The 2'-5' RNA Ligase of *Escherichia coli*: Purification, Cloning, and Investigations of *in vivo* Function

Thesis by

Eric Arn

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Defended October 10, 1995)

Acknowledgments

Special thanks are due to Dr.John Abelson for his support, mentoring, and faith in my unproven abilities.

This work could not have been completed without the valuable intellectual contributions of Drs. Jaime Arenas and Reinhard Rahaut (during the dark days of protein purification) as well as Drs. Jeff Sampson and Peggy Saks (on all matters of RNA chemistry and the vagaries of *E. coli*) as well as many other members of the Abelson lab past and present.

Thanks are also due to Christopher Trotta for many hours of fruitful discussion, companionship, and commiseration.

Thanks to Herbert J. Botts and Dr. John Bowman for being shining lights of inspiration.

Most of all, immeasurable thanks are due to my parents for their constant support of all my undertakings, scientific and otherwise.

Abstract

An RNA ligase activity has previously been detected in extracts of *E. coli* which is capable of joining *S. cerevisiae* tRNA splicing intermediates in the absence of ATP to form a 2'-5' phosphodiester linkage (36). In order to study the mechanism and function of this unusual enzyme in bacterial RNA metabolism, a purification of the ligase enzyme was undertaken. The ligase was purified to homogeneity from a soluble high-speed extract of *E. coli* utilizing standard chromatographic techniques and reconstitution of activity following separation by SDS-PAGE. A single active polypeptide of approximately 20 kiloDaltons (kD) was shown to provide RNA ligase activity. This protein was N-terminally sequenced, and the open reading frame (ORF) encoding it was identified by a database search. This ORF, which codes for a novel protein with a predicted molecular weight of 19.9 kD, was cloned by PCR and used for overexpression of active recombinant ligase in *E. coli* and *S. cerevisiae* . The single chromosomal gene encoding the ligase was disrupted by insertion, abolishing ligase activity. Cells lacking active ligase are viable and show growth kinetics identical to the parent strain. Either parent strain or ligase knockout expressing high levels of recombinant ligase grow slowly compared to wild type and are temperature sensitive. Computer analysis of the ligase protein sequence allowed prediction of antigenic peptides derived from it. These peptides were synthesized and injected into rabbits to elicit polyclonal anti-ligase antibodies. Such antibodies were purified from rabbit sera by affinity to immobilized ligase, and shown to specifically recognize the ligase protein on Western blots. Ligase minus strains, purified recombinant ligase, and anti-ligase antibodies have been utilized in a variety of experiments to attempt to identify the *in vivo* substrate of this enzyme.

Table of Contents

Chapter I - Review: RNA Ligases

Overview

Enzymatic activities specifically catalyzing the ligation of RNA molecules via covalent phosphodiester bond formation have been detected in organisms throughout all major divisions of the phylogenetic spectrum. Beginning with the discovery of oligoribonucleotide circularization in extracts of T-even phage-infected *E. coli* in 1972 (98), seven different types of RNA ligase enzymes falling into four distinct mechanistic classes have been discovered to date. Specific metabolic functions have been assigned to only three of these enzymes: T4 RNA ligase which repairs nicks introduced into tRNAs of restrictive *E. coli* host strains upon phage infection, tRNA splicing RNA ligase which joins tRNA half-molecules resulting from endonucleolytic removal of introns from eukaryotic tRNA precursors, and an RNA ligase responsible for the joining of guide RNA molecules to mRNA during RNA editing in kinetoplastids. The functions of the metazoan "animal pathway" and E. coli 2'-5' RNA ligases *in vivo* remain to be determined, as does the *in vivo* substrate range of the archaeal stable RNA splicing ligase.

Among the seven known types of RNA ligase, four general mechanistic classes have been defined. The first class, exemplified by the tRNA splicing ligases, follows a complex reaction pathway involving four separable catalytic activities utilizing two equivalents of nucleotide triphosphates for each bond formed, and incorporating a phosphate derived from a nucleotide cofactor into the ligated junction. A highly similar pathway is used by T4 RNA ligase, but the enzymatic activities of the T4 polynucleotide kinase (PNK) are required in trans to complete the entire reaction. Available evidence implies that the editing RNA ligase is mechanistically identical to T4 RNA ligase. The animal pathway ligase apparently operates using a simpler mechanism which formally requires only one equivalent of adenosine triphosphate (ATP) per bond formed and does not incorporate exogenous phosphate in the product. Lastly, the *E. coli* 2'-5' RNA

ligase appears to utilize a single transesterification mechanism which does not require an added energy source.

The cloning and sequencing of genes encoding the T4 RNA ligase, two tRNA splicing ligases, and the *E. coli* RNA ligase have allowed direct comparisons of the primary structures of these enzymes to each other and to sequence databases. Observed similarites between these sequences have helped to delineate functional domains within the proteins and identify amino acids essential for function. The availability of cloned RNA ligase genes has also allowed overexpression of these enzymes and manipulation of their amino acid sequences to produce native and modified ligase proteins for use in biochemical experiments. This has led to a rather thorough examination of the enzymatic mechanisms of these enzymes and insight into their functions *in vivo*. This review will attempt to provide a survey of the known RNA ligase enzymes and their characteristics, as well as an examination of common and unique enzymatic ligation mechanisms of these enzymes.

Known Ligases and Their Functions

1) T4 RNA Ligase:

Catalytic Activity - T4 RNA ligase and the reaction catalyzed by it have been subjected to close scrutiny over the past 23 years, and the mechanism is well understood. The RNA ligase of T4 bacteriophage was initially observed to ligate the 3'-hydroxyl and 5'-phosphate termini of polyriboadenylate (poly(A)) to produce circular molecules in a reaction dependent on ATP and Mg^{2+} (98). The ligation product is a 3'-5' phosphodiester bond containing the 5' phosphate of the substrate. An identical junction is formed during intermolecular ligation (114). This reaction is represented schematically in Figure I-1A. The ATP required for this reaction is subjected to pyrophosphorylytic

Figure I-1: Schematic Representation of Intermolecular Joining Reactions Performed by Four Classes of RNA Ligase:

A. T4 RNA ligase and the editing RNA ligase.
B. tRNA splicing RNA ligases.
C. Animal pathway RNA ligase and archaeal stable RNA splicing RNA ligase.
D. *E. coli* 2'-5' RNA ligase.

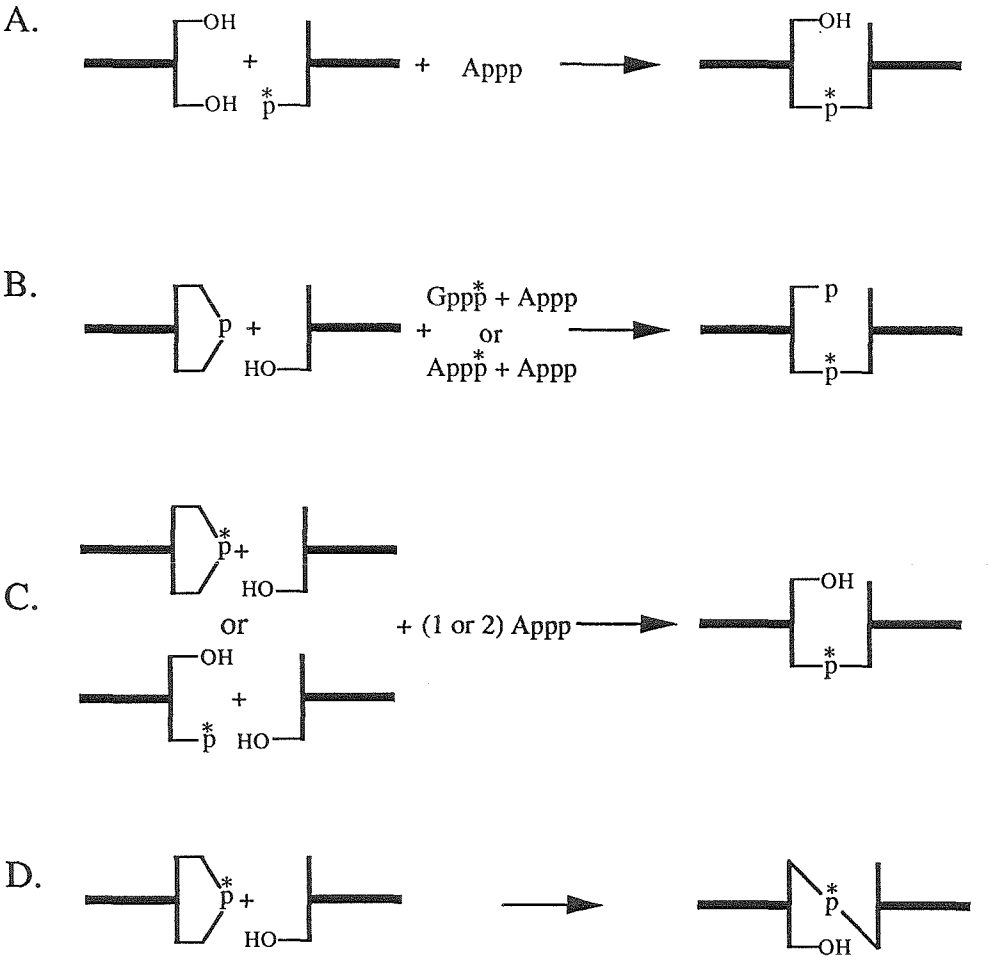

A.
OH
+
OH
p*
+
Appp
OH
p*
B.
p
+
HO
Gppp* + Appp
+
or
Appp* + Appp
p
p*
C.
p*+
HO
or
OH
+
p*
HO
+ (1 or 2) Appp
OH
p*
D.
p*+
HO
p*
OH

Figure I-1

cleavage, releasing AMP and inorganic pyrophosphate (PP_i) in amounts equivalent to the bonds formed (98). Although T4 DNA ligase also has the ability to ligate RNA molecules in a similar reaction, it does so at a much lower rate compared to DNA ligation and requires hybridization of the substrates to a template molecule, which the RNA ligase does not (25). True RNA ligase activity is restricted to T-even phages, with none detected in extracts after infection of *E. coli* with T-odd, Qβ, or λ phage (98).

Specificity- The overall substrate specificity of T4 RNA ligase is quite broad. The substrate molecule displaying a 5'-phosphate terminus, the donor, can be a deoxyribonucleotide or contain modified sugar or base residues without greatly affecting the reaction rate (7, 100). For RNA substrates, 5'-terminal pyrimidines are somewhat preferred over purines (66) The enzyme recognizes the first base and first two phosphates, making nucleoside 3',5'-bisphosphates the smallest possible donors, however it does not appear to have a length limit (24).

The requirements for the acceptor substrate (which provides the 3' hydroxyl) are more strict. Deoxyriboadenylate acceptors react over 200 times more slowly than riboadenylates, providing the enzyme's RNA specificity (68). The smallest reactive acceptor is a trinucleotide, indicating a recognition site encompassing three nucleosides and included bridging phosphates, and no limit to the length of the acceptor has been observed (48). Although any RNA sequence can be utilized as an acceptor at some rate, the acceptor site of the enzyme does display a marked sequence preference. 3'-terminal adenosines are the best acceptors, cytosine and guanosine show an intermediate reactivity, and terminal uridine residues are poor substrates (24). The ability to accommodate any RNA sequence at some efficiency has allowed this enzyme to be exploited for many RNA synthesis and labeling applications.

T4 RNA ligase is also sensitive to the secondary structure of RNA substrates. Donor terminal nucleotides which are base-paired in a duplex structure are generally

inactive (16). Conversely, duplex acceptors are active and in fact reaction yields are improved for acceptors in a duplex with a single 5' overhang that has the potential to base pair with the donor (102). Single stranded donors and acceptors which have the potential to form duplex structures with each other are however, poorly joined as the substrates cannot be properly aligned (68).

T4 RNA ligase shows a very high specificity for the base portion of the nucleoside triphosphate cofactor. Only ATP or dATP, of all the common nucleotides, can be productively utilized with a K_m for ATP of 12μM (20). Competitive inhibition studies with ATP analogs have shown that recognition of this cofactor takes place primarily at positions 1 and 6 of the heterocyclic adenine ring (47). The polyphosphate portion of the cofactor is not as stringently recognized, as many different forms can be utilized as long as they contain a hydrolyzable α–β phosphodiester bond (46). Thus, *in vivo* ligation by T4 RNA ligase may be sensitive to free ATP concentrations in the cell.

Purification- The availability of efficient assays of intra- or intermolecular RNA joining permitted purification of RNA ligase from extracts of non-lysogenic T4-infected *E. coli* soon after the activity was discovered. The use of nucleic acid affinity fractionation in concert with ion exchange and precipitation techniques allowed the isolation of a highly purified polypeptide with an apparent molecular weight of 43 kilodaltons (kD) (100). The identity of the protein was confirmed by the labeling of this polypeptide after formation of a covalent reaction intermediate with an α-^{32}P ATP cofactor. Although many refinements have been added to the purification protocol, RNA ligase preparations from phage-infected cells are often plagued with contaminating nucleolytic activities.

Cloning- The gene encoding the T4 RNA ligase was initially identified as gene 63 of the T4 genome by mapping of hydroxylamine-induced mutations which significantly

reduced RNA ligase activity in infected bacteria (99). Although this region of the genome was found to be resistant to cloning in *E. coli,* gene 63 was cloned in segments as pieces of sonicated DNA and the full coding sequence assembled in 1983 (83). The gene 63 open reading frame (ORF) was found to encode a 375 amino acid protein of molecular weight 43.5 kD with limited sequence homology to ATP-dependent DNA ligases. The ORF was cloned in *E. coli* under the control of a heterologous promoter and has been extensively utilized for the overproduction of T4 RNA ligase and also for site-directed mutagenesis of the enzyme in the time since (40).

Function- Before it was identified as the protein harboring RNA ligase activity, the product of T4 gene 63 was known as the tail fiber attachment (TFA) protein. TFA greatly accelerates the noncovalent joining of tail fibers to the phage body, but is not itself a part of the mature phage (116). Although contained within the same protein, there is convincing evidence that these two activities act independently from different active sites. Specific *in vitro* reaction conditions have been found which completely inactivate either activity while leaving the other intact, and mutations have been isolated which abolish RNA ligase activity but retain full TFA activity (87). Mutant phage lacking RNA ligase activity were found to infect laboratory strains of *E. coli* as efficiently as wild type phage, with no defects in protein synthesis, DNA replication or packaging. The physiological function of T4 RNA ligase therefore remained a mystery.

The one currently known metabolic function of T4 RNA ligase was identified utilizing a unique *E. coli* strain, CTr5x. This strain was created by crossing *E. coli* K-12 Hfr with a clinical isolate from the Cal Tech collection which was nonpermissive for infection by T4 phage mutants lacking PNK activity ((22). CTr5x was found to also be restrictive to infection by T4 mutants lacking RNA ligase activity (87). This restriction was found to be due to an activity in this strain which cleaves the host $tRNA^{Lys}$ immediately 5' of the anticodon in response to T4 infection, preventing expression of T4

late genes (1). Cleavage of tRNA by the anticodon nuclease (ACNase) of CTr5x leaves 2',3' cyclic phosphate and 5' hydroxyl termini which must be modified by the 2',3'-cyclic phosphodiesterase, 3' phosphatase, and 5' kinase activities of PNK to produce a substrate for ligation by T4 RNA ligase, the product of which can be utilized in translation (1) The phosphate added to the 5'-hydroxyl of the substrate by the kinase activity of PNK is incorporated into the 3'5'-phosphodiester bond produced in the ligation (1).

The ACNase of CTr5x was mapped to an optional chromosomal element (inserted at minute 29 in this strain) which upon sequencing was found to contain genes for the components of a classical Ic-Type DNA modification-restriction system as well as the ACNase (62). The ACNase enzyme was found to be latent as a component of the modification-restriction protein complex, but is activated upon disruption of that complex by the stp peptide expressed by T4 early in infection (2). Thus it seems that while T4 may have developed the stp peptide to inactivate a host DNA-restriction system, *E. coli* may have responded by evolving the ACNase to prevent infection in this eventuality. This could then have provided the requirement for phage PNK and RNA ligase activities, as religation of cleaved host tRNA represents the only confirmed *in vivo* function for either. This model does not explain however, the disparity between the ubiquity of PNK and RNA ligase among T-even phages and the detection of ACNase in only a single clinical isolate of *E. coli.* It remains a matter of speculation whether the optional chromosomal element of CTr5x represents a phage exclusion system which was somehow lost by laboratory strains of *E. coli*, or a mobile element recently inserted into the genome (11). The possibility therefore remains that the action of T4 RNA ligase in this strain is fortuitous and that the true physiological function of this enzyme remains to be discovered.

2) tRNA Splicing RNA Ligase:

Activity- In contrast to the RNA ligase of T4, the tRNA splicing ligase was detected in a search for activities capable of performing a function postulated to be required for the known physiological process of tRNA splicing. In this process, which is essential in all eukaryotes, intervening sequences are removed from the anticodon loop of precursor tRNA transcripts through cleavage at intron borders by a specific splicing endonuclease, and the remaining tRNA half-molecules are ligated together to form a mature functional tRNA.

The most studied of the tRNA splicing RNA ligases is that of the yeast *Saccharomyces cerevisiae*, which was first detected *in vitro* in 1978 (54). This enzyme is capable of ligating the 2',3'-cyclic phosphate and 5'-OH termini created by cleavage with splicing endonuclease to form a 3'-5' phosphodiester linkage which also bears a 2'-phosphate monoester on the 5' ribose, as shown in Figure I-1B (35, 72). The 2'-phosphate is subsequently removed by a specific NAD-dependent 2'-phosphotransferase leaving a standard 3'-5' phosphodiester at the ligation junction (65). An identical ligase activity which had been previously observed in extracts of wheat germ was also shown to participate in tRNA splicing, indicating a general mechanism for this process (32, 58).

The yeast and wheat splicing ligases require two equivalents of nucleotide triphosphates (NTPs) per bond formed, which are converted to one equivalent of NDP and one equivalent of NMP (35, 79). The γ-phosphate removed from an NTP to produce NDP is incorporated into the 3'5'-phosphodiester linkage of the product, while the 2'-monoester phosphate is derived from the cyclic acceptor substrate (32). The reaction also requires Mg^{2+} or Mn^{2+}. Except for the additional 2'-phosphate monoester of the product, the reaction catalyzed by the tRNA splicing RNA ligase is therefore strikingly similar in substrate, mechanism, and product to the ligation of $tRNA^{Lys}$ fragments by T4 PNK and RNA ligase in CTr5x. In addition to its catalytic function, the tRNA splicing ligases may also perform a transport function in the delivery of intron-containing tRNA precursors to specific nuclear sites for splicing and export. This is indicated by the yeast

enzyme's specific *in vitro* binding to precursor tRNA transcripts, which can be demonstrated by gel shift and crosslinking, as well as an *in vivo* localization of the ligase along the inner nuclear membrane (4, 19, 105).

Specificity- The yeast tRNA splicing ligase exhibits a high degree of specificity, requiring a 5'-hydroxyl donor and 2',3'-cyclic phosphate or 2'-phosphate acceptor as well as conserved tRNA structural elements in ligation substrates (37, 75). The enzyme can ligate oligonucleotide substrates, but only very inefficiently with the ligation of 5'-^{32}P oligo A_{16} (displaying a mixture of cyclic, 2', and 3' phosphate acceptor ends) occurring 10^4-fold less efficiently than that of a tRNA splicing substrate (75). The yeast ligase does not appear to specifically recognize the nucleoside base modifications added to tRNA precursors *in vivo*, as half-molecules produced from artificial *in vitro* transcripts are ligated as efficiently as those derived from transcripts produced in yeast nuclear extracts (94).

The wheat version of this enzyme displays somewhat less specificity, ligating oligo(A), oligo(U_nG), and intervening sequences excised from yeast tRNA precursors efficiently (93). The less discerning activity of plant tRNA splicing ligase has apparently been exploited *in vivo* by viroids and virusoids to circularize their unit length genomic RNA transcripts after rolling circle replication in wheat, tobacco, and *Chlamydomonas* (15, 50, 51).

Both the wheat and yeast tRNA splicing RNA ligases were originally thought to specifically utilize ATP or dATP for their NTP requirements in ligation. Recent binding and joining experiments, however, have shown that the yeast enzyme actually utilizes one equivalent of GTP and one equivalent of ATP per ligation reaction, with GTP preferred 500-fold as the source of γ-phosphate for phosphorylation of the donor 5'-hydroxyl (9). These two NTPs appear to be bound at independent nucleotide binding sites. The site utilized to bind a cofactor for the kinase reaction binds GTP with a K_D of 0.15μM but can

apparently accommodate ATP with a K_D of 1mM, while the second binding site is specific for ATP with a K_D of 4μM (9). Although detailed studies of this sort have not been carried out with the wheat enzyme, optimized assays of wheat germ 5'-hydroxyl kinase activity alone are carried out at μM concentrations of ATP, while whole joining reactions require mM ATP, as does yeast ligase in the absence of GTP (77). This suggests that the wheat ligase may have a similar dual nucleotide requirement.

Purification and Cloning- tRNA splicing RNA ligase was purified from high salt extracts of *S. cerevisiae* through salt precipitation and nucleic acid analog affinity chromatography, with a highly effective affinity elution from heparin-agarose by the addition of bulk tRNA providing nearly homogeneous enzyme (75). All enzymatic activities required for ligation of tRNA splicing intermediates including kinase, cyclic phosphodiesterase, and joining activities copurified throughout the procedure. The predominant polypeptide in the final fraction was a protein of 90 kD which could be labeled by the addition of α-^{32}P ATP (75). One interesting feature of this protein is an apparent modular organization of domains as demonstrated by the ability of various protease digestion fragments to separately perform the individual enzymatic activities required for the overall joining reaction (123).

The purified 90kD protein was subjected to N-terminal chemical sequencing and the peptide sequence determined was utilized to design oligonucleotides for hybridization screening of an *S. cerevisiae* genomic library which yielded a clone containing the ligase gene. Sequencing of this clone revealed an ORF encoding a basic polypeptide of 95.4 kD with no obvious similarities to known protein sequences (119). The cloned ligase gene has proven useful for overexpression of ligase in homologous and heterologous systems, and for the creation of genomic disruptions which show the ligase to be responsible for tRNA half-molecule joining *in vivo*, and to be essential for viability (76). Rescue of such a genomic ligase disruption by a clone from a *Candida albicans* genomic library has

allowed the cloning and sequencing of the tRNA splicing ligase from that organism, which is 42% identical to the *S. cerevisiae* protein (8).

The wheat tRNA splicing RNA ligase has been purified over 6,000-fold from an extract of wheat germ by ion exchange, phosphate affinity, hydroxyapatite, and ATP affinity chromatography, followed by sedimentation through glycerol gradients (78). As with purification of the yeast enzyme, all catalytic activities necessary for tRNA half-molecule joining are found to copurify. The predominant polypeptide in the final purification fraction is a doublet of approximately 110kD which can be labeled with α-^{32}P ATP (78). Although this protein has not been cloned, all available evidence suggests that it is a homolog of the yeast tRNA ligases.

Phylogenetic Distribution- Although the tRNA splicing RNA ligase has been most thoroughly studied in yeast and wheat, it has also been detected in a variety of eukaryotic systems. An activity in extracts of the primitive plant *Chlamydomonas* has been observed to ligate both endogenous tRNA half-molecules and viroid genomic RNA with formation of 2'-phosphomonoester, 3',5'-phosphodiester linkages (51, 110). In HeLa cells, an RNA ligase activity has been detected which is capable of ligating tRNA half-molecules to produce a 2'-phosphomonoester, 3',5'-phosphodiester bond with incorporation of an exogenous phosphate, although it was originally masked by the so-called "animal pathway" ligase activity (125). The incorporation of ^{32}P from exogenous α-^{32}P.ATP into RNA 5' termini and into internal sites within tRNA in mouse L cell nuclear extract is also consistent with the action of an RNA ligase of this sort (121). A specific NAD-dependent phosphotransferase capable of removing the 2'-phosphate from the unique linkages generated by tRNA splicing ligase has also been detected in extracts of both HeLa cells and *Xenopus laevi* oocytes as well as yeast and wheat germ (21, 125). Thus it appears that the tRNA splicing RNA ligase activity is conserved throughout eukaryotes, although an additional RNA ligase activity is also detected in vertebrates.

3) Animal Pathway Ligase:

In addition to the conserved tRNA splicing ligase activity, extracts of HeLa cells also contain another RNA ligase activity which is capable of ligating tRNA half-molecules derived by cleavage of human, *Xenopus,* or *S. cerevisiae* tRNA precursors by the HeLa tRNA splicing endonuclease (28, 112). This alternative ligase requires 2',3'-cyclic acceptor and 5'-hydroxyl donor substrates, but ligates them to form a simple 3',5'-phosphodiester with the bridging phosphate derived from the cyclic acceptor, as shown schematically in Figure I-1C (28, 112). An identical activity has been observed to ligate *S. cerevisiae* tRNA half-molecules in *Xenopus* oocyte nuclei (69). Because it has only been observed in vertebrates, this activity has been called the "animal pathway" ligase.

