300. http://doi.org/10.1093/emboj/16.20.6290

Ma, H., Qi, M.-Y., Zhang, X., Zhang, Y.-L., Wang, L., Li, Z.-Q., … Liu, D. (2014). HSPC117 Is Regulated by Epigenetic Modification and Is Involved in the

Migration of JEG-3 Cells. *International Journal of Molecular Sciences*, *15*(6), 10936–10949. http://doi.org/10.3390/ijms150610936

McCraith, S. M., & Phizicky, E. M. (1990). A highly specific phosphatase from Saccharomyces cerevisiae implicated in tRNA splicing. *Molecular and Cellular Biology*, *10*(3), 1049–55. http://doi.org/10.1128/MCB.10.3.1049

Moreira, M.-C., Barbot, C., Tachi, N., Kozuka, N., Uchida, E., Gibson, T., … Koenig, M. (2001). The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nat Genet*, *29*(2), 189–193. Retrieved from http://dx.doi.org/10.1038/ng1001-189

Nandakumar, J., & Shuman, S. (2004). How an RNA ligase discriminates RNA versus DNA damage. *Molecular Cell*, *16*(2), 211–221. http://doi.org/10.1016/j.molcel.2004.09.022

Nandakumar, J., Shuman, S., & Lima, C. D. (2006). RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward. *Cell*, *127*(1), 71–84. http://doi.org/10.1016/j.cell.2006.08.038

Odell, M., Sriskanda, V., Shuman, S., & Nikolov, D. B. (2000). Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Molecular Cell*, *6*(5), 1183–1193. http://doi.org/10.1016/S1097-2765(00)00115-5

Okada, C., Maegawa, Y., Yao, M., & Tanaka, I. (2006). Crystal structure of an RtcB homolog protein (PH1602-extein protein) from Pyrococcus horikoshii reveals a novel fold. *Proteins*, *63*(4), 1119–22. http://doi.org/10.1002/prot.20912

Olivera, B. M., & Lehman, I. R. (1967). Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(5), 1426–1433. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC224490/

Pascal, J. M. (2008). DNA and RNA ligases: structural variations and shared mechanisms. *Current Opinion in Structural Biology*. http://doi.org/10.1016/j.sbi.2007.12.008

Pascal, J. M., O'Brien, P. J., Tomkinson, A. E., & Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature*, *432*(7016), 473–478. http://doi.org/10.1038/nature03082

Phizicky, E. M., Schwartz, R. C., & Abelson, J. (1986). Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. *Journal of Biological Chemistry*, *261*(6), 2978–2986.

Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., … Martinez, J. (2011). HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science (New York, N.Y.)*, *331*(6018), 760–764. http://doi.org/10.1126/science.1197847

Popow, J., Jurkin, J., Schleiffer, A., & Martinez, J. (2014). Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors. *Nature*, *511*(7507), 104–7. http://doi.org/10.1038/nature13284

Popow, J., Schleiffer, A., & Martinez, J. (2012). Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences*. http://doi.org/10.1007/s00018-012-0944-2

Ray, a., Zhang, S., Rentas, C., Caldwell, K. a., & Caldwell, G. a. (2014). RTCB-1 Mediates Neuroprotection via XBP-1 mRNA Splicing in the Unfolded Protein Response Pathway. *Journal of Neuroscience*, *34*(48), 16076–16085. http://doi.org/10.1523/JNEUROSCI.1945-14.2014

Remus, B. S., Jacewicz, A., & Shuman, S. (2014). Structure and mechanism of E. coli RNA 2′,3′-cyclic phosphodiesterase. *RNA*, *20*(11), 1697–1705. http://doi.org/10.1261/rna.046797.114

Remus, B. S., & Shuman, S. (2013). A kinetic framework for tRNA ligase and enforcement of a 2'-phosphate requirement for ligation highlights the design logic of an RNA repair machine. *RNA*, *19*(5), 659–669. http://doi.org/rna.038406.113 [pii]\r10.1261/rna.038406.113

Sawaya, R., Schwer, B., & Shuman, S. (2003). Genetic and biochemical analysis of the functional domains of yeast tRNA ligase. *The Journal of Biological Chemistry*, *278*(45), 43928–43938. http://doi.org/10.1074/jbc.M307839200

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., … Müller, R. (2007). Complete genome sequence of the myxobacterium Sorangium cellulosum. *Nature Biotechnology*, *25*(11), 1281–1289. http://doi.org/10.1038/nbt1354

Sekine, S., Bessho, Y., & Yokoyama, S. (n.d.). Crystal structure of the RtcB-like protein from Thermus thermophilus. *To Be Published*. http://doi.org/10.2210/pdb2epg/pdb

Shamu, C. E., & Walter, P. (1996). Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. *The EMBO Journal*, *15*(12), 3028–3039.