The characteristics of the animal pathway RNA ligase have been examined in HeLa extracts and partially purified enzyme fractions. The enzyme can ligate a virusoid RNA fragment or oligoribonucleotide substrates as well as tRNA halves (29). The enzyme shows an apparent monomer molecular weight of 160 kD in glycerol gradients, requires Mg^{2+} for activity, and utilizes ATP or dATP with a K_m of 0.14μM for the ribonucleotide (73). It is not clear in what manner the ATP cofactor is utilized, however, as the reaction is inhibited by PP_i, indicating a pyrophosphorylytic cleavage, and also by AMP-PNP or ATP[γ-S] which would prevent γ-phosphate removal (73). The stoichiometry of ATP usage in this reaction has not been reported. The mechanism of the animal pathway ligase clearly differs significantly from that of the T4 or conserved tRNA splicing ligases, but the actual pathway utilized remains mysterious.

The animal pathway ligase activity is observed to join the vast majority of tRNA half-molecule substrates added to HeLa extracts or *Xenopus* oocytes. The relative proportions participation of this enzyme and the conserved pathway splicing ligase in tRNA splicing *in vivo*, however, are unknown. A more complete understanding of the

substrate ranges *in vitro* of the two vertebrate RNA ligases and of the distribution of introns in vertebrate tRNAs may help to shed light on this subject.

4) Archaeal RNA Ligase:

A distinct RNA ligase activity has been observed to participate in the splicing of introns from stable RNAs in archaea. These introns, which occur in 23S and 16S ribosomal RNAs as well as several different sites in various tRNAs, have in common a "bulge-helix-bulge" secondary structure at the intron boundaries which is sufficient for recognition by a single specific endonuclease (52, 107). The products of this endonuclease are most likely 3'-phosphate and 5'-hydroxyl termini (although there is a report of archeal tRNA intron removal leaving a 2',3'-cyclic phosphate and 5'-hydroxyl (107)) which are ligated by a presumably unique RNA ligase to produce mature tRNA and rRNA, as well as circular rRNA introns (53). This reaction is represented by Figure I-1C.

Splicing assays in crude extracts of the archaeote *Desulfurococcus mobilus* have revealed several features of the archaeal RNA ligase. This enzyme ligates 3'-phosphate acceptors and 5'-hydroxyl donors to produce a standard 3',5'-phosphodiester linkage in the absence of metal ions, but with a requirement for GTP (53). The junction phosphate is believed to be derived from the 3'-phosphate of the substrate as no radioactivity was incorporated into products from exogenous α or γ-^{32}P-labeled GTP or ATP (53). The ligase activity depends on extract protein, but not extract RNA, and can also detected in extracts of other extreme thermophile archaea. The archaeal RNA ligase therefore represents a new type of RNA ligase enzyme with a mechanism that has only begun to be investigated, but bears apparent similarity to that of the animal pathway RNA ligase.

5) Kinetoplastid Editing RNA Ligase:

Kinetoplastids such as *Trypanosoma brucei* and *Leishmania tarantolae* have been known for several years to contain an activity capable of circularizing endogenous ribosomal or messenger RNAs and adding cytidine 3',5'-bisphosphate to the 3' termini of various cytoplasmic RNAs (6, 43, 120). This reaction has been shown to require Mg^{2+} and ATP or dATP and to produce a standard 3',5'-phosphodiester linkage (6, 43, 120). A function for an RNA ligase in the editing of precursor mRNA transcripts of mitochondrial cryptogenes had been postulated, but editing via successive RNA-catalyzed transesterifications (which would not require an RNA ligase) could not be ruled out. (38).

Insight into the mechanism of kinetoplastid mitochondrial RNA editing was provided by the discovery of chimeric molecules formed of guide RNA transcripts attached at their 3' ends to truncated, partially edited mRNAs at known editing sites (14). Since these guide RNAs contain sequences complementary to edited mRNAs near editing sites as well as poly(U) tails which could serve as a source of uridine residues for insertion, these were regarded as editing intermediates which might have been produced via either proposed mechanism. A direct role for RNA ligase in guide RNA/mRNA chimera formation has, however, recently been demonstrated. Two distinguishable proteins of 50 and 57kD which can be covalently labeled by incubation with α–^{32}P ATP were found to be components of a chimera-forming ribonucleoprotein particle (RNP), and to release AMP concomitant with chimera formation (90). Both labeling of these proteins and chimera formation are inhibited by nucleotide cofactors with non-hydrolyzable α-β phosphodiester bonds, or by the addition of pyrophosphate (90). These results strongly imply the involvement in editing of two species of RNA ligase utilizing a mechanism similar to that of T4 or the conserved tRNA splicing ligase.

A cleavage/ligation mechanism of editing was confirmed by the discovery of an endonuclease activity associated with chimera-forming RNP's which cleaves pre-edited mRNAs at editing sites, leaving 3'-hydroxyl and 5'-phosphate termini (80). These cleaved mRNAs can be isolated and shown to be substrates for the formation of a specific

chimera by a mechanism resembling that of T4 RNA ligase, requiring ATP with a hydrolyzable α-β phosphate bond to produce a 3',5'-phosphodiester linkage (88). The likely mechanism of this enzyme is therefore the same as shown for T4 RNA ligase in Figure I-1A. The specificity of the joining reaction appears to lie in the conformation of the complementary mRNA and guide RNA substrates as T4 RNA ligase can successfully substitute for the *Trypanosome* enzyme in this reaction, and the *Trypanosome* ligase can efficiently join or circularize single-stranded oligoribonucleotides of random sequence (80).

6) *E. coli* 2'-5' RNA Ligase:

An RNA ligase activity capable of joining tRNA half-molecules produced by *S. cerevisiae* splicing endonuclease has been detected in extracts of several bacteria representing a wide range of phyla in the bacterial domain (36). When examined more thoroughly in *E. coli*, this enzyme was found to ligate only four of the ten different types of yeast splicing intermediates, to have no requirement for nucleotide triphosphates, and to incorporate the 2',3'-cyclic phosphate of the substrate into a 2',5'-phosphodiester linkage as seen in Figure I-1D (36). Further investigations with *E. coli* extracts and purified enzyme fractions indicate a marked specificity of this enzyme; a 2',3'-cyclic substrate terminus is required, short oligonucleotide substrates are not ligated, preferential ligation among the four yeast-derived substrates is seen, and nucleoside base modifications of the substrate are required (E. Arn, unpublished results). This specificity implies some interaction with endogenous tRNA substrates, and in fact an activity has been detected in *E. coli* C600 which can nick the anticodon loop of an *E. coli* $tRNA^{Lys}$ to produce a substrate for the 2'-5' ligase (E. Arn, unpub. res.). This enzyme does not appear to ligate the products of the CTr5x ACNase *in vivo*, as 2'-5' ligase activity levels are normal in this strain, yet tRNA fragments accumulate upon infection by T4 RNA ligase mutants (36).

The *E. coli* 2'-5' RNA ligase has been purified to homogeneity and derived N-terminal peptide sequence was used to clone the gene encoding it. The gene encodes a novel 20kD protein which shows a high degree of sequence similarity to an ORF of unknown function in *Bacillus stearothermophilus* (E. Arn, unpub. res.). Genomic disruptions of the ligase gene are viable, but show altered growth patterns and steady state RNA populations. The *in vivo* substrate and function of this enzyme remain to be elucidated.

7) Synthetic Action of Ribonucleases:

Ligation of RNA *in vitro* has also been observed to be catalyzed by various endoribonucleases performing a reverse reaction. For many microbial ribonucleases utilizing an RNase A-type mechanism, 2',3'-cyclic phosphate and 5'-hydroxyl cleavage intermediates can be covalently joined if the concentration of these is high enough (60). Oligoribonucleotides of considerable length can be produced by this method, and in fact manipulation of substrate concentrations and reaction conditions can even be used to favor 3'-5' or 2'-5' bonds in the product of such a reaction by ribonuclease T_1 (81). RNase T_1 has also been observed to cleave viroid genomic RNAs from multimeric transcripts and circularize them *in vitro* (109). This reaction is likely to be highly favored by the structure of the RNA, as some virusoids perform these reactions spontaneously in the absence of protein. The bacterial exonuclease RNase PH, which degrades RNA by a phosphorylytic mechanism, has also been shown to ligate nucleoside diphosphates to the 3' end of tRNA under certain reaction conditions (70). It should be noted, however, that the synthetic reactions of these enzymes have been shown to occur only *in vitro* under strikingly unphysiological reaction conditions. No evidence yet exists for ligation of an endogenous substrate *in vivo* by a ribonuclease.

8) RNA Catalyzed RNA Ligation:

It is now generally accepted that the ubiquitous splicing of introns from eukaryotic nuclear mRNA is likely to be RNA-catalyzed in a manner similar to that of the self-splicing Group I and II introns but dependent on the presence of a large number of proteins. In these reactions, concomitant cleavage and ligation of RNA via a single transesterification is performed by facilitated attack of a ribonucleoside 2' or 3' hydroxyl on a phosphoryl. While interesting in its own right, this does not constitute enzymatic ligation of independent RNA termini, and therefore falls outside the scope of this review. Recently, several classes of RNA structures catalyzing RNA ligation have been isolated from random sequences by *in vitro* selection. These are true ligases, capable of catalyzing intermolecular ligation of independent species with multiple turnover, and were found to produce either 2'-5' or 3'-5' phosphodiester linkages (23). It is interesting to note however, that the reaction catalyzed by these ribozymes is a templated joining of 3'-hydroxyl and 5'-triphosphate termini, resembling an RNA polymerase reaction much more than an RNA ligase. Although these ribozymes may shed light on ancient ligases existing in an RNA world, they do not appear to represent any present day biological phenomena, and will not be discussed further here.

Mechanisms of RNA Ligation

The mechanisms of action of several of the RNA ligase enzymes discussed above have been studied quite thoroughly and the reaction pathways known in some detail, although the specific catalytic residues of the enzymes are still mostly unknown . These highly characterized enzymes include the T4 RNA ligase and tRNA splicing RNA ligases, which display a general similarity in catalytic mechanism between themselves and also to DNA ligases and RNA capping enzymes. Potential reaction mechanisms of the remaining ligases can be proposed based on available biochemical data. The detailed mechanisms will be examined individually below.

1) T4 RNA Ligase

The mechanism of ligation by the RNA ligases of T-even bacteriophages, which is apparently also utilized by the kinetoplastid editing ligase, has been well understood for several years (23). This reaction consists of three steps: enzyme adenylylation, transfer of adenylyl to donor, and ligation of donor to acceptor.

Adenylylation - To begin the ligation reaction, the enzyme adenylylates itself by phosphorylytic cleavage of ATP with release of PP_i. This reaction, which in the presence of α-^{32}P ATP generates the labeled covalent enzyme intermediate exploited during purification, can proceed in the absence of RNA substrates and can be reversed in the presence of excess PP_i (20). Mass spectrometry of protease digestion products of adenylylated enzyme has revealed the site of adenylylation to be Lysine-99 in the T4 ligase (106). AMP is linked to the ε-amino group of this residue via a phosphoamide bond (45). An arginine substituted at position 99 in the ligase sequence can support adenylylation, but an enzyme with asparagine in this site is inactive for this step (40).

Although a variety of ATP derivatives can bind in the ATP site, the adenylylation reaction requires a nucleotide with a hydrolyzable α-β phosphodiester, and a suitable γ substituent(46). Chemical modification studies have also shown that the free hydroxyl groups of one to four of the ligase's tyrosine residues are involved in the catalysis of adenylylation (89). These results lead to a model for this step of the reaction where a tyrosine hydroxyl group on the enzyme hydrogen bonds to a nonbonded oxygen of the α phosphate of ATP, priming the a phosphorus for electrophilic attack on the e-amino group of Lys-99. This model is shown in detail in Figure I-2A.

Adenylyl Transfer- In the second step of the ligation reaction, the adenylyl group is transferred from Lys-99 of the ligase to the 5'-phosphate of the donor substrate to form an activated 5'-5' phosphoanhydride bond. This reaction is stimulated by the presence of

Figure I-2: Models for Catalysis of the Three Steps of RNA Ligation by T4 RNA Ligase.
A. Adenylylation.
B. Adenylyl transfer.
C. Phosphodiester formation.

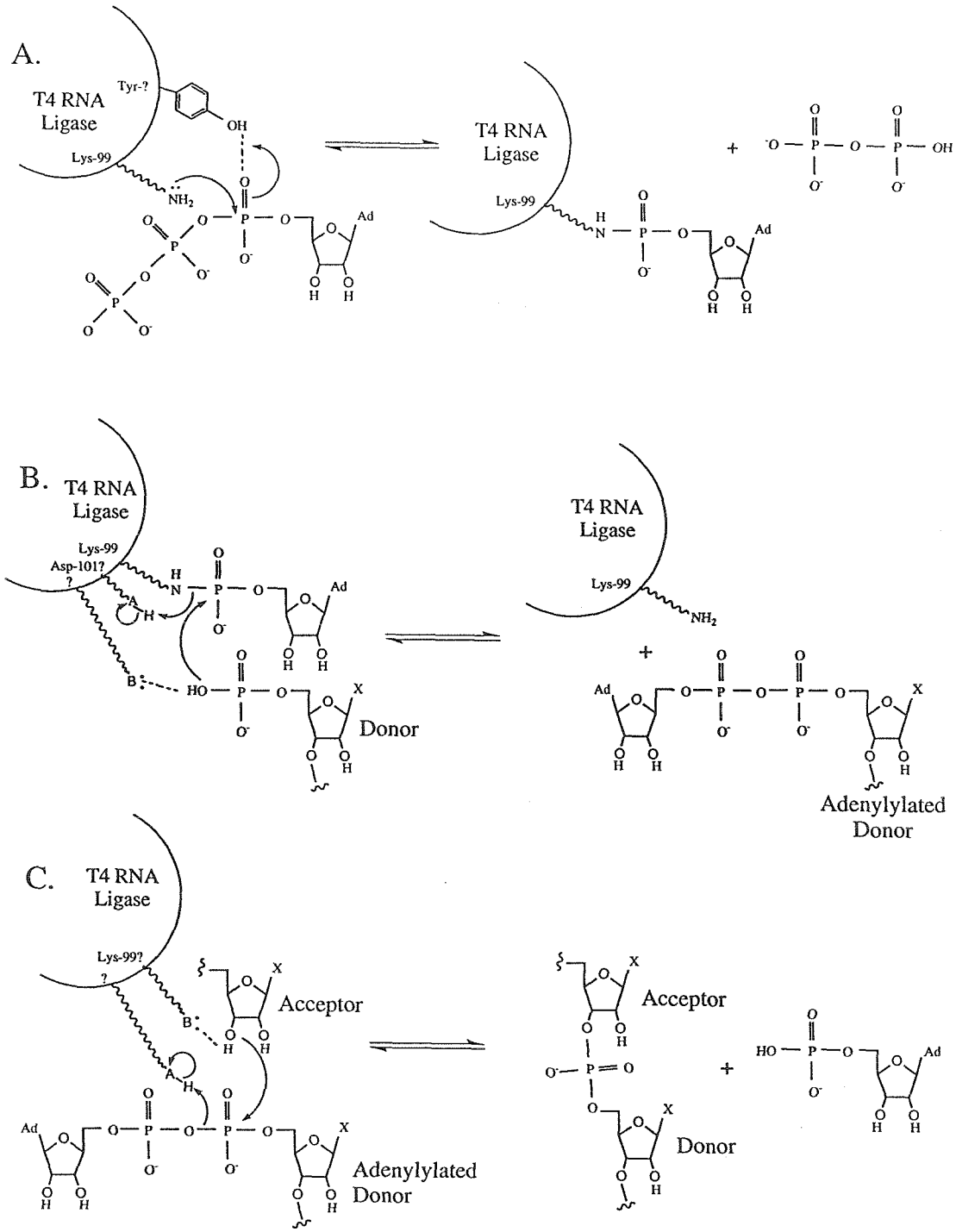

A.
T4 RNA Ligase
Tyr-?
Lys-99
T4 RNA Ligase
Lys-99
Ad
B.
T4 RNA Ligase
Lys-99
Asp-101?
Donor
T4 RNA Ligase
Lys-99
Adenylylated Donor
C.
T4 RNA Ligase
Lys-99?
Acceptor
Adenylylated Donor
Acceptor
Donor

Figure I-2

a potential acceptor substrate molecule as well as 5'-phosphorylated donor, and is reversible (61, 103). The use of preadenylylated donors for ligation in the absence of ATP significantly increases reaction velocity over the overall joining reaction, suggesting that the adenylyl transfer step may be rate-limiting (61, 103). The inability of adenylylated enzyme to utilize preadenylylated donor intermediates confirms the kinetic ordering of these steps (103). The low rate of this reaction in the absence of acceptor has made it difficult to examine the characteristics of this step alone without occurrence of the third and final step of the reaction, which is phosphodiester bond formation between donor and acceptor substrate. A proposed reaction mechanism for this step involving general acid and base catalysis is given in Figure I-2B.

Phosphodiester Formation- In this step, the 3'-hydroxyl of the acceptor substrate is primed for attack on the activated phosphoanhydride of the adenylylated donor, forming a 3',5'-phosphodiester with release of AMP. This occurs via direct nucleophilic displacement on the 5'-phosphorus of the adenylylated donor with inversion of stereochemical configuration, indicating that no covalent intermediate is formed in this step (17). The order of binding of acceptor and donor substrates is random, and apparently occurs under equilibrium conditions for good acceptor substrates (A) and steady-state conditions for poor acceptors (U) (124). Thus, the varying rate of phosphodiester bond formation provides the preferences seen amongst possible acceptors.

Some clues as to mechanism of catalysis of the second and third steps of the reaction have been obtained by site-directed mutagenesis. Aspartate-101 appears to be important in both the second and third steps which are completely blocked in mutants (including Glu-101 which retains the acidic moiety) that retain full adenylylation activity (40). The presence of a lysine or arginine at position 99 is also required for the third step of the reaction, indicating a multifunctional role for this residue (40). A possible

mechanism for this step which could involve general acid and base catalysis by the same residues active in earlier steps is shown in Figure I-2C.

Common Reaction Mechanism- This mechanism for nucleic acid joining via an adenylylated intermediate is common to other ligating enzymes. DNA ligases of organisms ranging from vertebrates to bacteria utilize an identical three-step mechanism for the ligation of deoxyribonucleotide 3'-hydroxyl and 5'-phosphate substrates, although the bacterial enzymes substitute NAD^+ as a source of AMP for adenylylation (25). In fact, oligodeoxynucleotide donors adenylylated by T4 RNA ligase can serve as substrates for phosphodiester bond formation by DNA ligase, if they are annealed to a template molecule(103).

A mechanism highly similar to the first two steps of the RNA or DNA ligase reactions is utilized by mRNA guanylyltransferase enzymes which function in mRNA-capping. These enzymes are guanylylated at a specific lysine residue with phosphorylytic cleavage of GTP, and then transfer GMP to a 5'-diphosphate terminus of mRNA (96). The general mechanisms of these two enzymes are represented schematically in Figure I-3A an B. The reaction mechanism conserved among these enzymes is reflected in conserved amino acid motifs which become apparent when the various sequences are aligned according to their active site lysines. These motifs will be discussed in a subsequent section.

2) Conserved tRNA Splicing RNA Ligase Mechanism:

As mentioned previously, the reaction catalyzed by the yeast and wheat germ tRNA splicing ligases is nearly equivalent to the sum of the reactions catalyzed by T4 RNA ligase and PNK. The ligation reaction proceeds from the 3'-hydroxyl and 5'-phosphate intermediates by the three-step mechanism described above. Additional cyclic phosphodiesterase and polynucleotide kinase activities which are necessary to generate

Figure I-3: Joining Reactions Catalyzed by Enzymes Related to RNA Ligase
A. DNA ligase.
B. mRNA capping enzyme, small subunit.

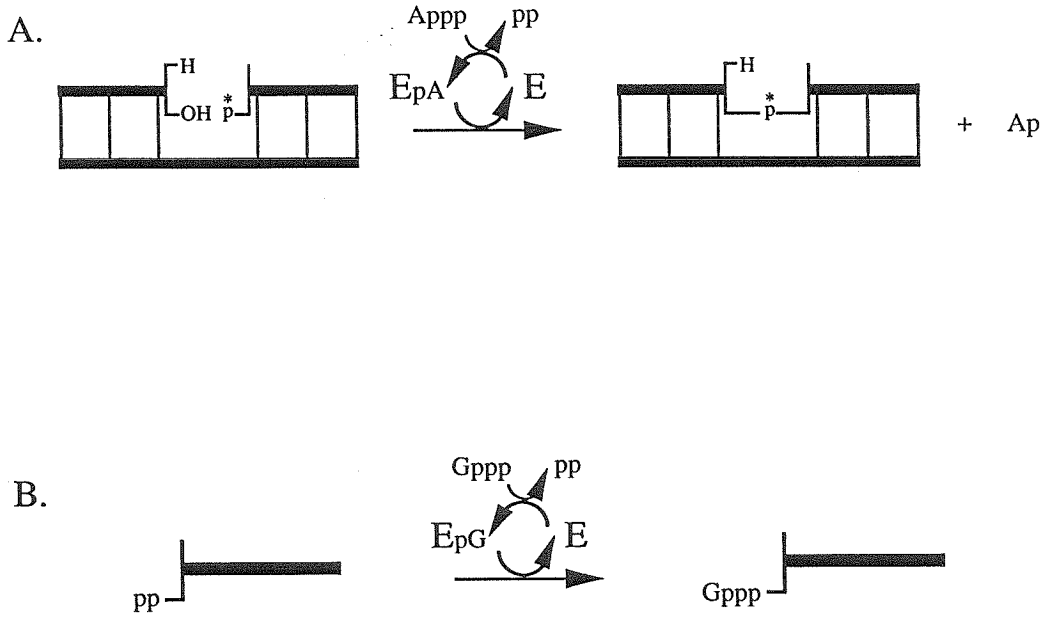


Figure I-3

these intermediates from 2',3'-cyclic phosphate and 5'-hydroxyl substrate termini are also present in the tRNA splicing ligases. These are utilized in the joining reaction as detailed below. The mechanisms of tRNA splicing RNA ligase and the T4 enzymes are compared in Figure I-4.

2',3'-Cyclic Phosphodiesterase- Opening of the 2',3'-cyclic phosphate acceptor terminus to a 2'-phosphomonoester appears to be a common first step of the reaction catalyzed by the tRNA splicing RNA ligases. This step can occur in the absence of ATP and donor substrate, but the rate is increased in the presence of 5'-hydroxyl (but not 5'-phosphate) termini (79). This reaction is not concerted with any of the subsequent steps, as 2'-monoester substrates are joined as efficiently as 2',3'-cyclic (79). This step represents the most significant deviation from the T4 ligase/PNK mechanism, as the phage PNK opens cyclic phosphates to a 3'-monoester which it removes by a phosphatase activity before ligation (117). In the *S. cerevisiae* tRNA splicing ligase, the cyclic phosphodiesterase activity has been localized to a 230 amino acid fragment of the carboxy-terminal region of the enzyme, which may bear weak similarity to other enzymes which exhibit this activity (4). Non-tRNA oligoribonucleotide substrates have been successfully utilized by several investigators to assay this activity, demonstrating that the substrate specificity does not lie in this step.

5'-Hydroxyl Kinase- The second step which must occur before ligation by the T4 ligase mechanism can occur is the addition of a phosphate to the 5'-hydroxyl of the donor, which is accomplished by a kinase activity integral to the tRNA splicing RNA ligases. In this reaction the ligase transfers the γ-phosphate of an NTP cofactor to the donor 5' end, where it will eventually become the bridging phosphate of the product linkage. Some hydrolysis of ATP to ADP by the phosphorylation activity is observed in the absence of substrate, but this activity is greatly stimulated by the addition of 5'-hydroxyl termini

Figure I-4: Comparison of the Overall Reactions Catalyzed by T4 PNK and RNA Ligase or by tRNA Splicing RNA Ligase.

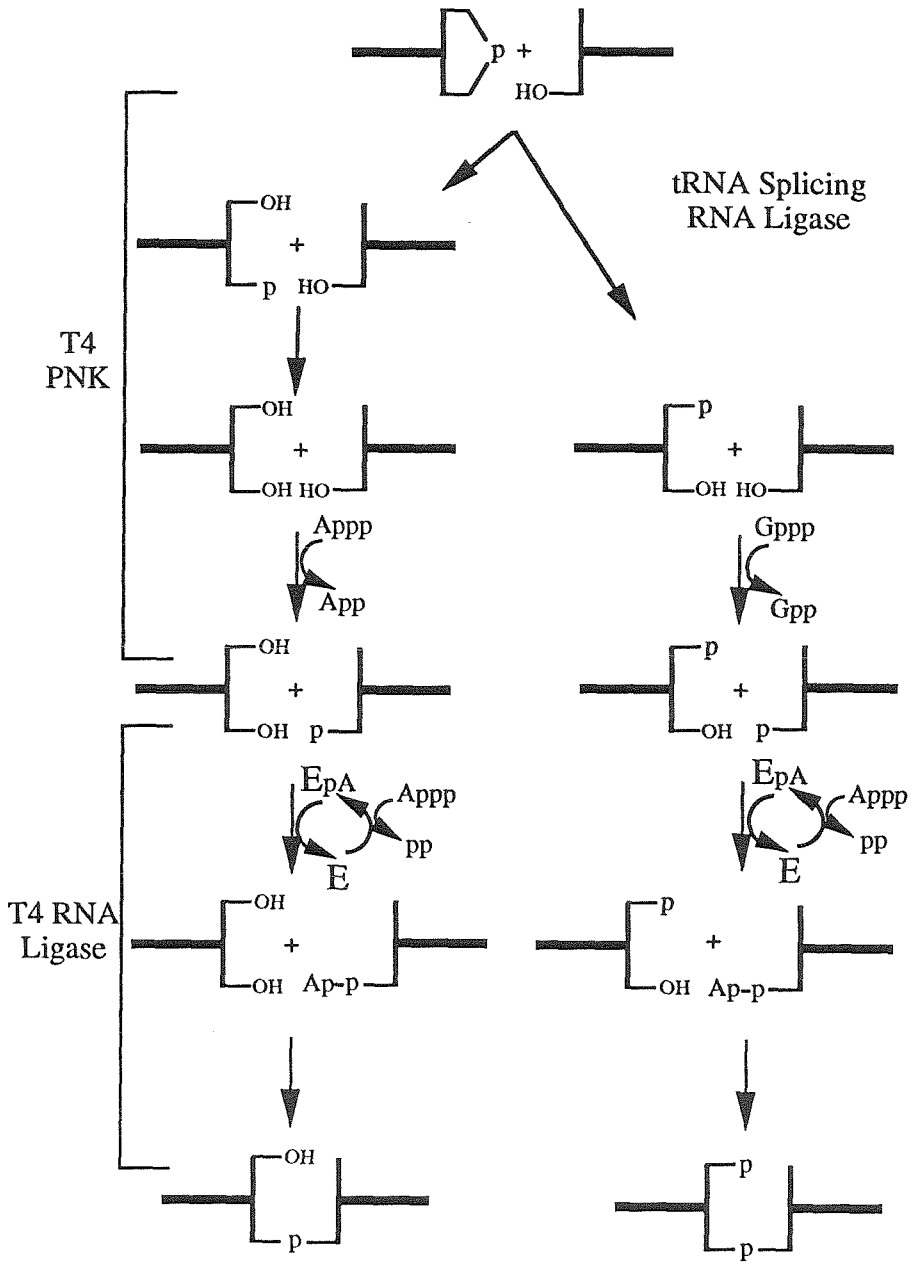

p +
HO
tRNA Splicing
RNA Ligase
OH
+
p HO
T4
PNK
OH
+
OH HO
p
+
OH HO
Appp
App
Gppp
Gpp
OH
+
OH p
p
+
OH p
EpA
Appp
pp
E
EpA
Appp
pp
E
T4 RNA
Ligase
OH
+
OH Ap-p
p
+
OH Ap-p
OH
p
p
p

Figure I-4

(although a suitable acceptor terminus is not required) (79). Although all the tRNA ligases assayed can utilize ATP exclusively to support this reaction, the *S. cerevisiae* enzyme (and likely the other splicing ligases) prefers to utilize GTP by at least two orders of magnitude for this step (9). The location of the kinase domain within the yeast enzyme is not precisely clear, but residues within the last 400 amino acids are required for activity, with the GTP-binding site apparently located within the C-terminal 200 amino acids (4).

Adenylylation- The tRNA splicing RNA ligases are adenylylated as are T4 RNA ligase and DNA ligases, with reversible pyrophosphorylytic cleavage of ATP. PP_i is released in equimolar amounts to the ADP produced by the kinase activity. The wheat germ tRNA ligase adenylylation activity is active only in the presence of RNA, although the yeast enzyme and T4 ligases can be adenylylated in the absence of substrates (35). A preferential stimulation of wheat ligase adenylylation by 5'-hydroxyl termini over 5'-phosphate donors suggests that adenylylation may actually precede the kinase reaction kinetically, but since both types of donors are stimulatory this is not conclusive (79). The adenylyl group is covalently bound to the tRNA splicing RNA ligases by a phosphoamide linkage to a residue which has been identified as lysine-114 in the *S. cerevisiae* sequence (123). Full adenylylation activity is provided by the N-terminal 400 amino acids of the S. cerevisiae ligase (4). The N-terminal 300 amino acids of the *S. cerevisiae* and *C. albicans* splicing ligases show significant similarity to the T4 RNA ligase when the sequences are aligned by the adenylylated lysines, as will be shown in a subsequent section.

Adenylyl Transfer- Following the formation of the tRNA ligase-AMP complex, the adenylyl group is transferred to the 5'-phosphate of the donor substrate to form an activated phosphoanhydride. Preadenylylated donors can be utilized for ligation in the absence, but not the presence, of ATP suggesting the same ordering of steps as for T4

RNA ligase (93). Substrate recognition for this step would appear to provide the specificity of the yeast tRNA ligase since cyclic phosphodiesterase, kinase, and adenylylation activities of this enzyme are efficient with model substrates. 2'-acceptor and 5'-donor phosphomonoester terminated substrates are seen to form a stable complex with the wheat tRNA ligase until the adenylyl transfer step, when they are rapidly ligated and released (79). This again suggests a close association between adenylyl transfer and phosphodiester bond formation, as for the T4 RNA ligase.

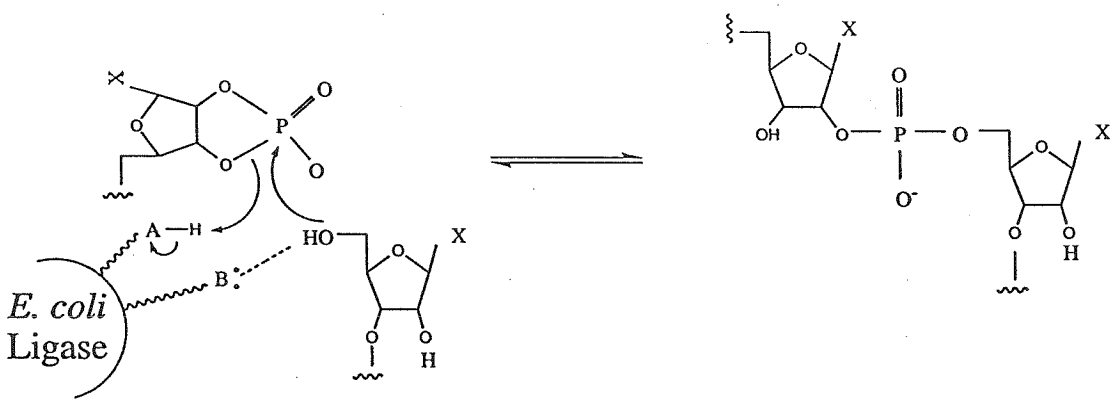
Phosphodiester Bond Formation- The final step of ligation by the tRNA splicing RNA ligases is the facilitated attack of the acceptor 3'-hydroxyl on the activated donor phosphoanhydride to form a 3',5'-phosphodiester, 2'-phosphomonoester bond. One equivalent of AMP is released per bond formed (79). The final product is then available for 2'-phosphate removal by 2'-phosphotransferase to produce mature tRNA.

3)Animal Pathway, Archaeal, and *E. coli* 2'-5' RNA Ligase Mechanisms:

Although the chemical mechanism of RNA ligation have not been thoroughly examined for these enzymes, model pathways can be proposed based upon the available evidence. The animal pathway and *E. coli* enzymes both require 2',3'-cyclic phosphate and 5'-hydroxyl substrates and incorporate the cyclic phosphate into a phosphodiester linkage product. The simplest mechanism that can be proposed for these enzymes would therefore be a facilitated attack by the 5'-hydroxyl on either the 2' or 3'-phosphoryl of the cyclic phosphate to generate either a 3'-5' or 2'-5' phosphodiester by a single transesterification. This is equivalent to the synthetic reverse reaction performed by RNase A-type endonucleases. Since the *E. coli* 2'-5' ligase does not require a nucleotide cofactor, this seems a likely mechanism. A proposed mechanism for this reaction utilizing general acid and base catalysis as in the ribonuclease mechanism is shown for the *E. coli* RNA ligase in Figure I-4A. Since this mechanism should proceed with inversion

Figure I-5: Proposed Mechanisms for Catalysis by Other RNA Ligases.
A. *E. coli* 2'-5' RNA Ligase.
B. Animal pathway and archaeal RNA ligases.

A.


E. coli
Ligase
A—H
B
HO
X
OH
O
P
H

B.

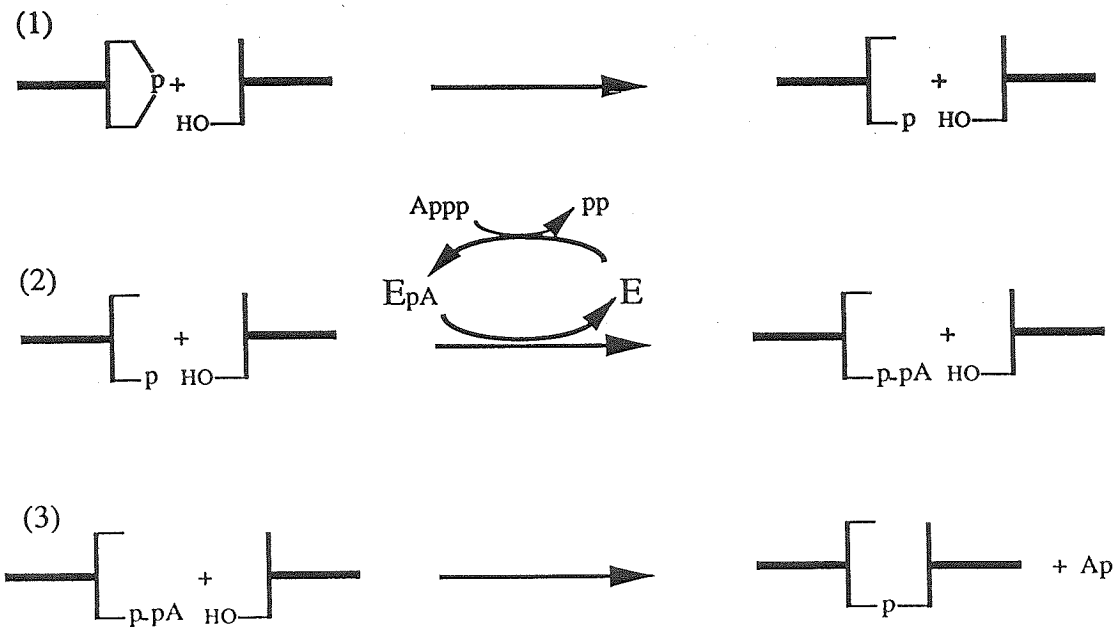

(1)
p +
HO
+
p HO
(2)
Appp
pp
EpA
E
+
p HO
+
p-pA HO
(3)
+
p-pA HO
p
+ Ap

Figure I-5

of configuration about the cyclic phosphorus, this mechanism may be investigated by the use of cyclic phosphorothioates of known configuration in the substrate. If instead a double transesterification with a covalent enzyme-substrate intermediate is formed, configuration should be retained. The availability of large amounts of pure recombinant 2'-5' ligase protein should facilitate the isolation of covalent enzyme intermediates or crosslinking of substituted substrate to the active site for identification of catalytic residues.

Unlike the *E. coli* enzyme, the animal pathway and archaeal RNA ligases have been shown to require ATP. Unless this nucleotide is simply necessary for induction of an unfavorable conformational change in enzyme or substrate conformation, a more complex mechanism must be proposed to explain this requirement. Also, since the archaeal ligase is reported to be active with 3'-phosphate acceptors as well as 2',3'-cyclic acceptors, the single transesterification mechanism would be unlikely. One possible mechanism to accommodate the known characteristics of these enzymes would be a three step process similar to that of T4 RNA ligase, but with reversal of the donor and acceptor functions. Such a possible mechanism is illustrated in Figure I-4B. The 2',3'-cyclic phosphate may be opened to a 3'-phosphate by a cyclic phosphodiesterase activity similar to PNK (but lacking 3'-phosphatase activity). The enzyme could be adenylylated, and then transfer the adenylyl to the 3'-phosphate to form an activated 3'-5' bond which may be attacked by the 5'-hydroxyl substrate to form a 3',5'-phosphodiester. This mechanism would require one equivalent of ATP per ligation with no inclusion of exogenous phosphate in the product. The plausibility of this proposed pathway is bolstered by the fact that 3'-phosphate termini can be adenylylated by T4 RNA ligase in a so-called "reverse transfer" reaction (111). Since *in vitro* assays using crude archaeal extracts or partially purified HeLa protein fractions are available for the investigation of these enzymes, this mechanism may perhaps be demonstrated by the production of a

covalent enzyme-AMP intermediate or the utilization of preadenylylated 3'-phosphate substrates for ligation in the absence of ATP.

Conserved Primary Sequence Motifs in RNA Ligases and Related Enzymes

1) T4 RNA Ligase, DNA Ligases, and 5'-Guanylyltransferases:

The availability of sequence information for T4 RNA ligase and several enzymes of similar mechanism has allowed the alignment of primary amino acid sequences to identify common sequence elements. Sequence elements in common among RNA ligase, DNA ligases, and mRNA capping enzymes may be involved in nucleic acid binding or a conserved mechanism of enzyme adenylylation or guanylylation and subsequent nucleotidyl transfer catalysis. Since the capping enzymes perform guanylylation of substrates but not ligation, they would not be expected to contain motifs necessary for the last step of the overall ligation reaction. Figure I-6A shows four conserved regions of sequence present in all three classes of enzymes when they are aligned according to their known active site lysines [adapted from (97)]. The identification of presumably functional conserved sequence elements in these enzymes provides a guide for direct tests of residue function by site-directed mutagenesis.

Motif I The K-X-D-G-X-R motif at the site of nucleotidyl addition has been most thoroughly investigated by mutagenesis. As mentioned, in T4 RNA ligase the active site lysine can only be weakly substituted by arginine, but not by other amino acids, the adjacent glutamate can be altered with little effect on activity, and the conserved aspartate is required for the second and third steps of the ligase reaction (40). Active site mutants of human DNA ligase I expressed in *E. coli* have been tested for both adenylylation activity and the ability to complement DNA ligase mutants of *E. coli*. In this enzyme, adenylylation activity is lost if the active site lysine is substituted by either arginine or histidine or if the conserved downstream glycine or arginine are replaced with alanine

Figure I-6: Amino Acid Sequence Comparisons of RNA Ligases and Related Enzymes.

A. Conserved motifs in capping guanylyltransferases (CE), DNA ligases (DNA), and RNA ligases (RNA, tRNA). SFV= Shope fibroma virus, Vacc.= vaccinia virus, ASF= African swine fever virus, S. cer.= *Saccharomyces cerevisiae*, S. pom.= *Schizosaccharomyces pombe*, D.amb.= *Desulfurolobus ambivalens*, H. sap.= human ligases 1 and 4, T4,T3= bacteriophage, C. alb.= *Candida albicans*. (adapted from Shuman and Ru, 1995)

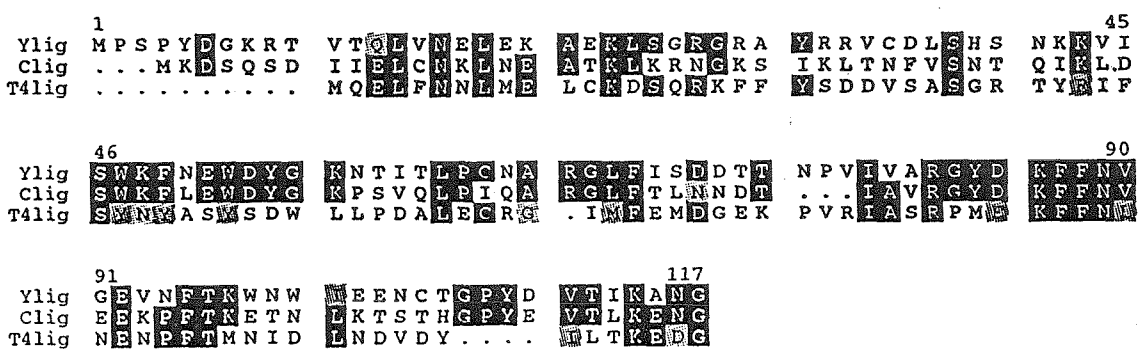
B. Optimal alignment of yeast tRNA ligases and T4 RNA ligase from N-terminus to Motif I. Highly similar or identical amino acids are underlined.

C. Conserved putative nucleotide binding motif. Cpde= human 3'-cyclic nucleotide 3'-phosphodiesterase. Pnk= T4 polynucleotide kinase. Mst= *Streptococcus* multiple sugar-binding transport ATP-binding protein M. S-cer= *S. cerevisiae* tRNA ligase. C-alb= *Candida albicans* tRNA ligase. Braat= *E. coli* branched-chain amino acid transport ATP-binding protein. Shkin= *E. coli* shikimate kinase I.

A.

		Motif I		Motif II		Motif III		Motif IV
SFV	CE	KTDG	-34-	VTLYGEAV	-68-	EGVVL	-9-	DYKIKLDNTTD
Vacc.	CE	KTDG	-34-	VVVFGEAV	-69-	EGVIL	-10-	DFKIKKENTID
ASF	CE	KADG	-32-	TILDGEFM	-78-	DGIIL	-13-	KWKPTWDNTLD
S. cer.	CE	KTDG	-53-	TLLDGELV	-90-	DGLIF	-17-	KWKPEQENTVD
S. pom.	CE	KSDG	-50-	TLLDGELV	-90-	DGLIF	-16-	KWKPKEMNTID
SFV	DNA	KYDG	-43-	FILDAELV	-92-	EGLML	-13-	WLKIKKDHLKT
Vacc.	DNA	KYDG	-43-	ILVDSEIV	-91-	EGLVL	-13-	WLKIKRDYLNE
ASF	DNA	KRNG	-43-	VYLDGELY	-85-	EGAIV	-20-	KLKPLLDAEFI
D. amb.	DNA	KYDG	-41-	FIVEGEII	-96-	EGVMV	-17-	WIKFKRDYQSE
S. cer.	DNA	KYDG	-44-	LILDCEAV	-96-	EGLMV	-18-	WLKLKKDYLEG
S. pom.	DNA	KYDG	-44-	FILDCEAV	-96-	EGLMV	-18-	WLKVKKDYLSG
H. sap.1	DNA	KYDG	-44-	FILDTEAV	-96-	EGLMV	-17-	WLKLKKDYLDG
H. sap.4	DNA	KLDG	-49-	CILDGEMM	-93-	EGIMV	-14-	WLKIKPEYVSG
T4	DNA	KADG	-49-	VLIDGELV	-124-	EGIIL	-16-	KFKEVIDVDLK
T3	DNA	KYDG	-50-	FMLDGELM	-127-	EGLIV	-14-	KLKPECEADGI
T4	RNA	KEDG	-4-	TYLDGDEI	-112-	EGYVA	-6-	HFKIKSDWYVS
S. cer.	tRNA	KANG	-6-	GLEDGTLV	-134-	EGFVI	-11-	FFKYKFEEPYL
C. alb.	tRNA	KENG	-6-	GLSTGDIV	-150-	EGFVI	-16-	FFKYKFEQPYL
consensus		K DG N		DGEAV ECDLM I I		EG V D I M		LK D F E W

B.



C.

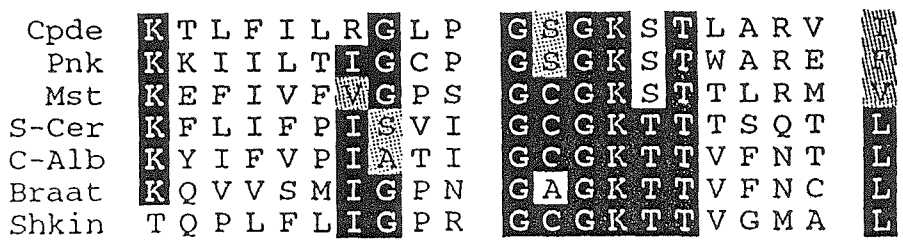


Figure I-6

(55). Various substitutions at the conserved aspartate are active for adenylylation but do not complement. Interestingly, the aspartate of this motif is replaced by an asparagine in the tRNA ligases and also the African Swine Flu virus DNA ligase. The function of the aspartate carboxylate is presumably fulfilled by another residue in these enzymes. Adenylylation activity is preserved if the tyrosine at the second position of this motif in human DNA ligase is changed to alanine or phenylalanine, but only the phe substitution can complement *in vivo*, indicating a requirement for the phenyl ring structure in the second or third reaction step (55). The variability at this position among various ligating enzymes may imply an involvement in specific substrate recognition.

The *S. pombe* mRNA capping guanylyltransferase has been shown to complement a mutant *in vivo* when the conserved motif I tyrosine or aspartate residues are replaced with alanine, but will not tolerate substitution of the lysine or glycine (95). This suggests that the function of the aspartate is only absolutely necessary for the third step of the reaction, and that nucleotidyl transfer can occur at some level in its absence.

Motifs II-IV- The functionality of conserved residues in the remaining motifs have to date been investigated only in two contexts. The first is the *S. pombe* capping enzyme. Alanine substitutions of the acidic glutamate or aspartate in motif II are lethal in vivo, although alterations of the lysine or valine are not (95). A glycine to valine change in this motif, however, confers temperature sensitivity. In the second context that a mutational analysis has been performed in, Vaccinia virus DNA ligase, changing the glutamate of motif II to alanine prevents adenylylation, but altering the aspartate to alanine has no effect on activity (97). Other positions which cannot be substituted by alanine in the capping enzyme include the aspartate and glycine of motif III, and the lysine and apartate of motif IV (95). Because only the ability of mutated enzyme to complement was assayed, these results provide little information on the precise function of these residues, but do indicate the necessity of many of these conserved residues. The conserved

glutamine of motif III and the universally conserved lysine of motif IV have, however, been shown to be essential for adenylylation of Vaccinia DNA ligase *in vitro* (97). Further *in vitro* biochemical analysis of capping and ligating enzymes mutated at these sites should provide additional functional information.

2) tRNA Ligases

The availability of a second tRNA ligase sequence, that of *C. albicans*, which is 26% identical to the sequence of the *S. cerevisiae* enzyme allows for comparisons to determine essential residues in these enzymes. An alignment of these two sequences with that of the T4 RNA ligase, shown in Figure I-6B, shows a significant similarity between all three in the N-terminal region up to and including the active site motif I. The conserved aspartate of this motif, however, is instead an asparagine in the tRNA ligases. An 18 amino acid region is highly conserved among all three, with 50% identity and an additional 17% similarity, and is not highly similar to any other sequences currently found in public protein databases. This therefore represent a novel RNA ligase motif. The critical residues of motifs II-IV as defined above are not as well conserved in the yeast tRNA ligase enzymes, especially motif II, as seen in figure I-6A. The mutagenesis results from the DNA ligase and mRNA capping enzymes demonstrate, however, that not all of the conserved residues in these motifs are essential for function. The addition of more enzyme sequences to multiple sequence alignments as they are determined should help to resolve this question.

Similarity between a central portion of the *S. cerevisiae* tRNA ligase and sequences of PNK, adenylate kinase, and cyclic nucleotide phosphodiesterases have been proposed to indicate an NTP binding site (4, 59). Inclusion of the *C. albicans* sequence and similar sequences from the databases in the alignment shows this region to be well conserved with several ATP-binding proteins and also with GTP binding proteins, as seen in Figure I-6C. This raises the possibility that this sequence represents the kinase

activity nucleotide cofactor binding site, which prefers GTP but can accept ATP (9). The sequence of two regions near the C-terminus of the yeast ligase which had previously been recognized for slight homology to PNK are not well conserved in the *C. albicans* enzymes. The tRNA ligases would therefore appear to have little detectable sequence similarity with PNK or cyclic phosphodiesterases in this region.

3) *E. coli* 2'-5' Ligase

Extensive database searching has revealed only a single sequence which is highly similar to the *E. coli* ligase.(my unpub. res.). Since this similar sequence is derived from a sequenced ORF of unknown function, it does not provide much functional insight. The alignment between the *E. coli* enzyme and the sequence of this ORF from *B. stearothermophilus* does reveal clusters of conserved residues which may prove useful as starting points for mutagenesis studies or for the design of degenerate oligonucleotide primers for use in cloning RNA ligases from other bacteria by PCR.

Conclusion

The RNA ligases are a class of enzymes which have been well-studied biochemically, but whose *in vivo* functions and properties remain for the most part poorly understood. The tRNA ligases, the one ligase type with a well-defined function, may become the best understood overall as more enzyme sequences become available for comparison and the available expression systems are exploited to produce reagent quantities of protein for biochemical experiments. The system utilized for cloning of the *C. albicans* tRNA ligase, expression library complementation of an *S. cerevisiae* ligase knockout, also holds great promise for the cloning of more tRNA ligases and possibly other types of ligases as well. Since the combination of T4 PNK and RNA ligase can

mimic tRNA ligase function *in vitro*, it would be interesting to see whether this combination could complement a yeast tRNA ligase mutant if expressed in tandem. If such a non-specific ligase could complement effectively, then it is conceivable that other ligases such as the animal pathway RNA ligase, archaeal RNA ligase, or mRNA editing RNA ligase could be cloned in the same way (in the presence of PNK if necessary), avoiding the potentially difficult purification of these enzymes which would otherwise be necessary to obtain protein sequence.

The bacterial RNA ligases represented by the *E. coli* and potential *Bacillus* enzymes appear to be a class of enzyme mechanistically and structurally distinct from the other known ligases. Although the in vivo structure and functions of these enzymes are unknown, cloning of the *E. coli* gene has made a variety of genetic and biochemical approaches to this problem possible. The lack of a homologous gene in the recently sequenced *Haemophilus influenzae* genome indicates that this enzyme is not likely to perform a basic housekeeping function of bacterial metabolism but is likely to perform a more specialized or conditional function.