Shuman, S. (2009). DNA ligases: Progress and prospects. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.R900017200

Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews. Microbiology*, *5*(11), 852–861. http://doi.org/10.1038/nrmicro1768

Shuman, S., & Hurwitz, J. (1981). Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme--guanylate intermediate. *Proc Natl Acad Sci U S A*, *78*(1), 187–191. http://doi.org/10.1073/pnas.78.1.187

Sidrauski, C., Cox, J. S., & Walter, P. (2016). tRNA Ligase Is Required for Regulated

mRNA Splicing in the Unfolded Protein Response. *Cell*, *87*(3), 405–413. http://doi.org/10.1016/S0092-8674(00)81361-6

Silber, R., Malathi, V. G., & Hurwitz, J. (1972). Purification and Properties of Bacteriophage T4-Induced RNA Ligase. *Proceedings of the National Academy of Sciences* , *69* (10), 3009–3013. Retrieved from http://www.pnas.org/content/69/10/3009.abstract

Smith, P., Wang, L. K., Nair, P. A., & Shuman, S. (2012). The adenylyltransferase domain of bacterial Pnkp defines a unique RNA ligase family. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(7), 2296–2301. http://doi.org/10.1073/pnas.1116827109

Spinelli, S. L., Malik, H. S., Consaul, S. A., & Phizicky, E. M. (1998). A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14136–14141. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC24339/

Sriskanda, V., & Shuman, S. (1998). Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Research*, *26*(15), 3536–3541. http://doi.org/10.1093/nar/26.15.3536

Subramanya, H. S., Doherty, A. J., Ashford, S. R., & Wigley, D. B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, *85*(4), 607–615. http://doi.org/10.1016/S0092-8674(00)81260-X

Tanaka, N., Chakravarty, A. K., Maughan, B., & Shuman, S. (2011). Novel Mechanism of RNA Repair by RtcB via Sequential 2',3'-Cyclic Phosphodiesterase and 3'-Phosphate/5'-Hydroxyl Ligation Reactions. *Journal of Biological Chemistry*.

Tanaka, N., Meineke, B., & Shuman, S. (2011b). RtcB, a Novel RNA Ligase, Can Catalyze tRNA Splicing and HAC1 mRNA Splicing in Vivo. *The Journal of Biological Chemistry*, *286*(35), 30253–30257.

Tanaka, N., & Shuman, S. (2011a). RtcB is the RNA ligase component of an Escherichia coli RNA repair operon. *Journal of Biological Chemistry*, *286*(10), 7727–7731. http://doi.org/10.1074/jbc.C111.219022

Tumbale, P., Appel, C. D., Kraehenbuehl, R., Robertson, P. D., Williams, J. S., Krahn, J., … Williams, R. S. (2011). Structure of an aprataxin–DNA complex with insights into AOA1 neurodegenerative disease. *Nature Structural & Molecular Biology*, *18*(11), 1189–1195. http://doi.org/10.1038/nsmb.2146

Tumbale, P., Williams, J. S., Schellenberg, M. J., Kunkel, T. A., & Williams, R. S. (2014). Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. TL - 506. *Nature*, *506 VN* - (7486), 111–115. http://doi.org/10.1038/nature12824

Uversky, V. N. (2007). Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. *Journal of Neurochemistry*. http://doi.org/10.1111/j.1471-4159.2007.04764.x

Vogel, U. S., & Thompson, R. J. (1987). Molecular cloning of the myelin specific enzyme 2′,3′-cyclic-nucleotide 3′-phosphohydrolase. *FEBS Letters*, *218*(2), 261–265. http://doi.org/10.1016/0014-5793(87)81058-X

Wakayama, T. (2007). Production of cloned mice and ES cells from adult somatic cells by nuclear transfer: how to improve cloning efficiency. *The Journal of Reproduction and Development*, *53*(1), 13–26. http://doi.org/10.1262/jrd.18120

Wang, L. K., Nandakumar, J., Schwer, B., & Shuman, S. (2007). The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair. *Rna*, *13*(8), 1235–44. http://doi.org/10.1261/rna.591807

Wang, Y., Hai, T., Liu, Z., Zhou, S., Lv, Z., Ding, C., … Zhou, Q. (2010). HSPC117 deficiency in cloned embryos causes placental abnormality and fetal death. *Biochemical and Biophysical Research Communications*, *397*(3), 407–412. http://doi.org/10.1016/j.bbrc.2010.05.105

Weiss, B., & 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. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(4), 1021–1028. http://doi.org/10.1073/pnas.57.4.1021

Weitzer, S., & Martinez, J. (2007). The human RNA kinase hClp1 is active on 3’ transfer RNA exons and short interfering RNAs. *Nature*, *447*(7141), 222–226. http://doi.org/10.1038/nature05777

Zhelkovsky, A. M., & McReynolds, L. A. (2014). Polynucleotide 3'-terminal phosphate modifications by RNA and DNA ligases. *Journal of Biological Chemistry*, *289*(48), 33608–33616. http://doi.org/10.1074/jbc.M114.612929

Zillmann, M., Gorovsky, M. A., & Phizicky, E. M. (1991). Conserved mechanism of tRNA splicing in eukaryotes. *Molecular and Cellular Biology*, *11*(11), 5410–6. http://doi.org/10.1128/MCB.11.11.5410.Updated