Although still the most widely exploited and thoroughly characterized RNA ligase at the molecular level, the T4 RNA ligase remains without a truly convincing metabolic function. The requirement of RNA ligase for infection of only one clinical isolate of *E. coli* out of all strains ever assayed is at odds with the strict conservation of this enzyme. Investigation of the true extent of occurrence of the anticodon nuclease enzyme in wild *E. coli* populations and in other bacteria as well as a search for RNA ligases in bacteriophage outside of the T-even series may help to shed light on this problem. The conserved nature of RNA ligases from phage to humans indicates that these enzymes are likely to perform some essential functions which remain to be discovered.

Chapter II -

Preparation for Protein Purification; Optimization of Ligase Assay and Source Material.

Introduction

Bacterial RNA Ligase:

The existence of RNA ligase activity in bacteria in the absence of bacteriophage infection was first reported by Greer, Javor, and Abelson in 1983. While assaying for tRNA splicing enzymes in extracts of archaea, these investigators observed an activity capable of performing the ligation step of tRNA splicing in control bacterial extracts. This activity was found to exist in a wide range of bacteria including members of the Alpha and Gamma subdivisions of proteobacteria, green sulfur bacteria, and low G+C content gram positive bacteria (36). Further investigation of this activity in *E coli* extracts revealed a substrate specificity restricted to four of the ten different *S.cerevisiae* tRNA splicing intermediates: tRNA Tyr, Phe, Lys_2, and Trp half-molecules. The reaction mechanism of the *E coli* RNA ligase was found to differ from that of the known eukaryotic and phage RNA ligases in that the reaction did not require a nucleotide triphosphate cofactor and that the product contained an unusual 2'-5' phosphodiester at the ligated junction (36).

The discovery of RNA ligase in bacteria implied the existence of a novel form of bacterial RNA processing. Although self-splicing introns have been found in the tRNA genes of certain cyanobacteria (12) and some proteobacteria (85), no intervening sequences of the type found in eukaryotic nuclear or archaeal tRNA genes (which require processing enzymes for excision and religation) occur in any known bacterial tRNA genes. This includes the full genomic complement of tRNA genes of *E.coli, Mycobacterium capricolum,* and *Haemophilus influenzae* (3, 57). In fact, no RNA processing event which would require the action of an RNA ligase is known to occur in bacteria.

Biological Significance:

The existence of an *in vivo* function for the RNA ligase activity observed *in vitro* is suggested by the occurrence of 2'-5' linkages in native *E coli* RNA. Researchers investigating the production of 2'-5' linked oligoadenylates, which are synthesized in eukaryotes as part of an antiviral defense pathway, found several forms of 2'-5' linked A_{2-4} in acid-soluble extracts of *E. coli* (108). The most abundant species of these oligoribonucleotides observed (at an intracellular concentration over 100nM) was A2'p5'Ap. As this dinucleotide does not correspond to any of the intermediates in the oligoadenylate synthesis pathway, it was postulated by these authors that such molecules are degradation products of RNA containing individual 2'-5' bonds among standard 3'-5' linkages (108). Since the formation of 2'-5' linkages is not catalyzed by RNA polymerase, these bonds must be added in a post-transcriptional processing event by an enzyme such as the 2'-5' RNA ligase.

The unusual 2'-5' structure of the ligation product of the 2'-5' RNA ligase may prove to have biological significance. The presence of a 2'-5' "kink" in the backbone of an otherwise 3'-5' linked RNA chain considerably lowers the potential for base stacking across the altered bond, precluding the participation of adjacent bases in a double-stranded helix, and also renders that bond resistant to most or all *E coli* ribonucleases (71). The function of the *E. coli* RNA ligase may not be simply to covalently join independent RNA molecules but to introduce this altered structural element into a molecule which has been specifically nicked for that purpose. Investigation of the *E coli* 2'-5' RNA ligase enzyme and identification of its *in vivo* substrate(s) should reveal a presently undiscovered RNA processing event and further understanding of the complex pathways of RNA metabolism in bacteria.

In addition to its function in metabolism, investigation of the *E. coli* 2'-5' RNA ligase enzyme may also reveal utilization of a novel mechanism of RNA ligation. The

lack of a requirement for ATP and the unusual 2'-5' linkage formed by this enzyme distinguish its mechanism from the complex energy-requiring reaction pathways of the phage T4 and eukaryotic tRNA splicing RNA ligases. In contrast, the reaction catalyzed by the *E coli* enzyme formally requires only a single phosphotransfer from the 3' position of the cyclic phosphate donor substrate to the 5' hydroxyl group of a poised acceptor substrate. The driving force for this reaction would presumably be provided by the favorable enthalpy of converting the strained geometry of a 2'-3' cyclic phosphate to a linear phosphodiester (33). This is equivalent to a reversal of the initial step of the RNA cleavage reaction performed by common endoribonucleases, except that the phosphoryl transfer is between 3' and 5' hydroxyl groups instead of 2' and 5'. Ligation by the *E. coli* enzyme may therefore proceed by a reversal of the same mechanism involving general base-catalyzed proton abstraction of the attacking hydroxyl and general acid stabilization of the leaving group (30). A variety of chemical modification studies of the enzyme and substrates can be performed to investigate the mechanism of the *E. coli* ligase, but these will require an abundant source of pure ligase protein.

Investigation Strategy:

All of the characteristics of the *E coli* 2'-5' RNA ligase discussed indicate that this enzyme warrants detailed study. Knowledge of the primary structure of the ligase protein, the ability to specifically alter that structure, and an abundant source of purified active enzyme would be crucial to a comprehensive investigation. Since all of these requirements can best be fulfilled by cloning the gene encoding the ligase, several strategies for doing so were envisioned. First, a genetic screen to identify ligase gene mutants in a bank of temperature sensitive strains produced by random mutagenesis was considered. With the *in vivo* substrate of this enzyme unknown, it was not possible to set up a genetic selection for ligase mutants. A genetic method would therefore have entailed screening very large numbers of strains by assaying crude extracts of each for

ligase activity. The success of this method would have depended on the assumption that RNA ligase activity is essential for viability under laboratory conditions. Because of the daunting amount of brute force screening involved and uncertainty about the essential nature of the ligase, this scheme was discarded. A second genetic strategy was also considered which would have involved infecting wild type *E. coli* with an *E. coli* genomic library constructed in bacteriophage λ and screening individual clones for an increased specific activity of RNA ligase. Since it was not sure that an increased ligase gene copy number in the context of a bacteriophage would give a readily detectable increase in activity, and this strategy would also employ a brute force screen of large proportions, it too was rejected.

The unsuitability of proposed genetic screens for cloning the *E. coli* RNA ligase gene prompted consideration of a "reverse genetics" strategy involving purification of the enzyme from *E. coli* extract using biochemical methods. If a sufficient molar amount of native ligase could be purified, a partial amino acid sequence of the protein could be determined by direct chemical sequencing. This amino acid sequence would provide the basis for deduction of possible DNA sequences encoding it and direct the synthesis of degenerate deoxyoligonucleotides which could be used to screen a genomic library for the ligase gene. Since the ligase appeared to be stable and readily detectable in crude extracts, this approach was deemed reasonable and was employed.

Several prerequisites are necessary for a successful enzyme purification. The most important is a reliable assay of enzymatic activity which is sensitive enough to changes in enzyme concentration so that it may be used to accurately monitor the levels of activity in various purification fractions. This assay should be optimized for the nature and concentration of substrate as well as the chemical composition of the reaction mixture and the time and temperature of reaction. It is also useful to determine the richest possible source of the enzyme to minimize the amount of starting material needed to isolate an amount of purified protein sufficient for amino acid sequencing. The identity of the

source organism and growth conditions utilized should be chosen with this in mind. This chapter describes the refinement of the activity assay and preparation of source material for purification of the *E. coli* 2'-5' RNA ligase.

Results

Substrate Preparation:

In order to facilitate the large numbers of ligase activity assays necessary during enzyme purification, a strategy for large scale production of ligation substrates was devised. Assays of bacterial RNA ligase activity had originally been performed as a coupled digestion and ligation reaction using intron-containing tRNA precursors (pre-tRNAs) isolated from splicing deficient *S. cerevisiae* mutants or pre-tRNAs transcribed by RNA polymerase III in *S. cerevisiae* extract from a cloned $tRNA^{Tyr}$ gene template (36). Both these methods of pre-tRNA preparation are laborious, utilize ill-defined systems, and provide low yields. It was therefore decided that the production of large amounts of ligation substrates would be facilitated by transcription of *S. cerevisiae* pre-$tRNA^{Phe}$ with T7 RNA polymerase using an artificial intron-containing $tRNA^{Phe}$ gene as a template. Radiolabeled transcripts produced in this fashion are excellent substrates for the *S. cerevisiae* tRNA splicing endonuclease and ligase (86).

In order to accurately measure the accumulation of ligation products in activity assays without interference from degradation products of pre-tRNA, it was decided to purify tRNA half-molecules by denaturing gel electrophoresis after cleavage with tRNA endonuclease instead of performing a concurrent cleavage/ligation reaction (36). These purified half-molecules could then be mixed and joined together by RNA ligase activities in bacterial extracts or enzyme purification fractions to produce mature tRNA. Intermolecular ligation of radiolabeled tRNA half-molecules can be simply monitored by autoradiography after denaturing polyacrylamide gel electrophoresis.

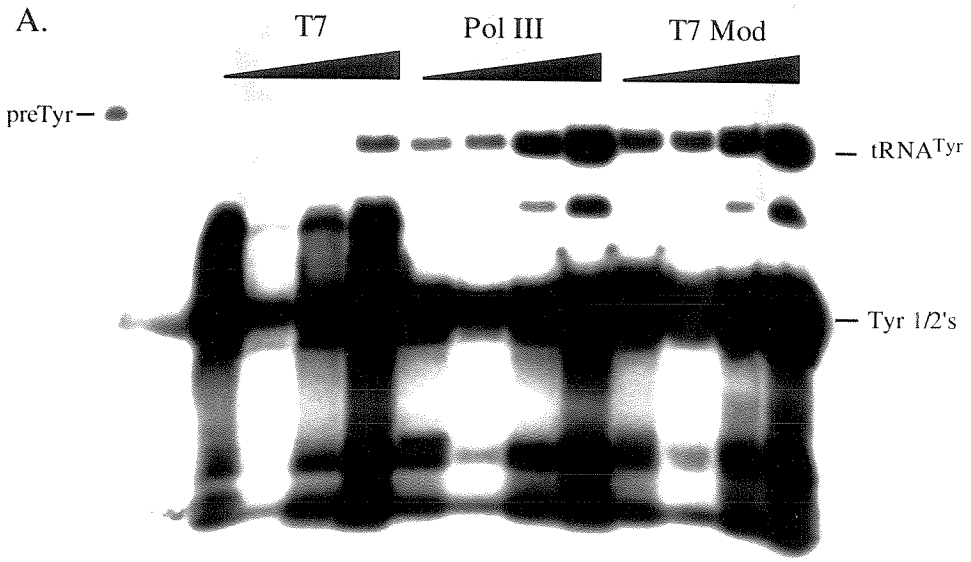
Gel purified $tRNA^{Phe}$ half-molecules produced from pre-$tRNA^{Phe}$ transcribed by T7 RNA polymerase were not ligated to any appreciable extent in *E. coli* extracts under an extensive range of reaction conditions (data not shown). As a control, $tRNA^{Tyr}$ half-molecule substrates were produced by splicing endonuclease digestion of pre$tRNA^{Tyr}$ transcribed by RNA Polymerase III in *S. cerevisiae* extract. Purified half-molecules produced in this manner were ligated in *E. coli* extract, implying that some modification of substrate transcripts was required by the *E. coli* RNA ligase.

In order to test the putative modification requirement, a pre-$tRNA^{Tyr}$ gene under the control of a T7 promoter was created by the polymerase chain reaction (PCR) and cloned into a plasmid vector. Unmodified transcripts produced from this gene using T7 RNA polymerase were extremely poor substrates for the *E. coli* ligase, although they were utilized effectively by the *S. cerevisiae* tRNA splicing ligase. When these transcripts were incubated in the same yeast extracts utilized for Pol III transcription, several nucleotide base modifications occurred, as confirmed by a 2-D thin layer chromatography (TLC) Nishimura analysis of complete nuclease digests (data not shown). Figure II-1A shows that these modified transcripts were then utilized by the *E. coli* RNA ligase. The presence of a 2'-5' linkage in ligated products was confirmed by nuclease digestion and TLC (not shown), indicating that the previously described *E. coli* 2'-5' ligase activity was in fact being observed. Quantification of the ligation product in the substrate titrations shown in Figure II-1A reveals that post-modified T7 transcripts are ligated as efficiently as those produced in yeast extracts (Figure II-1B). This confirmed that the bacterial enzyme requires some base modification of the substrate, although the precise location and identity of necessary modifications were not determined. When pre-$tRNA^{Phe}$ T7 transcripts were modified by incubation in yeast extracts they too became substrates for ligation by the *E. coli* enzyme, although Tyr half-molecules are preferred 3 to 4-fold over Phe as a substrate . Modified T7-transcribed $tRNA^{Tyr}$ half molecules were therefore chosen as the substrate for use in assaying ligase activity during purification.

Figure II-1: Comparative ligation of variously modified tRNATyr half-molecules.

A. Titrations of substrates in joining reactions performed in *E. coli* extracts. T7= transcripts produced by T7 RNA polymerase from an artificial gene, Pol III= transcripts produced by yeast RNA polymerase III in yeast extract, T7 Mod= transcripts produced by T7 polymerase and subsequently incubated in yeast extract.

B. Quantification of ligation products shown in part A.


A.
T7
Pol III
T7 Mod
preTyr
tRNA^Tyr
Tyr 1/2's

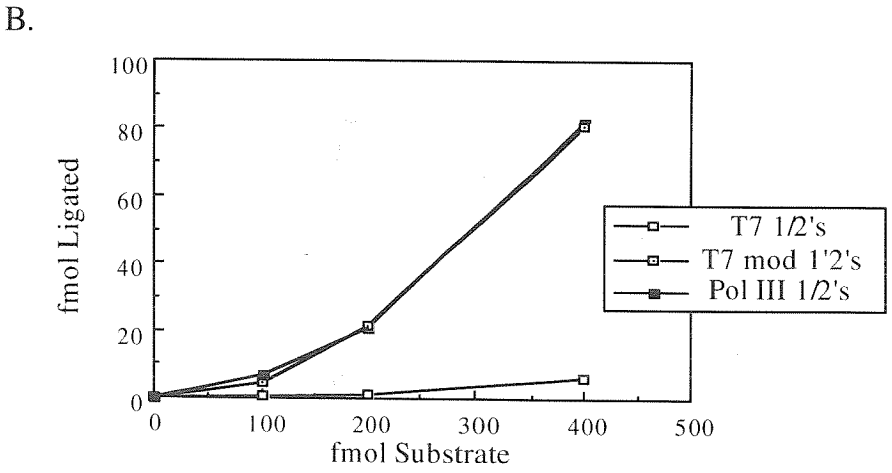

B.
fmol Ligated
fmol Substrate
0
20
40
60
80
100
0
100
200
300
400
500
T7 1/2's
T7 mod 1'2's
Pol III 1/2's

Figure II-1

The melting and denaturation of tRNA half-molecules during the gel purification step of substrate preparation raised concerns about the possible effects of this process on final substrate conformation and ligation efficiency. This was investigated by assaying ligation of gel purified $tRNA^{Tyr}$ half-molecules which were either utilized directly, annealed by heating to >80°C in the presence of Mg^{2+} and slowly cooling to room temperature, or heated to >80°C and snap-cooled on ice. Those substrates which had been annealed by slow cooling proved to be more efficiently ligated than either of the others (not shown). Native gel analysis of these variously treated substrates revealed a shift to a conformer similar to that of ligated product taking place upon productive annealing, suggesting that a particular three-dimensional conformation of hybridized half-molecules is necessary to position their end groups for ligation by the *E. coli* RNA ligase. Such a hybridization experiment is shown in Figure II-2A.

Substrate Specificity:

The specificity of ligation by the *E. coli* 2'-5' RNA ligase was examined to test the feasibility of using more easily prepared model substrates to assay activity and to characterize the ligation reaction. The utilization of tRNA half-molecules as a ligation substrate, inclusion of the cyclic phosphate of the substrate in the ligation junction, and absence of a nucleotide cofactor requirement by the *E. coli* ligase all argued that this enzyme would likely be specific for substrates ending in a 2',3'-cyclic phosphate. This was tested directly by ligation of substrates treated with enzymes to alter cyclic termini. *S. cerevisiae* tRNA splicing ligase was used in the absence of ATP to open the 2',3'-cyclic phosphate of purified $tRNA^{Tyr}$ halves to a 2'-monophosphate. In addition, termini lacking any phosphates were created by subsequent treatment with bacterial alkaline phosphatase. These alterations of terminal phosphate structure were confirmed by TLC of treated substrates after complete nuclease digestion (data not shown). Only those

Figure II-2: Structure and Sequence Specificity of *E. coli* RNA Ligase.

A. Native polyacrylamide gel of modified yeast Tyr half-molecules treated variously: B= boiled and quick-chilled. N= no treatment. H= annealed by heating and slow cooling in reaction buffer with magnesium. Y= annealed as for H and ligated with *S. cerevisiae* tRNA ligase. E= annealed and ligated with *E. coli* RNA ligase. EB= annealed, ligated with *E. coli* RNA ligase, boiled, and quick-chilled.

B. Digestion of *E. coli* $tRNA^{Lys}$ by an activity in *E. coli* extract. 0, 30, 60, and 90 minutes of incubation at 30°C.

C. Religation of $tRNA^{Lys}$ fragments by: mock = extract buffer only, T4 = T4 RNA Ligase, Yeast = partially purified *S. cerevisiae* tRNA splicing RNA ligase, coli = *E. coli* extract.

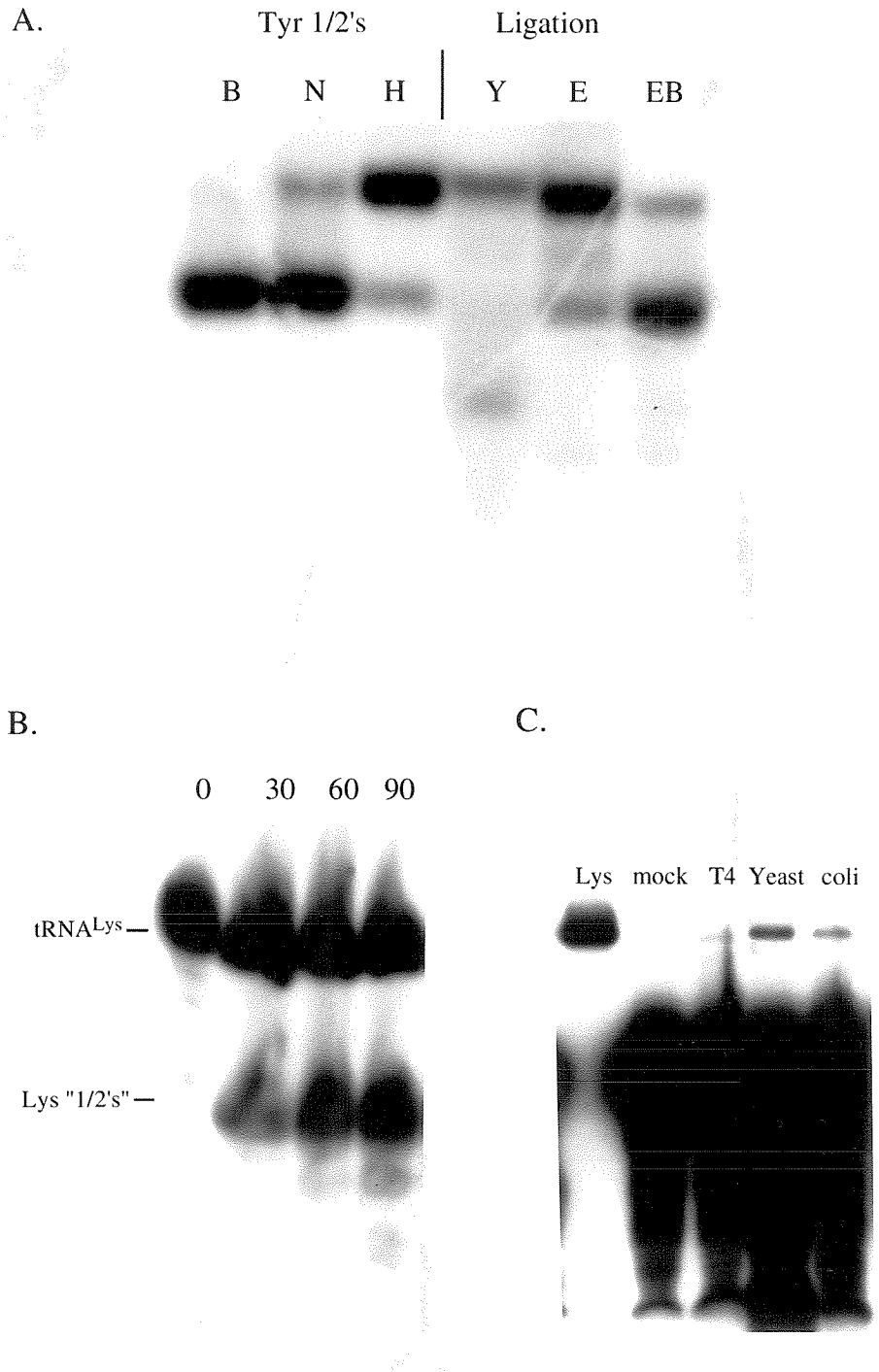

A.
Tyr 1/2's
Ligation
B
N
H
Y
E
EB
B.
0
30
60
90
tRNA^Lys
Lys "1/2's"
C.
Lys
mock
T4
Yeast
coli

Figure II-2

annealed substrates containing an intact 2',3'-cyclic phosphate at the 3' end of the 5' half-molecule were ligated by the *E. coli* RNA ligase (not shown).

The requirement of a 2',3' cyclic phosphate terminus for ligation severely limited the range of substrate molecules which could be tested. Short oligoribonucleotides with cyclic termini were produced by digestion of defined T7 transcripts with ribonuclease N_1, but these were not ligated by the *E. coli* ligase. The one substrate successfully ligated by the *E. coli* ligase which was not an *S. cerevisiae* tRNA splicing intermediate was a cleavage product of *E. coli* $tRNA^{Lys}$. This molecule was originally considered because it is the only known substrate for cleavage by the *E. coli* anticodon nuclease (ACNase), an enzyme known to cleave this tRNA in the anticodon loop, leaving 2',3'-cyclic phosphate and 5'-hydroxyl termini (1). Attempts to overexpress recombinant ACNase for use in the production of possible ligase substrates by *in vitro* digestion were unsuccessful due to the toxic nature of this enzyme. During these experiments, however, an activity was observed in crude extracts of wild type *E. coli* which cleaved T7 transcribed $tRNA^{Lys}$ transcripts (likely also subject to base modifications in the extract) somewhere in the anticodon loop to create the half-sized molecules seen in Figure II-2B. When these $tRNA^{Lys}$ "half-molecules" were gel purified and annealed, they proved to be substrates for ligation by the *E. coli* ligase as well as yeast tRNA splicing ligase but less efficient substrates for T4 RNA ligase alone, indicating the presence of 2',3'-cyclic and 5'-hydroxyl ends on a larger proportion of the molecules (Figure II-2C). Since the enzyme producing these ligatable cleavage products was unidentified and relatively inefficient in crude *E. coli* extracts, this substrate was not adopted for use in ligase activity assays.

Reaction Conditions:

Reaction conditions for the ligation assay utilizing modified yeast $tRNA^{Tyr}$ halves were optimized to allow quantification of ligase activity over a wide range of concentrations with a linear range of detection. Various parameters such as pH, ion

concentration and identity, and temperature were varied to find conditions which enhanced ligation. The results of this optimization are summarized in Table II-1. The reaction was found to proceed better at 30°C than at 37°C, and to be effective across a range of pH from at least 7 - 10, with an optimum around 7.8. The presence of cations was required for activity, with the multivalent polyamine spermidine supporting ligation at concentrations of 2mM while the divalent cation Mg^{2+} alone was required at concentrations greater than 4mM. Ca^{2+} could substitute for Mg^{2+} in the reaction, Zn^{2+} substituted less efficiently, and Ni^{2+} did not support ligation. High concentrations of monovalent salts were inhibitory to the reaction, with concentrations of NaCl above 100mM or KCl above 180mM significantly repressing ligation. Therefore, the only monovalent salt in the standard activity assays was that present in extracts or purification fractions, and this was kept below 100mM. The concentration of substrate used in typical activity assays was 200-250nM $tRNA^{Tyr}$ halves. For critical quantitative assays utilizing partially purified ligase concentrations closer to substrate saturation, i.e., 1.5 μM, were used to insure accurate rate measurements.

	Range Tested	Standard Reaction
Temperature	24-37°C	30°C
pH	7.0-9.5	7.8 (40mM HEPES)
[monovalent salt]	0-200mM	50mM
[Mg^{2+}]	0-50mM	3mM
[spermidine]	0-3mM	2mM
[$tRNA^{Tyr}$ 1/2's]	2-800nM	200nM
[glycerol]	-	5%
Time	10sec-24hrs	2-5 min.

Table II-1: The range of various parameters tested and optimized conditions chosen for standard RNA ligation assays are shown.

A time course of ligation in crude *E. coli* extract under the optimal conditions defined is shown in Figure II-3A. The reaction appears to follow standard Michaelis-Menton kinetics, with the initial rate slowly declining as soon as one minute after initiation and the yield plateau reached after approximately 20 minutes. Typical activity assays were therefore performed for 0.5-5 minute durations (depending on estimated amounts of activity).

Regardless of substrate concentration and time of reaction, a maximum of about 25% of substrate was converted to product. The limiting factor in the ligase assay was identified by a "bump" experiment where a large scale ligation reaction was run to completion, aliquotted, and then supplemented with various reaction components. As Figure II-3B shows, only the addition of annealed substrate was able to support the generation of more product. The ratio of product to substrate at completion remained nearly constant. Although only a minority percentage of added substrate could be ligated in this reaction, the assay provided a reasonable linear range of detection for quantification of RNA ligase activity in purification fractions.

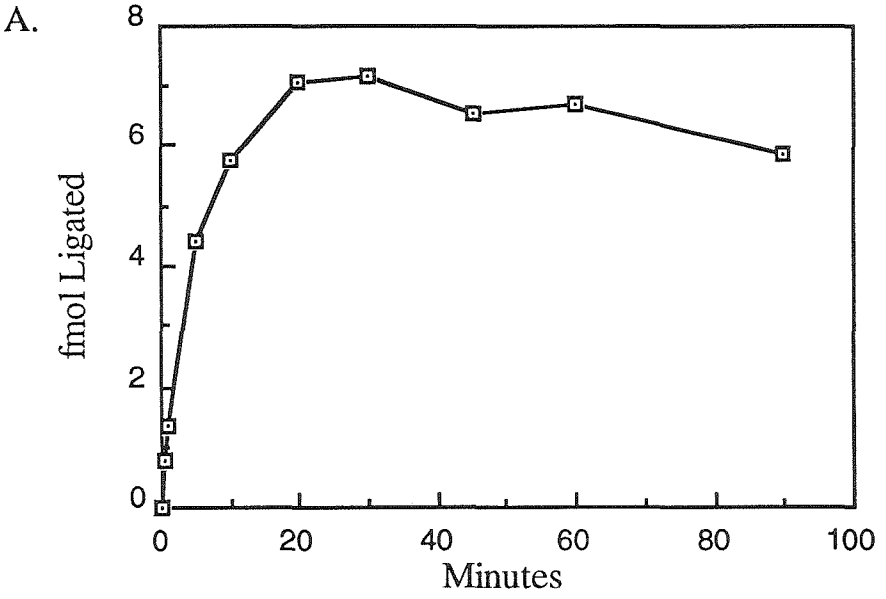
Source Material:

Escherichia coli was chosen as the source organism for purification of RNA ligase even though greater amounts of RNA ligase activity had been observed in extracts of some other bacterial species (36). *E. coli* is attractive as an experimental organism due to the vast amount of information which has been accumulated about its biochemistry and genetics and the fact that the 2'-5' ligase activity had only been previously characterized in this species. Test extracts were prepared from various of laboratory strains of *E. coli* , representing both the B and K lineages, and these were assayed for ligation of $tRNA^{Tyr}$ half-molecules. All strains tested displayed comparable amounts of ligase activity, and none contained excessive amounts of ribonuclease active against the substrate (not shown). *E. coli* strain HB101 was chosen as a source of ligase for purification due to its

Figure II-3: Characteristics of RNA Ligation in Crude *E. coli* Extracts.

A. Time course of $tRNA^{Tyr}$ half-molecule ligation by crude *E. coli* extract at 30°C: quantification of ligation product.

B. Bump experiment - A large scale ligation reaction by partially purified *E. coli* RNA ligase was incubated for 30 min. at 30°C, aliquoted, the indicated reagents added, and incubated for an additional 30 min. 30 min = reaction stop solution, buffer = 1x ligation buffer, enzyme = partially purified *E. coli* ligase, ATP = 2mM ATP, Nucleotide mix = 1.2 mM each ATP,GTP,CTP,UTP, NADH = 2mM NADH


A.
8
6
4
2
0
fmol Ligated
0
20
40
60
80
100
Minutes

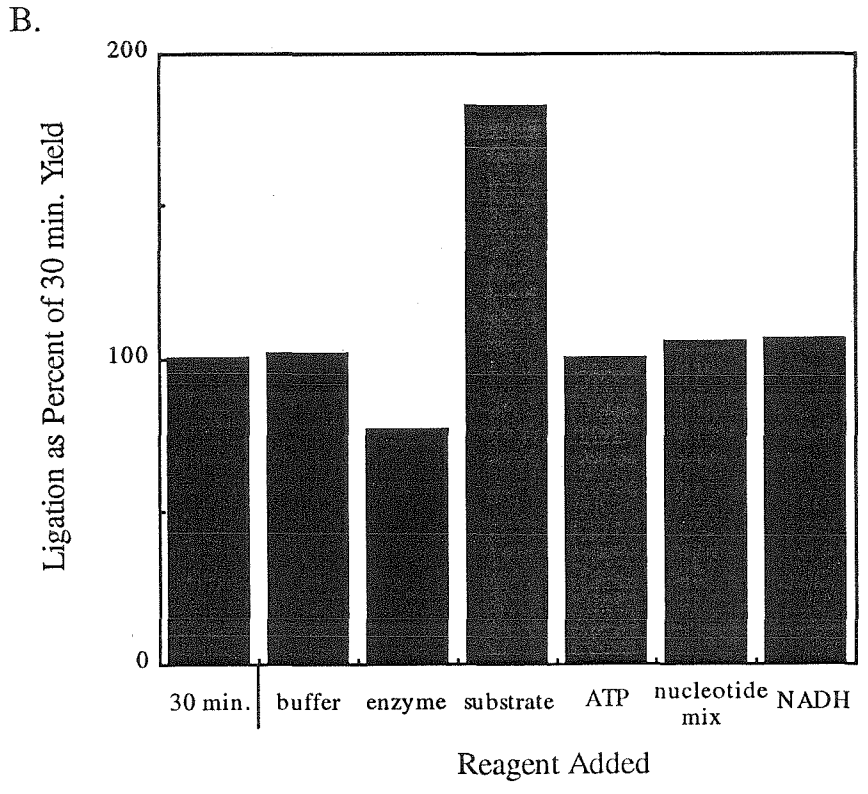

B.
200
100
0
Ligation as Percent of 30 min. Yield
30 min.
buffer
enzyme
substrate
ATP
nucleotide mix
NADH
Reagent Added

Figure II-3

robust growth, ease of handling, and streptomycin resistant phenotype which allowed for growth of wild type cells under selective conditions.

In an effort to minimize the amount of cell mass required for purification of a significant amount of ligase enzyme, HB101 extracts were tested for increased RNA ligase activity after growth under various conditions. For instance, cultures were grown in minimal media to induce the expression of anabolic enzymes but no increase in RNA ligase activity was observed in extracts of these cells. Another possible method of induction of increased ligase activity was suggested by the hypothesis that a ligase might possibly be involved in the repair of damaged RNA, and therefore might be induced under stressful growth conditions. Liquid cultures of HB101 cells were therefore subjected to a variety of physical and chemical stressors which are known to activate expression of distinct but overlapping sets of stress response genes (113). As shown in Figure II-4A, the only condition tested in this screen which resulted in an increase in RNA ligase specific activity was the addition of ethanol to the culture medium. Given this result, a variety of short chain alcohols including methanol, propanol and butanol were also tested for the ability to induce ligase, but only ethanol and methanol were effective. Titrations of the amount of ethanol and methanol added to media showed a maximal increase in specific activity of 50% over uninduced in 5% methanol (Figure II-4B). An assay of ligase activity in extracts prepared at various times after methanol addition showed the induction of activity to be stable for at least two hours (Figure II-4B). The addition of 5% methanol in culture media was therefore adopted to enrich cells for ligase during large scale growth in preparation for purification.

Discussion

Specificity of Ligation:

The variable activity of *E. coli* 2'-5' RNA ligase toward the substrates tested indicates a highly specific substrate preference. This is in stark contrast to, for instance,

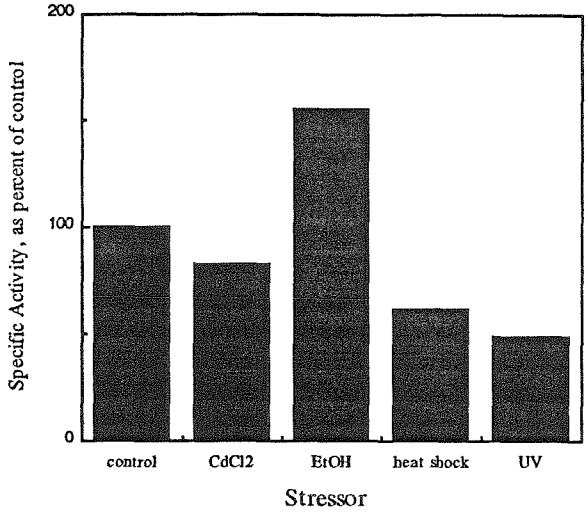
Figure II-4: Induction of RNA Ligase Activity by Alcohol Stress.

A. Specific activity of RNA ligase activity in extracts of *E. coli* cultures treated with various stressors and grown an additional 30 min. before harvest. Control= no stressor, $CdCl_2$= 0.5mM cadmium chloride, EtOH= 5% ethanol, Heat Shock= 43°C, UV= 254nm ultraviolet light, 10 sec.

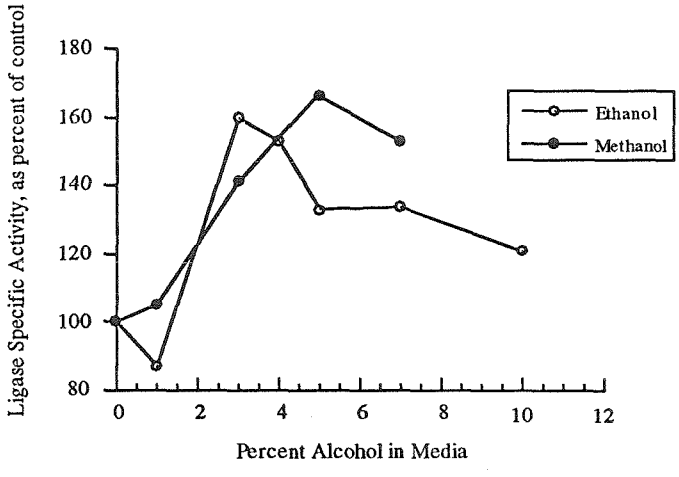
B. Specific activity of RNA ligase in extracts of cultures grown in the various concentrations of alcohol indicated.

C. RNA ligase specific activity in extracts of cells grown in the presence of 5% methanol or LB alone for the times indicated.

A.



B.



C.

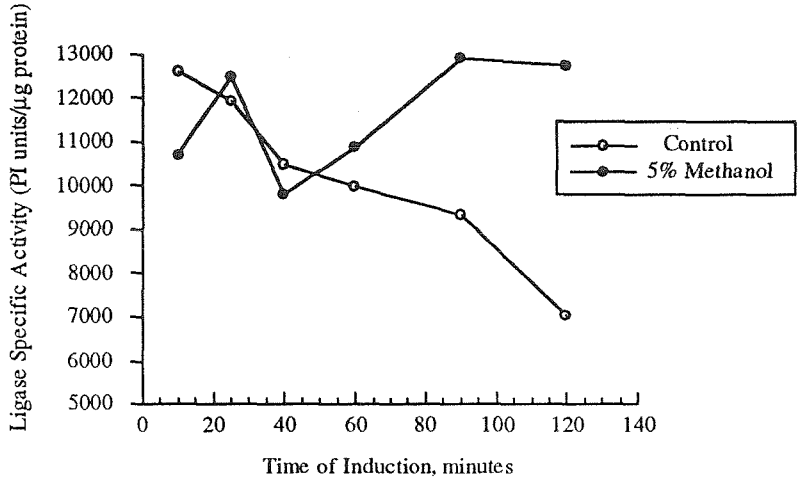


Figure II-4

the bacteriophage T4 RNA ligase which can ligate RNA of almost any sequence or structure. The *E. coli* ligase appears to show an even greater specificity than the *S. cerevisiae* tRNA splicing ligase. This enzyme has been thought to efficiently ligate only the ten different splicing intermediates produced by tRNA splicing endonuclease, but in this work has been shown to also act upon a nicked *E. coli* $tRNA^{Lys}$. It may be risky to speculate too broadly about the properties of an enzyme with unknown function based on activity towards fortuitous heterologous substrates, but some general conclusions can be made based upon consideration of these results. For instance, the formation of a 2'-5' linkage in this reaction insures that it will be specific for ribonucleic and not 2'-deoxyribonucleic acid substrates.

The requirement of *E. coli* 2'-5' RNA ligase for modified nucleosides in its substrates *in vitro* does narrow the range of possible *in vivo* substrates as these modifications occur exclusively in stable RNAs such as tRNA, rRNA, and RNaseP in *E. coli*. (13). These modified residues may be recognized directly by the enzyme for substrate identification, as has been shown to occur in the interactions between some tRNAs and tRNA amino-acyl synthetases (74). A close relationship between tRNA base modification and tRNA splicing has been observed in S. cerevisiae, where a crucial pseudouridine modification in the anticodon of $tRNA^{Tyr}$ is absent in transcripts of a recombinant intronless $tRNA^{Tyr}$ gene (44). It is also possible that base modifications are not necessary for direct recognition by the *E. coli* 2'-5' RNA ligase, but are required to structurally stabilize tRNA splicing substrates in conformers which can be recognized by this enzyme. Modified nucleosides have been demonstrated to stabilize biologically active conformers of mature tRNAs in other systems (5, 82).

The requirement of *E. coli* 2'-5' ligase for nucleoside modifications in substrates, the preference shown by this enzyme for *S. cerevisiae* $tRNA^{Tyr}$ over $tRNA^{Phe}$ substrates, and the sensitivity of the ligase to tRNA conformation after annealing suggest that the *E. coli* ligase is likely to act upon a tRNA or tRNA-like molecule *in vivo*. The ability of the

E. coli ligase to discriminate among tRNA species was first demonstrated by the utilization of only four of the ten *S. cerevisiae* tRNA splicing substrates by this enzyme (36). Comparison of these four species and the nicked *E. coli* $tRNA^{Lys}$ ligated by the *E. coli* ligase to the six yeast tRNAs which are not ligated does not reveal any obvious consensus of sequence or base modifications which might be recognized.

The requirement of *E. coli* RNA ligase for a 2',3'-cyclic terminus in ligation substrates also narrows the range of molecules it might act upon. These 2',3'-cyclic ends on an RNA would almost certainly be the product of an endoribonuclease cleavage. 2',3' cyclic phosphates are an intermediate in the RNAse A-type ribonuclease reaction mechanism, and such intermediates are often released before they can be converted to the final 3'-phosphate product (27). Few ribonucleases actually produce 2',3'-cyclic phosphate as a final end product, most notably perhaps the eukaryotic tRNA splicing endonuclease and the *E. coli* anticodon nuclease, both of which act upon tRNA. The cyclic phosphate requirement, therefore, suggests that the *E. coli* 2'-5' RNA ligase acts to ligate ends produced by ribonuclease action upon a tRNA or similar stable RNA *in vivo*. The existence of an activity in *E. coli* extracts capable of creating a substrate for the 2'-5' RNA ligase from $tRNA^{Lys}$ transcripts also provides more evidence of a biological role for ligation by this enzyme.

Optimization of reaction conditions for the RNA ligase assay was successful in creating a sensitive, quantitative measure of enzymatic activity for use in purification and provided some insight into the nature of the enzyme or reaction mechanism. The ability of spermidine to support the ligation reaction as the sole cation present indicates that multivalent cations are likely to be promoting correct RNA structure or screening like charge repulsion and probably do not participate directly in catalysis. The lack of an absolute divalent metal requirement suggests that the metal catalysis mechanism proposed for many phosphotransfer reactions is not employed by the *E. coli* ligase (101).

The minority percentage of substrate ligated at completion of the reaction by *E. coli* RNA ligase may represent a true feature of the ligase's kinetic behavior, with the ratio of product to substrate at maximal ligation reflecting the internal equilibrium of the enzyme. This would be consistent with a first order phosphotransfer proceeding by a reverse ribonuclease mechanism, as discussed earlier. Another possible explanation for the failure to ligate all substrate is that a majority of the substrate added was not competent for ligation. Incompetent substrates might include half-molecules whose cyclic terminus has been disrupted by cyclic phosphodiesterase activity or whose hydroxyl termini has been phoshorylated in the crude *E. coli* extracts, transcripts which were not subject to all necessary base modifications during exposure to *S. cerevisiae* extract, or half-molecules which formed unligatable conformers during the annealing procedure. Further meaningful investigation of the kinetics and mechanism of the *E. coli* RNA ligase will require highly purified enzyme preparations and will benefit from the identification of a natural *in vivo* substrate.

Induction of Ligase Activity by Alcohols:

An increase in specific activity of 2'-5' RNA ligase activity in *E. coli* cells was observed after addition of ethanol or methanol to liquid culture media. This apparent induction of activity provides a useful means of enriching cell extracts in this enzyme. The functional implications of this response for the RNA ligase however, are unclear. The presence of environmental ethanol, which alters water activity and may directly cause disruption of plasma membrane structure, has long been recognized as a canonical stressor for bacterial cells, eliciting expression of stress response genes and subsequent increased resistance. At least one way this response is activated is through the transmembrane kinase osmoreceptor KdpD, which activates the positive transcription factor KdpE in response to low concentrations of environmental ethanol (104). Within minutes after exposure to ethanol the expression of a number of proteins is induced,

including all members of the heat shock regulon but none typical of the SOS or oxidative stress responses (113). Exposure to ethanol or other non-mutagenic stresses also induces an increase in UV resistance and a lowering of UV-induced mutation frequencies via activation of an error-free DNA excision repair system (31) Despite the overlap of response mechanisms to different stressors, the induction of RNA ligase activity must involve some response pathway specific to alcohols as it did not occur after heat shock or UV radiation of *E. coli* cultures. Since alcohols are unlikely to be causing direct damage to RNA, and because the *in vivo* ligation substrate is likely to be created by specific endonuclease digestion, a function for the *E. coli* ligase in the repair of damaged RNA seems unlikely. The reason for the observed connection between alcohol-induced stress and 2'-5' RNA ligase levels therefore remains mysterious.

Materials and Methods

Strains and plasmids:

E. coli strains utilized - HB101: F', hsdS20(r_B', m_B'), recA13, ara-14, proA2, lacY1, galK2, rspL20(Sm^r), xyl-5, mtl-1, supE44, λ^-. XL1-Blue (for cloning): recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, [F' proAB, lac I^q, ZDM15, Tn10(Tet^r)].

Substrate Preparation: Pol III transcription of pre-$tRNA^{Tyr}$ in yeast extract was performed by the method of Evans and Engelke, using extracts prepared as described (26). The template for Pol III transcription of pre-$tRNA^{Tyr}$ was pYSUP6 (115).. T7 transcription was performed as described by Sampson and Saks (91). pre-$tRNA^{Phe}$ was transcribed from the artificial pre-tRNA gene of Reyes, cloned into pUC13 (86). Template for pre-$tRNA^{Tyr}$ transcription was created by PCR amplification from pYSUP6, adding a T7 promoter at the 5' end and a BstNI site at the 3' end of the pre-tRNA, as well as altering C1->G, T2->G, A87->C, and G88->C, and cloning into pBlueScript (Stratagene). Wild-type *E. coli* $tRNA^{Lys}$ was amplified from genomic DNA

and cloned similarly. Yeast tRNA substrates were modified in the extracts and conditions described for Pol III transcription, for 45 minutes at room temperature. tRNA precursors were cleaved as described by Peebles, using partially purified endonuclease fractions from the hydroxyapatite step or later, as described by Rauhut (72, 84). T7-transcribed *E. coli* $tRNA^{Lys}$ was modified and cleaved in an S-30 extract of *E. coli* C600, prepared as described for ACNase extracts (2).

All RNA was gel purified in 8%(20:1 acryl:bis) acrylamide, 1xTBE, 7M urea gels and visualized by autoradiography. RNAs were eluted from crushed gel slices in 0.6M NH_4OAc, 2mM EDTA, 0.005% NP-40 at room temperature for 15-20 minutes. All RNA was phenol/chloroform and chloroform extracted, then ethanol precipitated in the presence of 0.75M NH_4OAc and 60μg/ml glycogen carrier, then resuspended in distilled water before use.

Ligation Assay: (Ligations performed as described in Results.) Annealing of half-molecules was performed in 2x ligation buffer by heating to 85°C, and slow cooling to 30°C over 20-30 minutes. Reactions were stopped by the addition of 1.5 mg/ml proteinase K, 0.04% SDS, 5mM EDTA and 0.001 mg/ml *E. coli* total RNA, then incubated for 10-20 minutes at 30°C. An equal volume of 90% deionized formamide plus tracking dyes was added, and reactions incubated at 65°C for 5 minutes before cooling on ice and loading onto 10% gels as described for purification. After electrophoresis, gels were visualized by autoradiography or by exposure to a storage phosphor plate for quantification and analysis on a Molecular Dynamics Phoshorimager using Imagequant software.

Culture Growth and Stress Conditions: Cultures were grown in 1x LB media at 37°C unless otherwise stated. Absolute ethanol or anhydrous methanol was added to various percentages (vol/vol) as indicated.

Chapter III:

Purification, Cloning, and Genomic Disruption of the 2'-5' RNA Ligase.

Introduction

In order to perform detailed studies of the function, mechanism and structure of the *E. coli* 2'-5' RNA ligase, it was desirable to clone the gene encoding it by the method of reverse genetics. A purification of the enzyme from crude extracts was therefore undertaken to provide material for peptide sequencing. Small scale testing of a variety of centrifugation, precipitation and chromatographic methods for protein fractionation was utilized to determine general properties and behavior of the ligase enzyme. Ligase activity was monitored throughout the procedure by the optimized $tRNA^{Tyr}$ half-molecule ligation assay. Various combinations of purification steps which promised to provide effective separation were then tested using larger amounts of cell mass to assemble a purification scheme.

Results:

Development of a Purification Scheme:

A wide variety of bulk protein separation techniques were tested for effectiveness in the purification of *E. coli* 2'-5' RNA ligase using small scale (0.5-5 ml) preparations of crude extract. Initially, various precipitation methods were tested in order to remove nucleic acids from extracts and provide a crude separation of proteins. However, ligase activity was found to precipitate across a wide range of concentrations of ammonium sulfate, streptomycin sulfate, and polyethyleneimine rendering these procedures inappropriate. Tests of high-speed centrifugation showed that the ligase remained in solution after an S-100 spin (100,000xg for 1 hour), which should clear the extract of

ribosomes and chromatin as well as membrane debris. This spin was therefore adopted as the initial step in purification.

A variety of ion exchange, affinity, and gel filtration matrices were also tested for utility in the purification of *E. coli* RNA ligase. The ligase behaved consistently as a highly basic protein during ion exchange chromatography, allowing both anionic and cationic resins to be utilized for purification. Tests of ligase binding to various affinity resins commonly used to purify nucleic acid binding proteins, however, showed considerable differences in affinity. A low affinity was seen for DNA- and poly(ribo)adenine- affinity resins, but tight binding was observed for phosphate, heparin, and tRNA affinity resins which were therefore chosen for use in large scale purification. Binding to artificial blue dye resins proved to be too avid for recovery of active protein in a reasonable yield. The elution profile of *E. coli* RNA ligase from hydroxyapatite was inconsistent and this matrix was not utilized for purification. The results of gel filtration at various salt concentrations indicated the ligase protein was likely to be a low molecular weight globular protein, and not an integral part of any large macromolecular complex. Gel filtration on matrices optimized for separation of low molecular weight proteins was therefore employed for purification. The results from these various tests allowed selection and ordering of various fractionation procedures to form an effective scheme for large scale purification of the *E. coli* 2'-5' RNA ligase.

Preparative Purification:

Several large scale purifications utilizing 100-1000g of frozen HB101 cells were undertaken, and the results used to refine the purification scheme. HB101 cells used as source material for purification were produced by fermentation of cultures were grown in rich media utilizing glycerol as a carbon source, and fed continuously to replace nutrients. RNA ligase activity was induced by the addition of methanol to the medium. Harvested cells were stored frozen at -70°C until use.

An optimized protocol for purification was designed and followed, as detailed in Materials and Methods. Briefly, cells were disrupted by sonication and crude extracts were subjected to an S-100 spin. S-100 supernatant was mixed with DEAE anion exchange resin and the unbound supernatant loaded onto a cellulose phosphate cation exchange and phosphate affinity column. RNA ligase activity was eluted by a salt gradient and active fractions pooled for further purification by sequential binding to and salt gradient elution from heparin and *E. coli* tRNA-sepharose affinity columns. Peak active fractions from affinity purification were subjected to gel filtration through Superdex 75 media. Peak activity fractions from gel filtration were observed by silver stained SDS-PAGE to contain several polypeptides of molecular weights from 15-30 kD (not shown). Superdex 75 peak fractions were pooled for binding to a column of *S. cerevisiae* tRNA-sepharose. When ligase activity was eluted from this matrix and peak activity fractions examined by SDS-PAGE and silver staining, several polypeptide species remained. A quantitative analysis of this purification procedure and demonstration of the complexity of the protein mixture at each step is given in Figure III-1.

Since the purification scheme followed did not provide homogeneous ligase enzyme for sequencing, separation by denaturing electrophoresis was utilized with the aim of reconstituting ligase activity from individual excised protein bands. Ligase activity could be reconstituted after detergent denaturation and electrophoresis, but the small amounts of protein left for analysis did not allow for identification of the polypeptide responsible. A larger scale purification was therefore attempted, utilizing 4 kilograms of methanol-induced HB101 cells as starting material, with the aim of purifying enough protein for preparative SDS-PAGE and reconstitution. The purification procedure followed for this very large scale preparation was essentially the same as for the previous preparation, except that the *S. cerevisiae* tRNA column was substituted for the initial *E. coli* tRNA affinity step and the tRNA affinity pool was run over Superdex 75 twice. The activity peak fractions were pooled, concentrated by dialysis versus PEG 20k, and a large

Figure III-1: Purification of *E. coli* RNA Ligase Enzyme.

A. Purification profile for large scale preparation of *E. coli* RNA Ligase described in Materials and Methods. ND= not determined because total protein amounts unavailable.

B. Silver-stained SDS-PAGE gel of various purification fractions from this preparation. MW= molecular weight standards, 1 = crude extract, 2 = S-100 supernatant, 3 = DEAE flow through, 4 = cellulose phosphate, 5 = heparin hyper-D, 6 = *E. coli* tRNA sepharose, 7 = heparin II, 8 = superdex 75, 9 = *S. cerevisiae* tRNA sepharose.

A.

Fraction	Specific Activity fmol ligated/mg protein/min.	Purification Factor	Yield
Crude Extract	17.5	1	(77%)
S-100 Supernatant	22.8	1.3	100%
DEAE Flow-Through	246.5	14	ND
Cellulose Phosphate	1517	87	60%
Heparin Hyper-D	2381	136	37%
E. coli tRNA Sepharose	2780	159	23%
Heparin Hyper-D II	5826	333	ND
Superdex 75	8606	492	23%
S. cerevisiae tRNA Sepharose	18930	1082	17%

B.

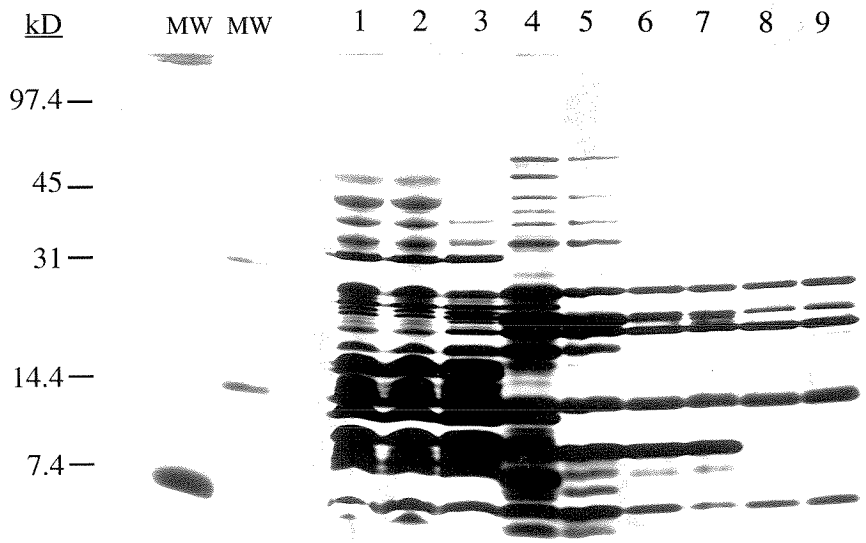


Figure III-1

sample separated by SDS-PAGE. Individual protein bands were visualized by negative staining with $CuCl_2$, excised, crushed, and eluted into an SDS-containing buffer. Eluted protein was precipitated by acetone in the presence of BSA carrier and resuspended in a buffer containing guanidine-HCl. After microdialysis, samples were tested for ligase activity and examined by SDS-PAGE and silver staining. Figure III-2 shows that only an eluted fraction containing a 20kD protein reconstituted RNA ligase activity, and that this activity was not affected by mixing with other eluted fractions. This 20kD protein was designated as the *E. coli* RNA ligase.

Ligase Protein Sequencing and Identification of the Ligase Gene:

In order to identify the *E. coli* RNA ligase gene, amino acid sequence was obtained from the protein identified by reconstitution. A sample of the concentrated Superdex activity pool was subjected to SDS-PAGE, electroblotted onto PVDF membrane, and individual protein bands visualized by staining with Coomassie blue. The 20kD ligase band was excised from the membrane and submitted to the Caltech Protein Analysis Facility for sequencing. 15 residues of N-terminal amino acid sequence were obtained and are shown in Figure III-3A. The sequence did not match any known proteins or predicted ORFs in the current sequence databases. The degenerate oligodeoxynucleotides corresponding to opposite strands of possible coding sequences for internal segments of the ligase sequence fragment as shown in Figure III-3A were synthesized for use in PCR. Figure III-3B shows that these oligos successfully amplified a product of approximately 55bp from *E. coli* whole cells or genomic DNA, but produced no product from *S. cerevisiae* cells or in reactions lacking template. The amplified DNA fragment was cloned and sequenced. Several DNA sequences were obtained encoding a portion of the amino acid sequence obtained from the ligase protein, confirming the presence of the ligase gene on the *E. coli* chromosome. Although several mismatches were found in the primer regions of these clones (presumably due to the low

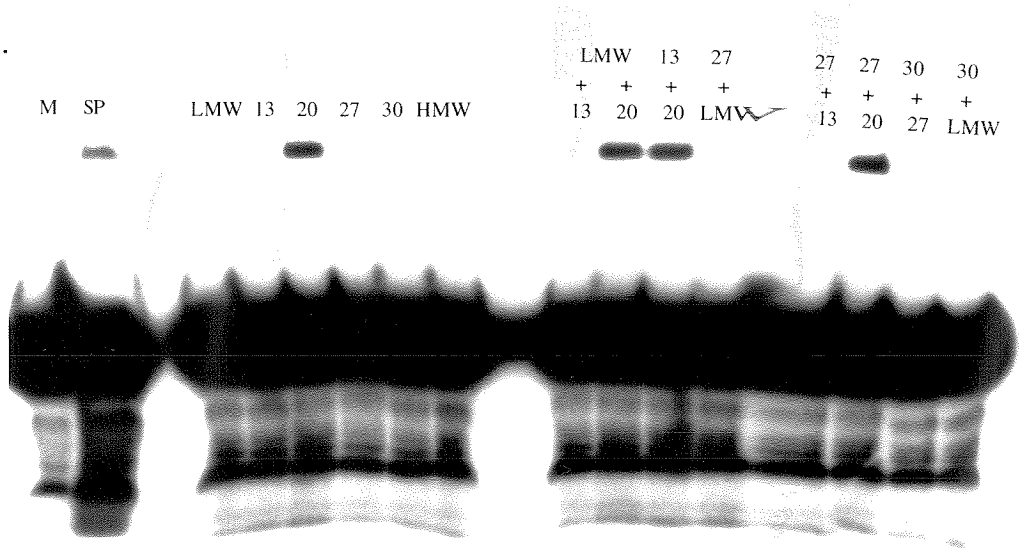
Figure III-2: Identification of the RNA Ligase Polypeptide.

A. Ligation activity assay of polypeptides and mixtures of polypeptides of stated molecular weights (in kD) after elution from a preparative SDS-PAGE gel and renaturation. M = mock, SP = Superdex pool, LMW = 3-12kD, HMW =35-100kD.

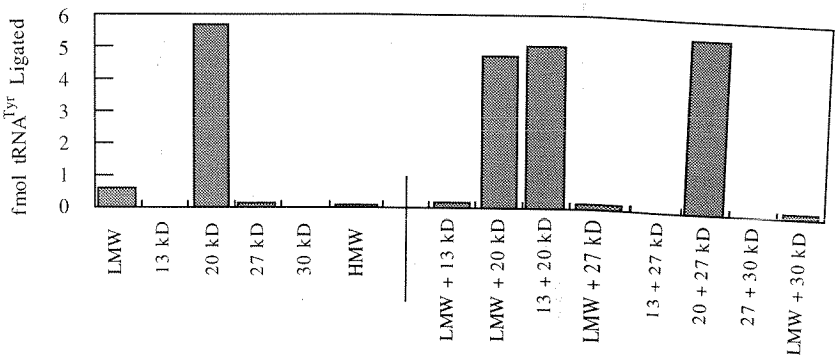
B. Quantification of ligated product in reactions from A.

C. Silver-stained SDS-PAGE of renatured fractions. MW = molecular weight standards, BA = bovine serum albumin

A.



B.



C.

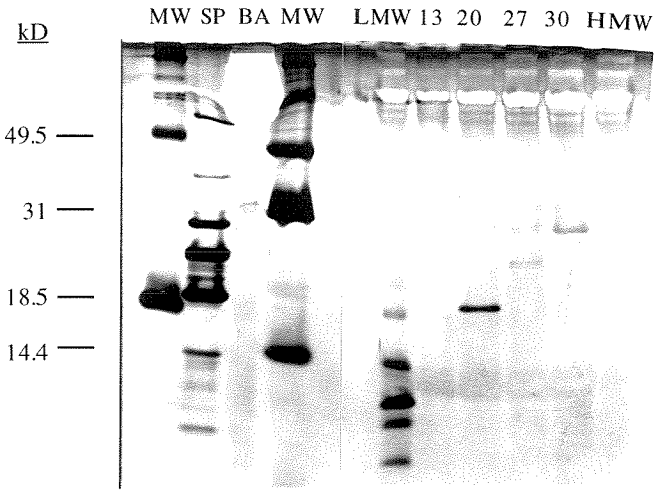


Figure III-2

hybridization temperatures used for PCR with complex degenerate oligos), the 14 bp of sequence internal to the primers was identical in all cases. The 14bp unambiguous DNA sequence did not match any entries in the available sequence databases. It was therefore thought to be necessary to clone the gene from a genomic library.

In order to identify genomic clones carrying the RNA ligase gene, a radiolabeled probe for the ligase gene was created for hybridization to a commercially obtained *E. coli* genomic mapping membrane. PCR was performed with the previously utilized degenerate oligos in the presence of P^{32}-labeled dATP and the product hybridized to the mapping blot containing an array of ordered, overlapping genomic phage clones covering approximately 95% of the *E. coli* genome. Hybridization of labeled PCR product to this membrane revealed two positive clones from the 3.5 minute region of the *E. coli* chromosome which overlapped by approximately 10 kbp (Figure III-3C+D). Since complete sequencing of the 0-4 minute region of the *E. coli* genome had recently been announced by a research group at the Institute for Viral Research in Kyoto, the 15 amino acid peptide sequence and 14bp of unambiguous DNA sequence were then sent to them for a search of their private database. These researchers found exact matches to both sequences in a theoretical ORF at 2.8 minutes on the chromosome, and kindly provided nucleotide sequence for a 3kbp region surrounding this ORF.

Analysis of the 2'-5' RNA Ligase Gene:

The DNA sequence of the ligase ORF and flanking regions allowed prediction of the complete sequence of the ligase protein as a 176 amino acid polypeptide with a molecular weight of 19,934 D (see Figure III-4), corresponding exactly to the size predicted from migration on SDS-PAGE. A methionine is encoded immediately prior to the first residue of the N-terminal sequence obtained from the purified protein, most likely representing the N-terminal of the protein as translated. When the second residue encoded in a gene is a serine, as in the ligase ORF, the N-terminal methionine is

Figure III-3: Localization of the *E. coli* RNA Ligase Gene.

A. Determined N-terminal sequence of 20kD ligase protein and degenerate oligo sequences used to amplify a genomic fragment.

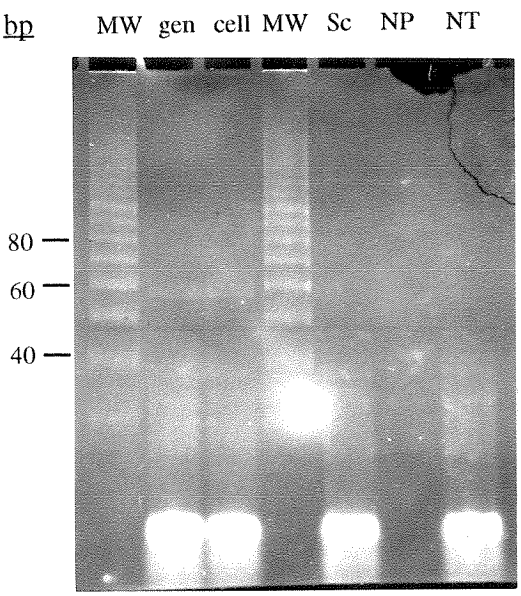
B. PCR amplification of an *E. coli* genomic DNA fragment using the oligos findicated in A as primers. MW = molecular weight standards, gen = *E. coli* genomic DNA template, cell = *E. coli* whole cells as template, Sc= *S. cerevisiae* cells as template, NP= no primers, NT= no template.

C. Hybridization to an *E. coli* genomic mapping membrane with labeled PCR product from B. Positive clones are indicated.

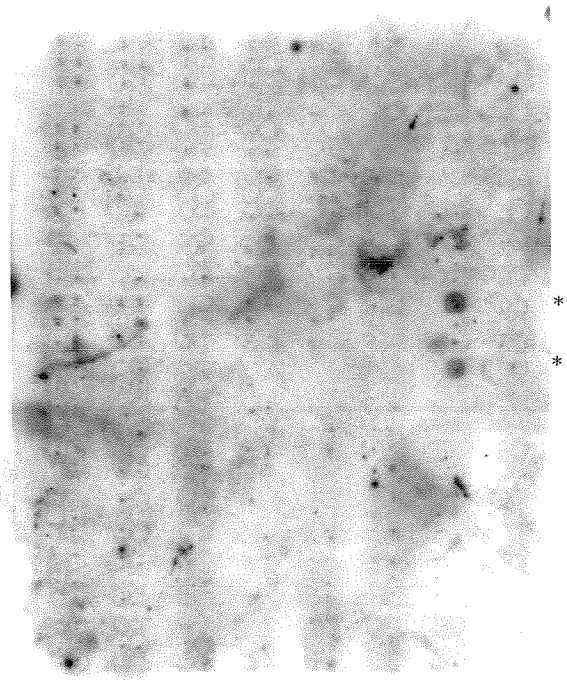
D. Location of the positive phage clones from mapping membrane hybridization on the physical map of the *E. coli* chromosome.

A.

B.



C.



D.

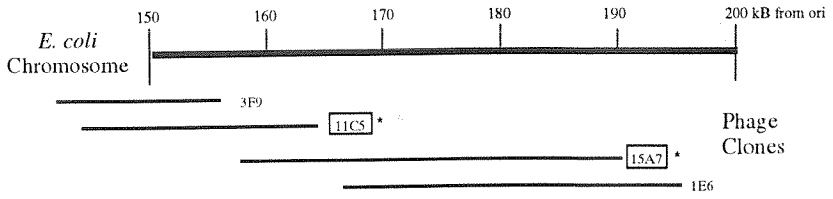


Figure III-3

commonly removed *in vivo* by methionine aminopeptidase *in vivo* (67). Based upon this sequence, *E. coli* RNA ligase mRNA would be transcribed from the *E. coli* chromosome in the counterclockwise direction and would contain a Shine-Delgarno ribosomal recognition sequence beginning 14 bp upstream of the coding region (Figure III-4). A possible weak match to the gearbox promoter consensus sequence can be found beginning at -119 and a possible σ^S -35 region (which is often found without a -10 consensus) can be seen at -293. The nearest potential σ70 housekeeping-type promoter consensus is found at -749bp. The more distal sigma 70 promoter region also contains a potential CRP binding site starting at -798. The ligase does not appear to be expressed as part of a multicistronic operon as the nearest upstream ORF transcribed in the same direction is located about 12 kb away and the adjacent downstream ORF (separated by only 15 bp) encoding the Sfs1 protein has its own promoter and regulatory elements (49). The ligase ORF predicts a very basic protein with a net charge of +7 and an isoelectric point of 11.2 composed of 22 mole percent charged amino acids (14 acidic and 26 basic residues) and 38 mole percent small hydrophobic residues (I,L,M,V) with only 10 percent aromatic residues.

Comparison of the ligase protein to sequence databases using various algorithms identified only one highly similar sequence: that predicted from a theoretical ORF of unknown function found adjacent to the *malA* gene in *Bacillus stearothermophilus* (63). When aligned, the *E. coli* and *Bacillus* sequences were found to be 24% identical and almost 50% similar over the entire length of the *E. coli* ligase (Figure III-5A). The major difference between the two is that the *Bacillus* protein has a C-terminal extension of 129 amino acids. This C-terminal extension alone does not have any significant matches in the protein databases. The fact that the global alignment between these two sequences contains short blocks of high similarity as well as a lower dispersed similarity throughout suggests that the alignment is meaningful and may represent a homologous origin and function for these proteins. The two blocks of sequence which are most highly

Figure III-4: Sequence of the *E. coli* genomic region including the RNA ligase gene, with translations. Starts of surrounding ORFs including a potential helicase homolog and the *Sfs1* gene are indicated. S-D= Shine-Delgarno ribosome binding site. -35 and -10= consensus *E. coli* promoter sequences, gbox -35 and -10 = potential gearbox promoter consensus sequences, σ^S -35= potential stress-induced promoter consensus sequence.

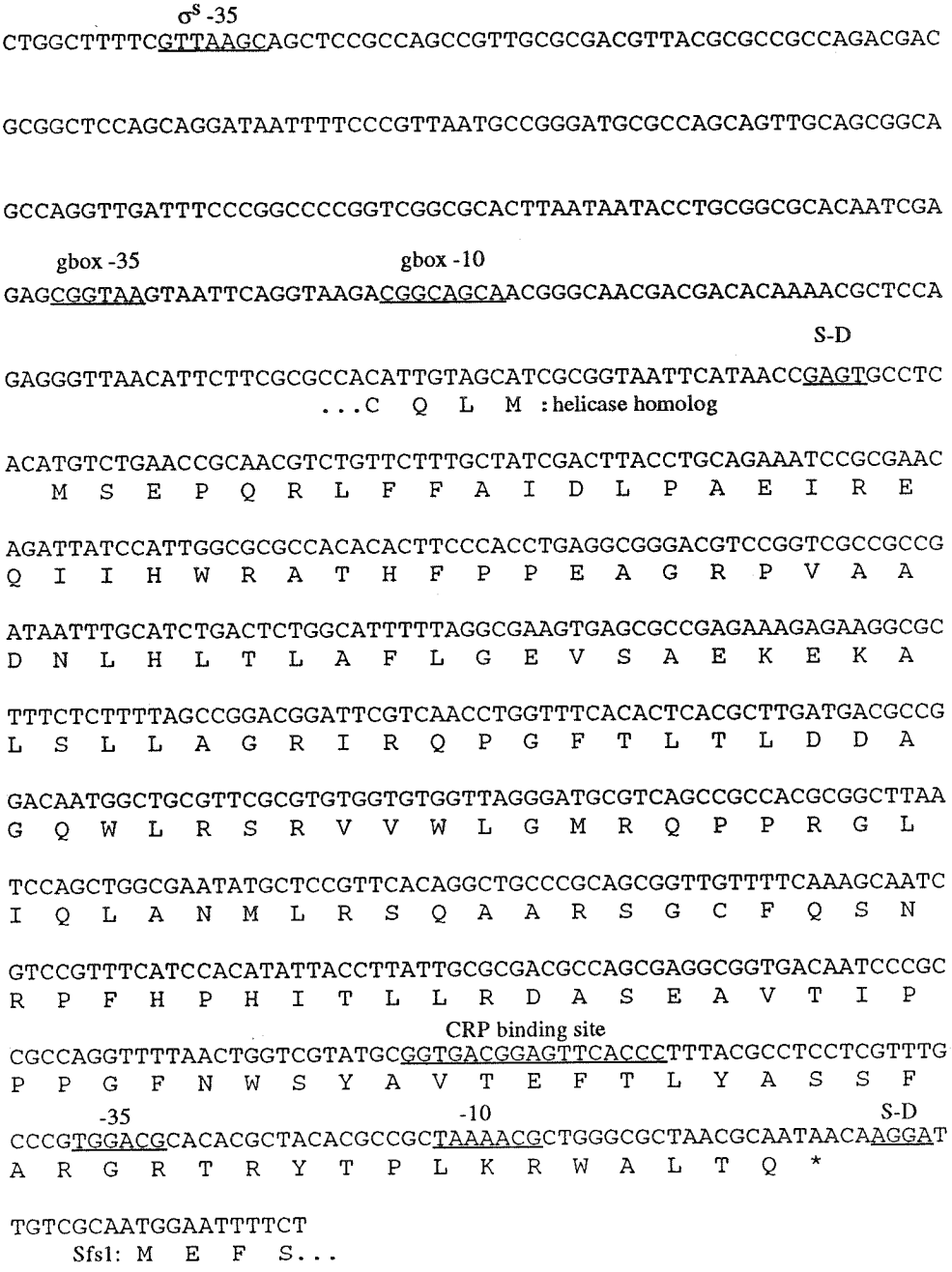

σS -35
CTGGCTTTTCGTTAAGCAGCTCCGCCAGCCGTTGCGCGACGTTACGCGCCGCCAGACGAC
GCGGCTCCAGCAGGATAATTTTCCCGTTAATGCCGGGATGCGCCAGCAGTTGCAGCGGCA
GCCAGGTTGATTTCCCGGCCCCGGTCGGCGCACTTAATAATACCTGCGGCGCACAATCGA
gbox -35
gbox -10
GAGCGGTAAGTAATTCAGGTAAGACGGCAGCAACGGGCAACGACGACACAAAACGCTCCA
S-D
GAGGGTTAACATTCTTCGCGCCACATTGTAGCATCGCGGTAATTCATAACCGAGTGCCTC
...C Q L M : helicase homolog
ACATGTCTGAACCGCAACGTCTGTTCTTTGCTATCGACTTACCTGCAGAAATCCGCGAAC
M S E P Q R L F F A I D L P A E I R E
AGATTATCCATTGGCGCGCCACACACTTCCCACCTGAGGCGGGACGTCCGGTCGCCGCCG
Q I I H W R A T H F P P E A G R P V A A
ATAATTTGCATCTGACTCTGGCATTTTTAGGCGAAGTGAGCGCCGAGAAAGAGAAGGCGC
D N L H L T L A F L G E V S A E K E K A
TTTCTCTTTTAGCCGGACGGATTCGTCAACCTGGTTTCACACTCACGCTTGATGACGCCG
L S L L A G R I R Q P G F T L T L D D A
GACAATGGCTGCGTTCGCGTGTGGTGTGGTTAGGGATGCGTCAGCCGCCACGCGGCTTAA
G Q W L R S R V V W L G M R Q P P R G L
TCCAGCTGGCGAATATGCTCCGTTCACAGGCTGCCCGCAGCGGTTGTTTTCAAAGCAATC
I Q L A N M L R S Q A A R S G C F Q S N
GTCCGTTTCATCCACATATTACCTTATTGCGCGACGCCAGCGAGGCGGTGACAATCCCGC
R P F H P H I T L L R D A S E A V T I P
CRP binding site
CGCCAGGTTTTAACTGGTCGTATGCGGTGACGGAGTTCACCCTTTACGCCTCCTCGTTTG
P P G F N W S Y A V T E F T L Y A S S F
-35
-10
S-D
CCCGTGGACGCACACGCTACACGCCGCTAAAACGCTGGGCGCTAACGCAATAACAAGGAT
A R G R T R Y T P L K R W A L T Q *
TGTCGCAATGGAATTTTCT
Sfs1: M E F S...

Figure III-4

Figure III-5: Sequence Comparisons with the *E. coli* RNA Ligase.

A. Alignment of *E. coli* RNA ligase protein sequence and a potential homolog from a *Bacillus*. Identical residues are shaded. Ecoli = *E. coli* ligase, Bstea = *B. stearothermophilus* ORF.

B. A potential nucleotide binding motif conserved between the *E. coli* RNA ligase, a potential homolog, and several GTP-binding proteins. Eclig = *E. coli* RNA ligase, Bslig = *B stearothermophilus* ORF, P05453 = *S. cerevisiae* omnipotent suppressor (G1 to S transition protein), P32769 = *S. cerevisiae* EF1-α-like protein, P26751 = *Pyrococcus woesii* EF1-α, P31582 = *Arabidopsis thaliana* RHA1 (Ras-related protein)

A.

```
Ecoli  MSEPQRLFFA  IDLPAEIREQ  IIHWRATHFP  PEAGRP.VAA  DNLHL
Bstea  .MKRSHYFIA  VPLTSEAKQA  ISRFSGDASS  SLPFRTWVHE  EDYHI

Ecoli  TLAFLGEVSA  EKEKALSLLA  GRIRQPGFTL  TLDDA.....  GQWLR
Bstea  TLAFLGDVPP  GKMAPLCEAM  AAVAAKSAPF  SLALAGLGTF  GERTA

Ecoli  SRVVWLGMRQ  PPRGLIQLAN  MLRSQAARSG  CFQSNRPFHP  HITLL
Bstea  PRIFNQGVKE  EA.ALNELRR  DVYEACLSLG  FSLDRRPFAP  HITIA

Ecoli  RDASEAVTIP  PPGFNWSYAV  TEFTLYASSF  ARGRTRYTPL  KRWAL
Bstea  RKWQGEAPFQ  PEAL.RSLPA  ASTVFSVPEI  VLYRTNMEKT  PKYET

Ecoli  TQ........  ..........  ..........  ..........  .....
Bstea  IAAFPLLGAP  DGRTGEGMGQ  LLKLRDYISR  YETDVYHYVP  EFIRL

Ecoli  ..........  ..........  ..........  ..........  .....
Bstea  KQWQWEQAKA  RWEAERDADG  ARREPGETWD  FLLDKPSWWE  RLIGR

Ecoli  ..........  ..........  ..........  ..........  .....
Bstea  WRRGPEPEMD  EERSPAPSLS  RAATLDELKW  QFLDDLFELQ  GMPAG

Ecoli  ..
Bstea  RL
```

B.

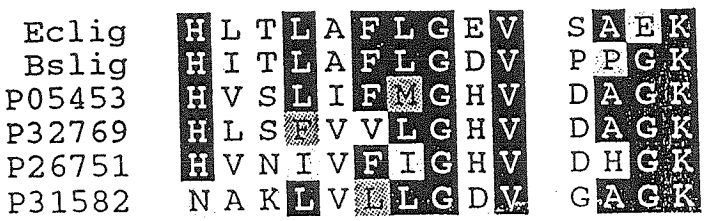


Figure III-5

conserved between these proteins (*E. coli* ligase residues 43-56 and 118-131) were used as query sequences for database searches, and were found to show some similarity to motifs conserved in GTP-binding proteins and succinate dehydrogenases, respectively as shown on Figure III-5B. The positive identification of functionally important and conserved residues by sequence comparison will therefore await the sequencing of homologous genes from additional organisms.

Genomic Disruption of the Ligase Gene:

In order to confirm that the protein which was sequenced from a highly purified ligase preparation was in fact the ligase enzyme and not a contaminant migrating identically during SDS-PAGE, a disruption of the putative ligase gene was undertaken. This would also allow detection of any other RNA ligase activities in *E. coli,* and a test of whether RNA ligase activity is required for cell viability. Genomic disruption was performed according to the method of Kushner as shown schematically in Figure III-6A, utilizing homologous recombination and resolution of a temperature sensitive (ts) plasmid bearing the gene to be replaced (39). First, the ligase ORF and 200bp of upstream flanking sequence were amplified from genomic DNA using unique oligo primers and cloned. When sequencing showed the cloned segment to be error-free, a kanamycin resistance cassette was inserted into a unique restriction site 45 bp into the ligase ORF. This site was chosen to interrupt the ligase gene as it should preclude expression of a functional truncated version of the gene without interfering with the downstream portion of the ORF which overlaps the promoter of the *sfs1* gene immediately downstream. The interrupted gene was then subcloned into a plasmid containing a chloramphenicol resistance marker and a ts replicon. After the cells were subjected to successive growth at 43°C and 30°C, as detailed in Materials and Methods, cells were plated and individual colonies tested for antibiotic resistance at various temperatures.

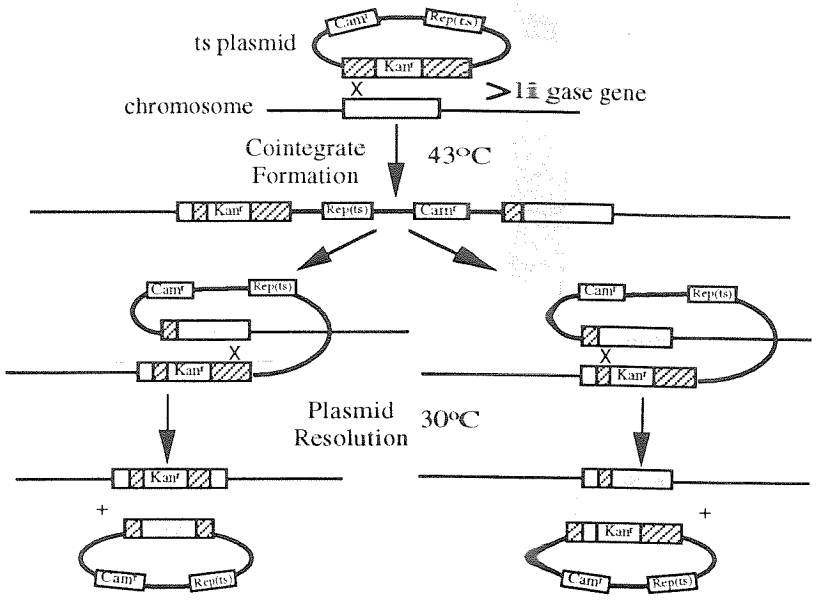
Figure III-6: Genomic Knockout of the RNA ligase Gene.

A. Schematic representation of the method utilized for disruption of the genomic RNA ligase gene using a temperature sensitive plasmid.

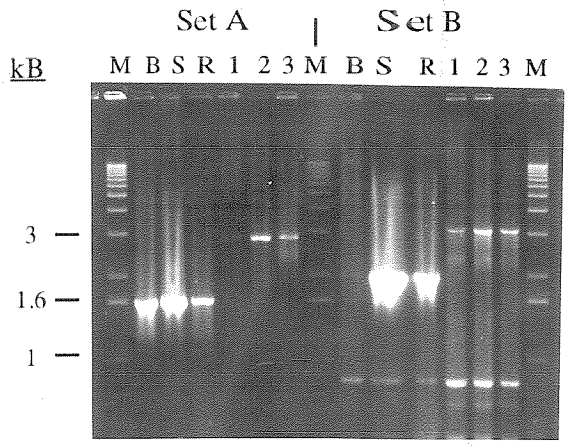
B. Disruption confirmed by PCR amplification of ligase gene from various genomic DNA preparations. Products separated by electrophoresis through a 1.5% agarose gel and stained with ethidium bromide. M = molecular weight standards, B = HB101, S = HS947 (knockout parent strain), R = non-disrupted resolved cointegrate, 1-3 = disrupted isolates. Set A = primers located 1kb upstream of the ligase ORF and at the 3' end of the ligase ORF, Set B = primers 1.5 kb downstream of the ORF and at the 5' end of the ligase ORF.

C. RNA ligation assay of extracts from various strains. m = mock reaction, L = partially purified *E. coli* RNA ligase 1,2 = disrupted, resolved isolates containing ts plasmid, 3-7 = disrupted, resolved isolates cured of plasmid.

A.



B.



C.

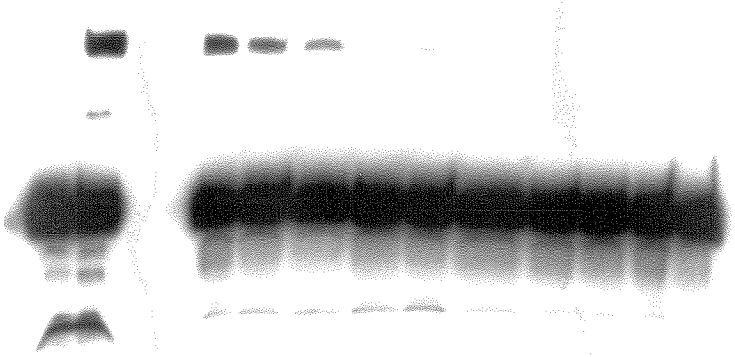


Figure III-6

Of 64 different colonies screened, one was found to be kanamycin resistant and chloramphenicol sensitive at 43°C. This was taken to be a disruption of the chromosomal RNA ligase gene by the kanamycin resistance cassette. The insertion of this cassette into the chromosomal copy of the ligase gene was confirmed by PCR amplification of the ligase gene and flanks from the genomic DNA of individual isolates as shown in Figure III-6B. When isolates of this knockout strain were assayed for RNA ligase activity, none was detected, as shown in Figure III-6C.

Discussion

Purification of *E. coli* RNA Ligase:

The purification of the *E. coli* 2'-5' RNA ligase over 1000-fold by the chromatographic methods utilized provided highly purified fractions but not a single homogeneous protein. In light of the behavior of the ligase during gel filtration and the shifting of peak concentrations of individual polypeptides during affinity chromatography, this is probably not because the ligase is part of a tightly associated macromolecular complex. More likely, the contaminating bands seen at the final stages of purification represent proteins with very similar biophysical properties. This is not completely unexpected given that many of the purification steps exploited affinity for nucleic acids. If it was necessary to improve the purification, it might be possible to add steps providing separation based on other properties such as hydrophobicity or metal ion affinity. It may also be possible to specifically elute the ligase from heparin or tRNA sepharose by exposure of bound proteins to tRNA half-molecules or cyclic nucleotides. Regardless, the procedure utilized was adequate to narrow the possible ligase candidates down to a small number which could be separated by SDS-PAGE for individual analysis. Some biological information was also obtained from the chromatographic purification, including the rarity of the ligase protein in *E. coli* even after alcohol induction, with an estimated number of y molecules per cell. The high activity in these fractions

demonstrates that the ligase protein is truly catalytic with turnover of $tRNA^{Tyr}$ half-molecules. The tight binding of *E. coli* RNA ligase to immobilized tRNA from two species provides additional evidence that the ligase recognizes a tRNA-derived substrate *in vivo*.

Reconstitution after SDS-PAGE:

The reconstitution of ligase activity after complete denaturation and separation by SDS-PAGE shows the enzyme to be quite stable and able to achieve an active conformation using only information contained in the primary sequence. The small size and excellent stability of this protein make it an excellent candidate for a future crystallographic structure determination, if a suitable overexpression system and complete purification protocol can be determined. The fact that activity of the renatured ligase protein was not affected in any manner by the addition of other eluted bands gives further evidence that this single polypeptide is responsible for all the RNA ligase activity observed in *E. coli* extracts.

Ligase Gene Sequence:

The ORF which was found to match the derived sequence of the candidate ligase protein identified by reconstitution after SDS-PAGE was predicted to encode a polypeptide with properties in excellent agreement with those observed for the ligase protein throughout purification. The sequence itself was not excessively informative however, as it does not appear to be a member of any known sequence families or to contain common structural or functional sequence motifs. The discovery of a highly similar protein in the sequence databases did not provide much insight as the function of this predicted *Bacillus* protein has not been investigated.

The conservation of the RNA ligase protein between such distantly related bacterial species as *E. coli* and *B. stearothermophilus* suggests that this enzyme exists in

diverse bacterial taxa. This protein would appear to have been lost from some evolutionary branches between the Low G+C Gram positives and the Gamma division Proteobacteria however, since the activity was not detected in *Desulfovibrio*, *Paracoccus*, or *Rhodopseudomonas* Proteobacterial species although this may have been due to some artifact of the extract preparation and assay (36). No obvious homolog can be found in the completed sequence of the *Haemophilus influenzae* genome, which although closely related to *E. coli* by rRNA sequence comparisons, has an extremely streamlined genome less than half the size of the chromosomes of other Proteobacteria. *Haemophilus* may have discarded the ligase activity or transferred this function to another polypeptide due to the selection pressures which caused the extreme downsizing of its genome. This would imply (as did the lack of detectable activity in several bacterial species) that the RNA ligase does not perform a function absolutely necessary for bacterial survival, but may be conditionally required under growth conditions encountered by a wide variety of other species.

Genomic Disruption:

Interruption of the genomic locus encoding the putative ligase protein was sufficient to confirm that the correct polypeptide had been sequenced, and to reveal that this was the only enzyme in *E. coli* capable of ligating yeast tRNA half-molecules. The fact that genomic ligase knockout isolates were viable at 37°C and at 43°C confirms that RNA ligase is not absolutely required for survival under laboratory growth conditions. Although the disrupted strains do not display a lethal phenotype, they can be examined for more subtle effects on growth and RNA metabolism. The availability of viable knockouts will also provide a useful null background for the expression of tagged or mutagenized ligase protein for biochemical experiments.

Materials and Methods

Large-scale cell growth: Cells were fermented by the Caltech Fermentation Facility in rich media containing 16g/L Bacto-tryptone, 6 g/L yeast extract, 50mM phosphate buffer pH 6.8, 4 g/L $(NH_4)_2SO_4$, 0.5 g/L $MgSo_4$ anhydrous, and 20 ml/L glycerol with thorough aeration and constant feeding of 2x media for 24 hrs. 5% methanol was added 2 hrs before harvest, and harvested cell paste was stored at -70°C.

Ligase Purification:

200g-1kg of HB101 cells frozen at -70°C were thawed in extraction buffer (EB) containing 10% glycerol, 40mM HEPES pH 7.8, 2mM EDTA, 125mM KCl, 1mM pefabloc, 1mM PMSF, and 1mM benzamidine, and evenly dispersed using a hand blender. This cell suspension was subjected on ice to sonication with a Branson Sonifier using the large tip at 80% power for 10-15 one-minute periods with cooling in-between. The crude extract formed was subjected to ultracentrifugation at 35,000 rpm in a Beckman 45Ti rotor at 3°C for one hour. The supernatant was mixed batchwise with 500ml of DEAE sepharose CL-6B (equilibrated in EB) and allowed to bind for at least 30 minutes on ice. All subsequent steps were performed at 4°C in the cold room. The DEAE slurry was poured into a column and the effluent loaded directly onto a 200ml bed of cellulose phosphate (Whatman) equilibrated in EB at approximately 1 ml/min.

The phosphocellulose column was washed extensively with 500mM KCl and ligase activity eluted with a linear KCl gradient in EB from 0.5 to 1.5M. Peak ligase activity fractions were pooled, diluted to 150mM KCl, and loaded onto a 25 ml heparin hyper-D (Biosepra) column. The ligase was eluted from heparin with a 150mM to 700mM KCl gradient in EB, and peak activity fractions pooled. The heparin affinity pool was diluted to 150mM KCl in EB and loaded onto a 15 ml column of *E. coli* tRNA linked to a sepharose support. tRNA sepharose substituted to 3 mg RNA per ml was prepared

as described by Rauhut, but substituting E. coli tRNA (Sigma) for yeast tRNA (84). Ligase was removed from this column using a gradient from 150mM to 1M KCl in EB, followed by a 2M KCl wash. The ligase eluted in a broad peak beginning around 400mM KCl, but continuing into the 2M wash.

Active fractions were pooled, diluted, concentrated on a small 2 ml column of heparin hyper-D, eluted in , and adjusted to 600mM KCl. The concentrated activity pool was then passed through a bed of superdex 75 gel filtration medium (Pharmacia) at a flow rate of 0.04ml/minute. Peak active fractions were diluted to 150mM KCl in EB and loaded onto a column of *S. cerevisiae* tRNA linked to sepharose. Ligase activity was eluted with a 150mM - 800mM KCl gradient.

SDS-PAGE elution and renaturation: Electrophoresis for preparative elution, as well as all analytical protein electrophoresis, was performed on 14-16% acrylamide/piperazine diacrylamide (2.67% crosslinking) gels under the modified discontinuous SDS-PAGE conditions of Schagger and von Janow (92). Bands were visualized for excision by negative staining with 0.3M $CuCl_2$ for 15 min at room temperature, and destained after cutting bay successive washes in 0.25M Tris-HCl pH 8.8 plus 0.25M EDTA, and in distilled water. Excised bands were crushed, and eluted overnight at 4°C in 50mM Tris-HCl pH 7.5, 150mM NaCl, 0.1% SDS, and 0.1 mg/ml acetylated BSA (Gibco). Protein was precipitated by the addition of 4 volumes cold acetone, with 30min incubation at 0°C, and the pellet recovered by centrifugation at 15,000rpm, 4°C, for 20min. in an SS-34 rotor. Protein pellets were resuspended in EB containing 6M guanidine HCl and microdialyzed versus EB plus 125mM KCl at 4°C. Silver staining of SDS-PAGE gels was performed according to the method of Wray (122).

PCR conditions: PCR was performed in 1x PCR Buffer with Mg^{2+} (Boehringer Mannheim), 0.4mM dATP,dCTP, dGTP, and dTTP, 10μM each primer, and 0.7 U/μl

Taq DNA polymerase. For reactions using degenerate primers, 400μM total primers was used. For labeled probe production, 300μM dCTP, dGTP and dTTP was used in the presence of varying amounts of α-^{32}P-dATP.

Southern Hybridizations: Labeled PCR product was heated to 95°C in 6xSSC, 10mM phosphate buffer pH 6.8, 1mM EDTA, 0.5% SDS, 100μg/ml sonicated calf thymus DNA, and 0.1% dry milk, then cooled on ice. The probe was hybridized to blots in the same buffer at 37°C, O/N. The blot was washed with 6xSSC, 0.1% SDS and with 5xSSC at room temperature, then exposed to film for autoradiography.

Knockout: Gene disruption was performed according to the method of Kushner (39).

The ts plasmid bearing an interrupted ligase gene was transformed into a RecA$^+$ strain of E. coli and plated at 43°C on chloramphenicol to select for cointegration of the plasmid into the chromosomal ligase gene by homologous recombination. Plasmid cointegrants were resolved from the chromosome by growth for several generations in liquid culture at 30°C, at which temperature the plasmid replicon is active and interferes with chromosomal replication. These cultures were then transduced with a P1 phage containing the recA locus to disrupt the chromosomal recA gene and thereby prevent any further homologous recombination events. P1 transducants were screened for resolution of the ts plasmid by streaking separately on chloramphenicol, kanamycin, and nonselective LB plates at 43°C. Cells which had resolved the plasmid from their chromosome would now be sensitive to chloramphenicol at the nonpermissive temperature. If the interrupted copy of the ligase gene had been left in the chromosome after resolution, however, kanamycin resistance would be retained. If the interrupted gene was essential for viability, then no cells would survive at 43°C, even on nonselective media.

Chapter IV

Investigations of *in vivo* Ligase Function

Introduction:

The successful purification of *E. coli* 2'-5' RNA ligase and cloning of the gene encoding it have provided a starting point for research into the function of this enzyme in bacterial metabolism. Knowledge of the DNA sequence of the ligase gene and surrounding chromosomal region allows the creation of powerful reagents for genetic, biochemical, and molecular biological studies. The use of genomic knockouts of RNA ligase, overexpressed ligase protein in both native and affinity-tagged forms, and specific polyclonal antibodies directed against ligase should yield clues about this enzyme's function, and lay a foundation for further research into this problem.

Results:

Ligase Knockout Growth:

Effects of the disruption of RNA ligase expression in *E. coli* on overall fitness were examined by assaying bacterial growth under a variety of conditions. Ligase knockouts were viable at temperatures ranging from at least 23-43°C and were able to grow in the presence of 5% methanol. The effect on growth of restoring RNA ligase activity to knockout strains was also tested. Ligase knockout and wild type *E. coli* were transformed with a clone of the ligase ORF in pBS which constitutively expressed ligase activity at about 10 times wild type levels. Both of these overproducing strains (but not those transformed with vector alone) were temperature sensitive, being viable at 37°C but unable to grow at 43°C, as shown in Figure IV-1A. Thus the overproduction of *E. coli* RNA ligase has a toxic effect at elevated temperatures.

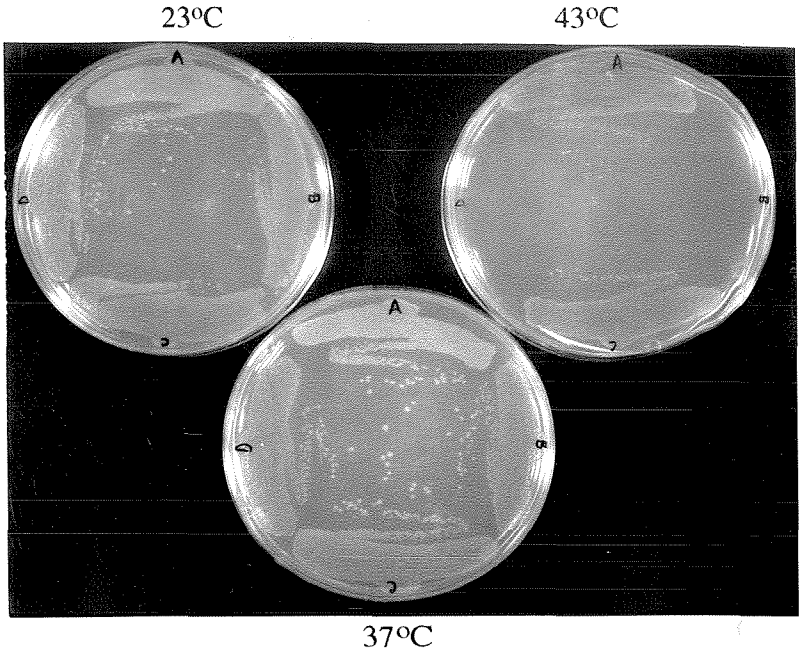
In order to further investigate the effects of ligase knockout and overexpression on growth, the growth curve of these strains was examined. Liquid cultures were

Figure IV-1: Growth Phenotypes of RNA Ligase Knockout and Overexpressing Strains.

A. Growth of strains at indicated temperatures. A = HS947/RecA- (knockout parent strain), B= HS947/RecA- + pBS-lig, C = KO, D = KO + pBS-lig.

B. Time course of cell growth as monitored by viable plating on selective media. Four replicate cultures of each strain were assayed, and one standard deviation at each data point is represented by error bars. KO = strain with a genomic ligase disruption, KO+pBS-lig = genomic disruption complemented by the RNA ligase ORF cloned into pBS.

A.



B.

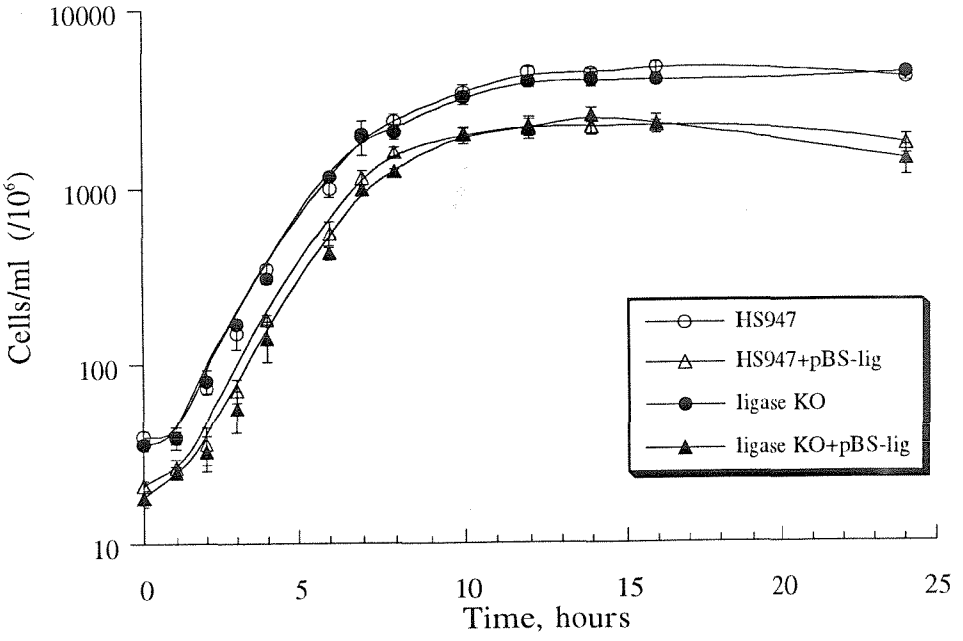


Figure IV-1

inoculated to approximately equal densities (as judged by optical density) from stationary phase cultures, and cell growth at 37°C monitored by dilution and viable plating at various times afterwards. The results are shown in figure IV-1B. Ligase knockouts grew at approximately the same rate as wild-type *E. coli* throughout the entire growth curve. Both ligase knockout and wild-type strains which were overexpressing RNA ligase, however, showed somewhat altered growth characteristics as shown in Figure IV-1B. These strains displayed a slightly extended lag phase and slower growth characteristics through the exponential phase of growth. Overproducing strains showed a carrying capacity of about half the cell densities of wild-type or knockout cultures during the stationary phase of growth. These effects on growth show that while the *E. coli* RNA ligase is not absolutely required for cell viability, it appears that proper expression levels of this enzyme are necessary for the regulation of growth. Since the ligase is being overexpressed at a level only about ten times that of wild-type, it is likely that the growth defects are specific to the ligase and not due to a major portion of the cellular energy and resources being directed into production of this protein.

In order to examine more specific effects of alterations in ligase activity *in vivo*, total RNA was isolated from wild type and knockout strains as well as knockouts overexpressing ligase grown at permissive or nonpermissive temperatures. Gross characteristics of these RNA populations were observed by denaturing gel electrophoresis and methylene blue staining of equivalent amounts of RNA as shown in Figure IV-2. Ligase$^-$ cells show a slight accumulation of tRNA-sized molecules, but otherwise closely resemble wild-type. Overexpressing cells show a depletion of tRNA-sized molecules and accumulation of a range of RNA fragments in the 5-20N size range. An overexpressing strain incubated at 43°C shows a less severe depletion of tRNAs but a more pronounced accumulation of low molecular weight fragments.

Figure IV-2: Total RNA Populations of Ligase Knockout and Overexpressing Strains. 50μg each RNA separated on 10% acrylamide and stained with methylene blue. HS = HS947 (knockout parent strain), KO = genomic ligase knockout, KO + pBS-ligase overexpressing clone grown at 37°C, or incubated at 43°C for 2hrs prior to harvest.

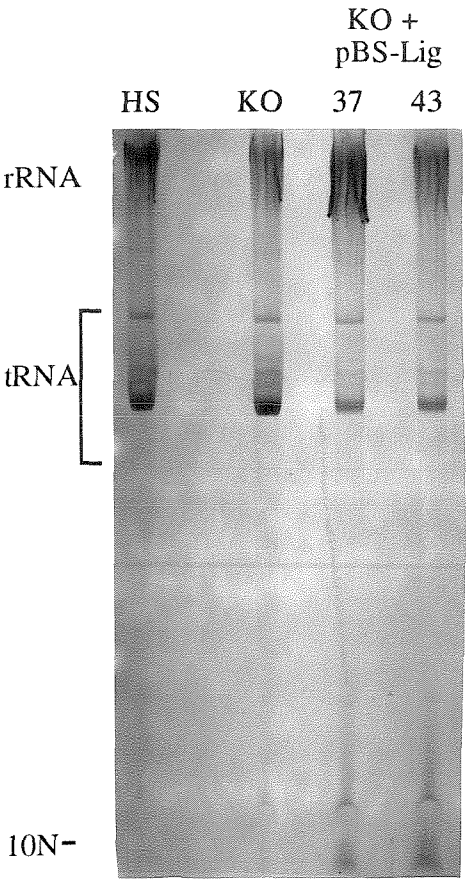


Figure IV-2

Examination of Enzyme Equilibrium:

The highly purified final fractions of large scale purifications of *E. coli* 2'-5' RNA ligase were utilized for further investigation of the biochemical properties of the enzyme for information on the reaction mechanism which might also provide clues to *in vivo* function. Toward this end, the equilibrium of the ligation reaction *in vitro* was examined. $tRNA^{Tyr}$ which had been produced by ligation of half-molecules by the *E. coli* ligase, and therefore should contain a 2'-5' linkage in the anticodon loop, was gel purified and reannealed by heating and slow cooling. $tRNA^{Tyr}$ produced by ligation using T4 RNA ligase and PNK was purified analogously as a control. Figure IV-2A shows that 2'-5' linked substrates were specifically cleaved in the anticodon loop by purified ligase fractions with the same kinetics as ligation of half-molecules by those fractions. Beginning with either pure half-molecules or ligated product, the final molar ratio was approximately 4 or 5:1, halves:full length. tRNA ligated by the T4 enzymes, which should contain 3'-5' linkages exclusively, was not cleaved. This demonstrates the specific reversibility which would be expected of this reaction if it occurs via the proposed ribonuclease mechanism. The direction of the equilibrium *in vivo*, and therefore the function of this enzyme for joining or cleavage, will have to be resolved by identification of the natural substrate(s) of this enzyme.

Ligase Overproduction:

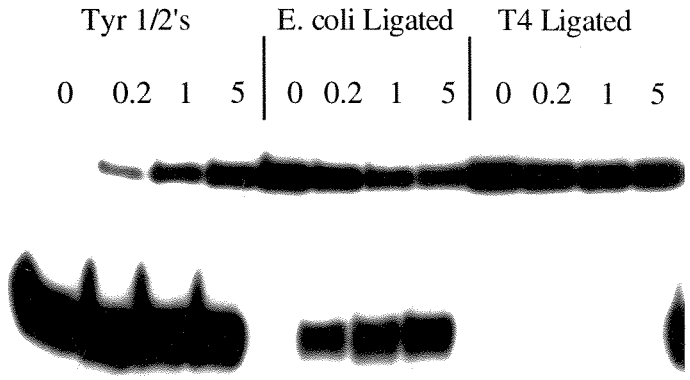
Several strategies were undertaken to overexpress and purify *E. coli* RNA ligase in the amounts necessary for use as a reagent in biochemical experiments. The first expression system utilized was the result of a fortuitous discovery. When the PCR-amplified RNA ligase open reading frame alone was cloned into pBluescript (pBS) for sequencing, it was found to cause a 10-fold increase in ligase activity in host strains. This clone lacked any endogenous upstream sequence, and was extended at the 5' end only by an EcoRI site added for cloning. This level of expression was achieved from this

Figure IV-3: Equilibrium of the *E. coli* RNA Ligase .

A. Time course of forward and reverse reactions of highly purified *E. coli* RNA Ligase. E. coli Ligated= tRNATyr produced by ligation of Tyr 1/2's with *E. coli* RNA ligase, T4 Ligated= tRNA produced by ligation with T4 RNA ligase and PNK.

B. Quantification of the reactions in A, graphed according to the molar ratio of tRNA to half-molecules at each time point.

A.



B.

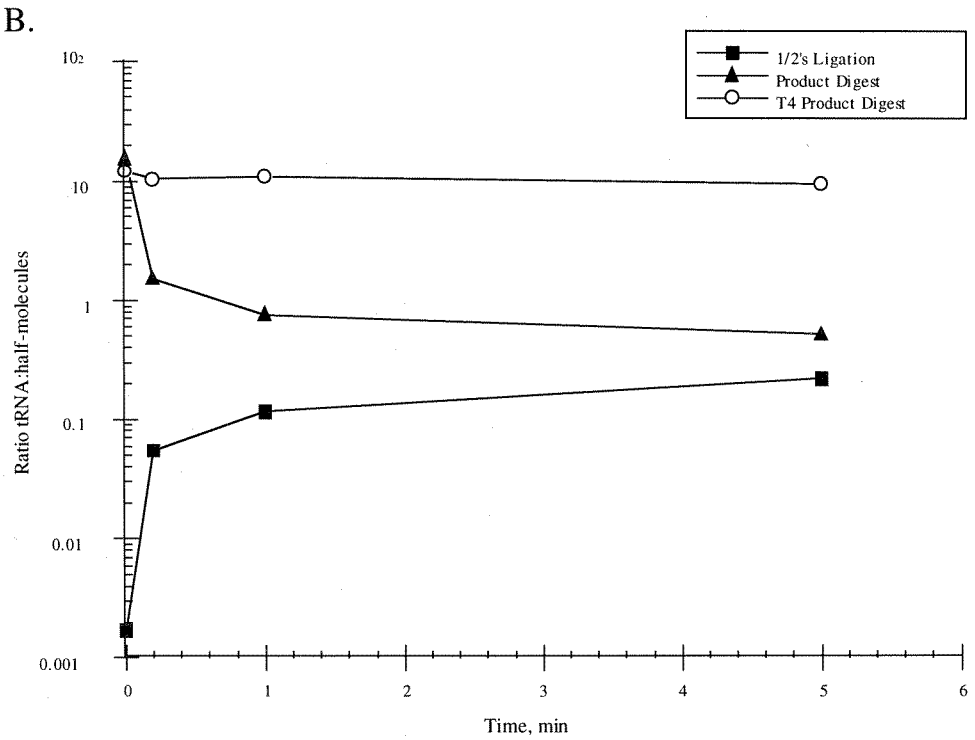


Figure IV-3

clone in both wild-type and ligase$^-$ strains. The ligase gene in this construct was interrupting the β-galactosidase coding region, and oriented in the same direction. It was therefore considered that the expression of some highly active ligase might be due to the low level of transcription allowed in the absence of an inducer by the *tac* promoter controlling β–galactosidase expression in this vector. When this promoter was induced by the addition of 1mM IPTG to culture media however, only a slight increase (up to 15%) or a net drop in ligase activity was seen in various experiments. The inclusion of approximately 200bp of upstream genomic sequence ahead of the ORF in these constructs did not significantly alter expression patterns.

pET vectors:

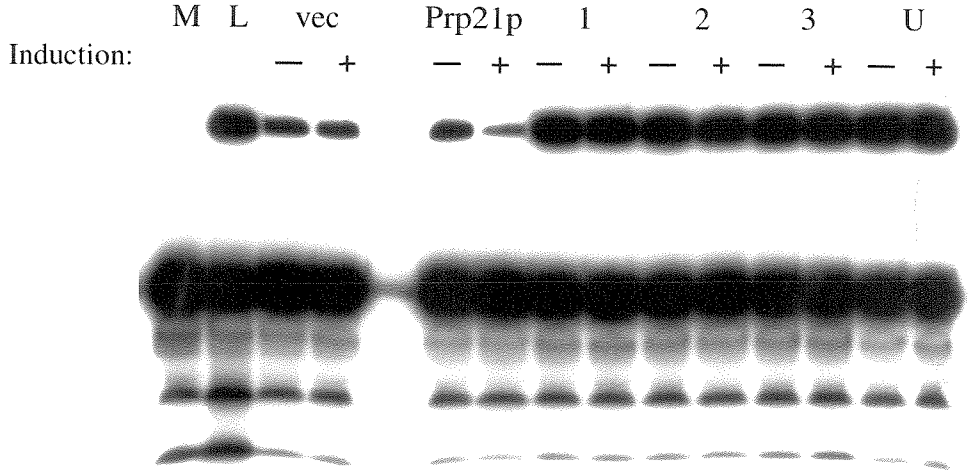
In an effort to express the ligase at higher levels and in a more controlled fashion, the ORF alone or with the 200bp upstream region was subcloned into two versions of the pET vector (to allow expression of native or His_6 affinity-tagged protein) under the control of a phage T7 promoter. When expression of ligase from these constructs was monitored, as shown in Figure IV-4A, the results were the same as for the pBS vector. At least a 10 fold increase in RNA ligase activity was seen in strains carrying these vectors, even when such strains also expressed T4 lysozyme to block transcription from the T7 promoter. The addition of 5% methanol to culture media induced a slight increase in ligase activity, but never as much as the increase of 50% or more typically observed for expression from the endogenous gene. When the expression of T7 polymerase in these strains was induced with IPTG, activity was not significantly increased. The slight induction of activity by exposure to methanol or IPTG was not additive in the presence of both. Inducible expression of a control protein, *S. cerevisiae* PRP21p, in the same vector and host strain indicated that the T7 system was functional, as seen in Figure IV-4B. The level of overexpression of *E. coli* RNA ligase in this system, remained constant in the presence of the ORF on a multicopy plasmid as monitored by SDS-PAGE. Some

Figure IV-4: Overexpression of Cloned *E. coli* RNA Ligase in *E. coli.*

A. Overexpression from pET 11 vector. Ligase activity in extracts of wild-type *E. coli* harboring various clones. vec = pET11 alone, Prp21p = *S. cerevisiae* PRP21 gene cloned into pET11 as a control, 1,2,3 = isolates of *E. coli* RNA ligase ORF in pET11, U= ligase ORF + 200bp upstream sequence. Induction= culture growth in the presence of 1mM IPTG.

B. Silver stained SDS-PAGE gel of cultures assayed in part A. s indicates soluble extract, all other lanes are whole cells. MW = molecular weight standards.

A.



B.

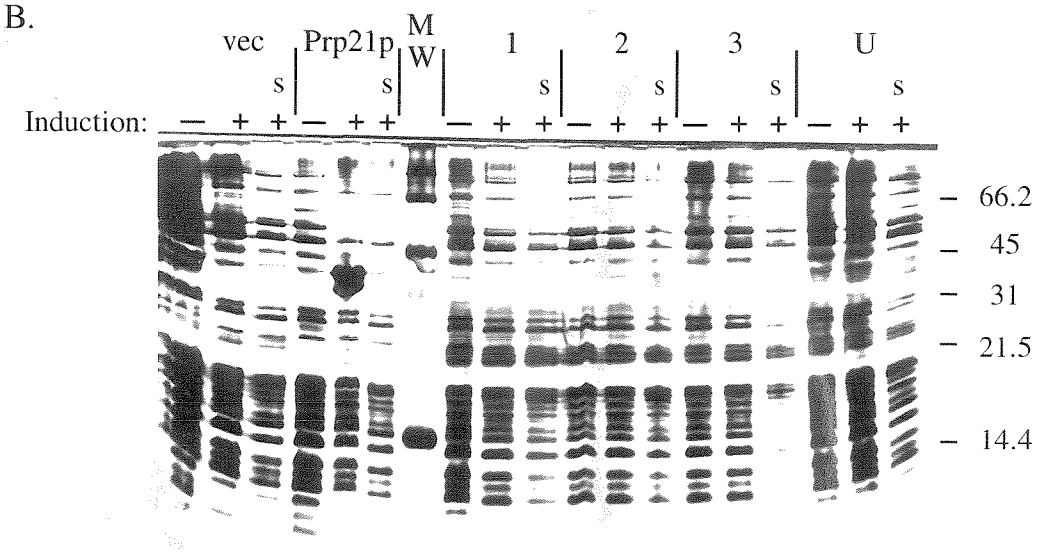


Figure IV-4

recombinant protein could be purified from strains expressing ligase with a His_6 metal-affinity peptide tag fused to its amino terminus. The small amount of protein purified from crude extracts on nickel-NTA-agarose was of the predicted molecular weight for tagged ligase but did not show appreciable RNA ligase activity (not shown). These results seemed to indicate that expression of the recombinant ligase was being driven by a constitutive promoter element within the cloned ORF which interfered with activation by an upstream inducible promoter.

pATH vector:

As an additional test of the possibility of an internal promoter, the ligase ORF was subcloned into a pATH vector for expression of the enzyme as a TrpE fusion protein under the control of the inducible *trp* promoter. Such fusions are usually insoluble and accumulate in inclusion bodies *in vivo*. When ligase-TrpE constructs were transformed into host strains grown in rich media supplemented with tryptophan to repress *trp* promoter expression, an increase in ligase activity of approximately 10-fold was observed (not shown). Cells containing these constructs were washed, transferred to minimal media lacking tryptophan, and challenged with indole acrylic acid to maximally induce expression from the *trp* promoter. Once again, upon induction of expression from the upstream promoter, total ligase activity was the same or somewhat lowered. The expression of some ligase-TrpE fusion protein could be detected in crude extracts upon induction, but the amount was far less than that observed for a control construct expressing the TrpE protein alone (not shown).

Taken together, the data on ligase expression in these various inducible systems would appear to indicate the presence of a constitutive promoter of transcription somewhere within the ligase coding region. Although the unusual location and behavior of the ligase promoter is interesting and worthy of study in its own right, the implications for the immediate goal of controllable, high-level expression in *E. coli* were not

promising. Since it appeared that the ligase could not be highly overexpressed as a significant portion of total cellular protein in *E. coli,* expression in a heterologous organism was attempted.

E. coli RNA ligase Expression in *S. cerevisiae*::

The yeast *Saccharomyces cerevisiae* was chosen as a host for heterologous expression. The extensive information on protein expression and extract preparation in *S. cerevisiae* at hand, as well as the availability of useful strains and vectors made this an attractive organism for this purpose. Although *S. cerevisiae* does contain an endogenous RNA ligase activity in the form of the tRNA splicing ligase, this enzyme is found exclusively in the nucleus and should be easily separable from a cytoplasmically expressed heterologous protein. The *E. coli* RNA ligase ORF was subcloned into the expression vector YpKRAP (for Konvertible Rapid Affinity Purification) and this construct was transformed into *S. cerevisiae* strain YPH274. Expression of the ligase in this vector was directed by the CUP1 promoter of the *S. cerevisiae* metallothionein gene which allows a low constitutive level of expression that can be quantitatively induced to high levels in the presence of copper ions (42). Cytoplasmic extracts of YPH274 harboring vector alone show negligible ligase activity, while those containing the *E. coli* RNA ligase ORF in YpKRAP have gained considerable ligase activity, which could be further induced by the addition of copper to liquid culture media.

Optimization of Copper-induced Expression:

The major advantages of protein expression from the yeast CUP1 promoter are the quantitative response to extracellular copper concentration and the fact that CUP1 expression is not coupled to the physiological state of the cell. Therefore, high levels of expression may be obtained throughout all phases of diauxic growth. In order to determine culture conditions for maximal production of RNA ligase protein, induction of

ligase activity and cell density at various concentrations of extracellular copper were monitored over a long time course of growth in minimal media. The results of this experiment are shown in Figure IV-5. Initial induction of ligase activity was roughly proportional to the amount of copper sulfate added to media, but induction by 2mM or higher extracellular copper proved detrimental to cell growth. This growth defect may be due to the toxicity of high cytoplasmic levels of the *E. coli* RNA ligase, or to sensitivity of the host strain to high copper concentrations in minimal media. Regardless, it is clear that the largest total amounts of ligase could be produced by long term growth in the presence of 1.5 mM copper sulfate. The specific activity of ligase in these cells ranged up to 100 times that endogenously expressed in *E. coli*, and 10 times the maximum amount overexpressed in that organism. This induction protocol was utilized for preparative growth of ligase-expressing strains in flasks or a fermentor.

Purification of Heterologously Expressed Ligase:

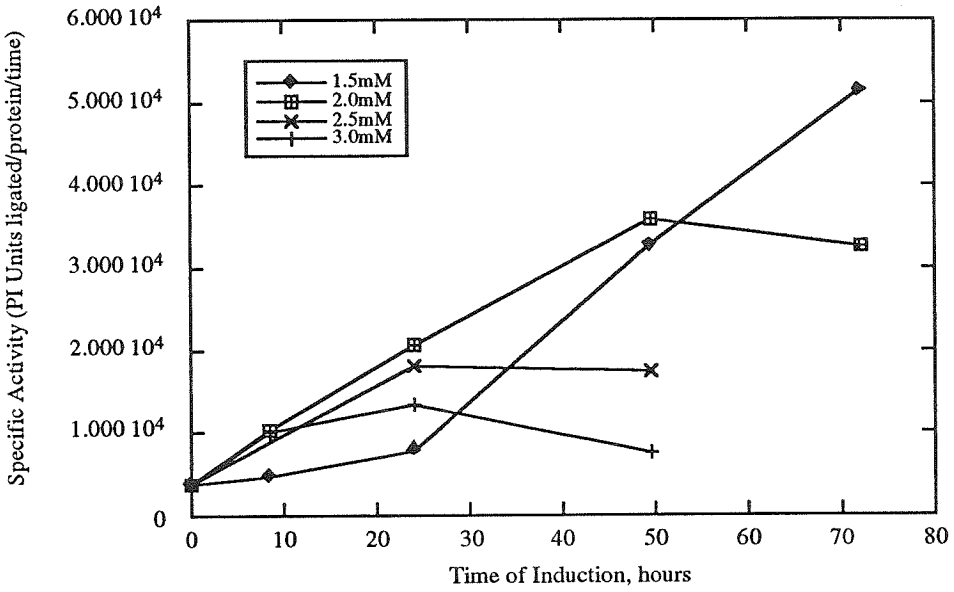
In order to avoid the development of a new and possibly laborious protocol for the purification of *E. coli* RNA ligase from yeast extracts, the ligase was expressed as a fusion with a short, bifunctional peptide tag for use in quick, highly selective protein purification on specific affinity matrices. A bifunctional affinity tag which can be fused to either terminus of a recombinant protein is a feature of the YpKRAP expression system utilized for ligase expression. This peptide tag contains both the His-X-His_8 motif which has the ability to bind immobilized metals, and the FLAG peptide which is recognized and tightly bound by a commercially available monoclonal antibody (Mab). Since ligase protein previously expressed in *E. coli* with an amino-terminal His_6 tag was found to be inactive, the ORF was cloned into YpKRAP so that the dual tag was fused to its carboxyl terminus. Tagged ligase was expressed in *S. cerevisiae* with a similar time course of induction as untagged enzyme, as shown in Figure IV-7A. Tagged ligase could be bound in yeast extracts by either nickel-NTA-agarose or immobilized M2 anti-FLAG Mab, as

Figure IV-5: Overexpression of *E. coli* RNA Ligase in Yeast.

A. Specific activity of *E. coli* RNA ligase in extracts of *S. cerevisiae* induced with indicated amounts of $CuSO_4$ in minimal growth media.

B. Growth of $CuSO_4$-induced strains as monitored by optical density.

A.



B.

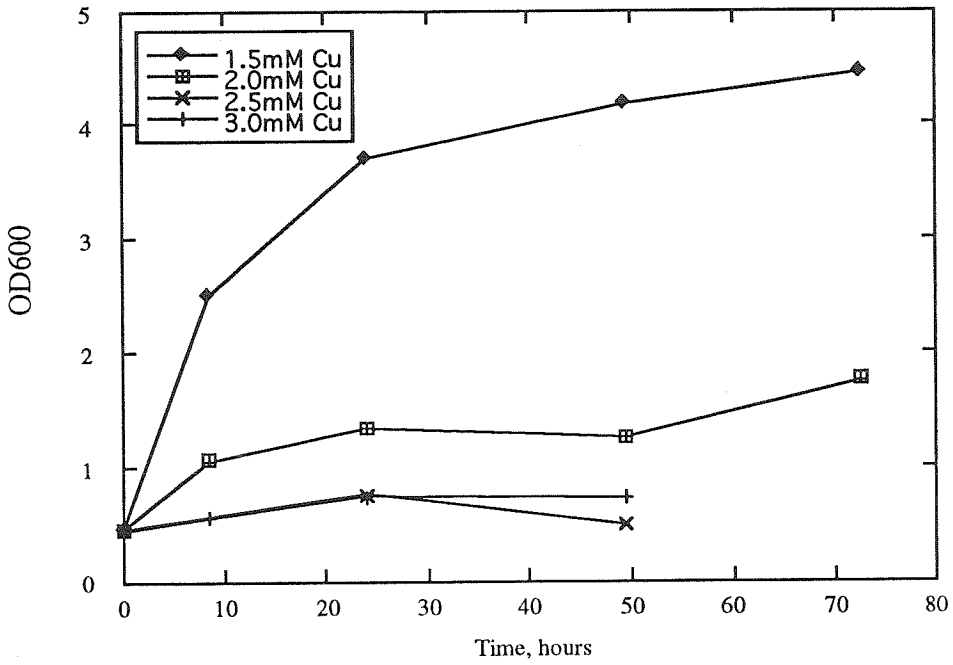


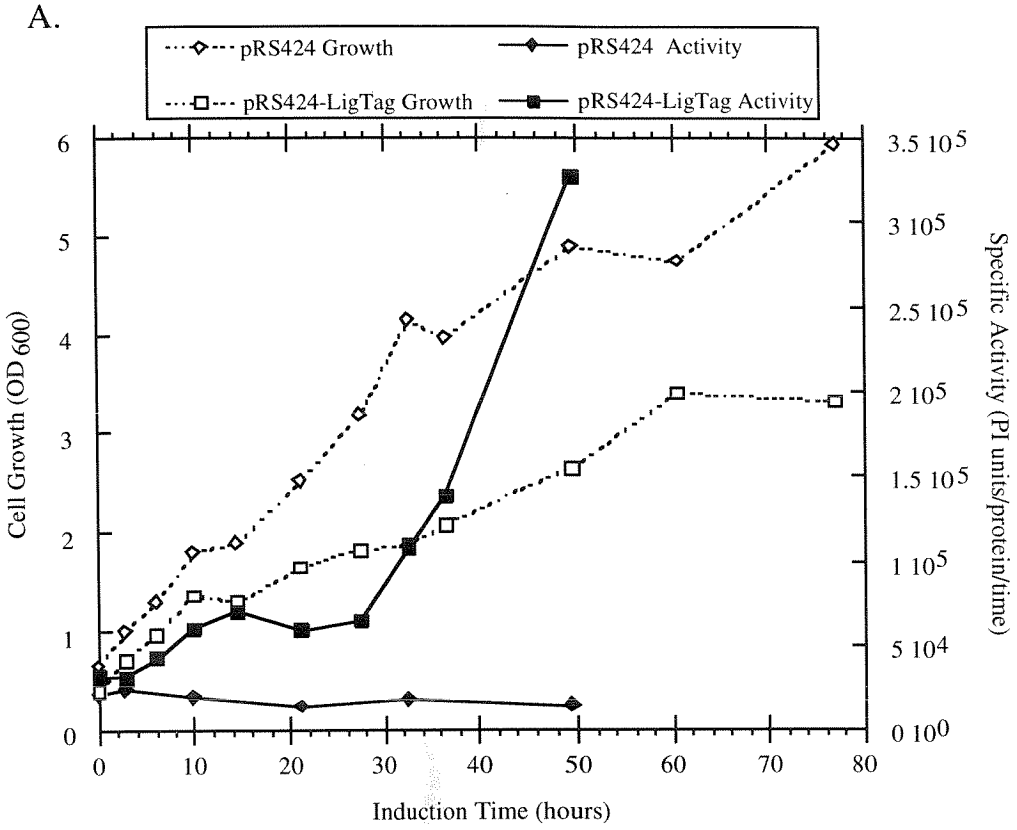
Figure IV-5

Figure IV-6: Overexpression of Tagged RNA Ligase in Yeast.

A. Growth and RNA ligase specific activity in vector only or tagged ligase clone-containing *S. cerevisiae* induced with 1.5mM $CuSO_4$.

B. Purification of tagged ligase from yeast extract using tag affinity resins. C = crude extract. Ni-NTA: F = flow through, W = wash, I = imidazole elution, E = EDTA elution. M2: F = flow through, W = wash, P = free peptide elution.

A.



B.

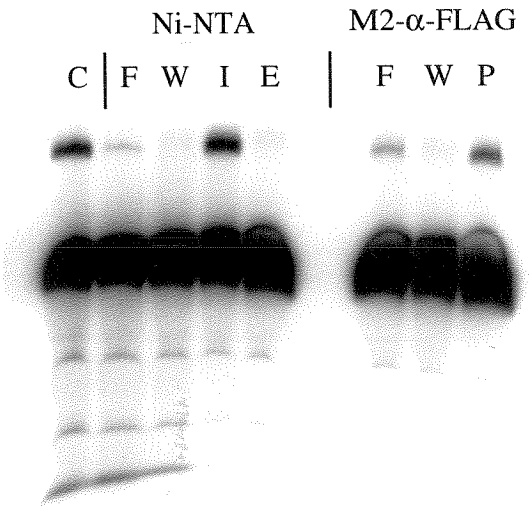


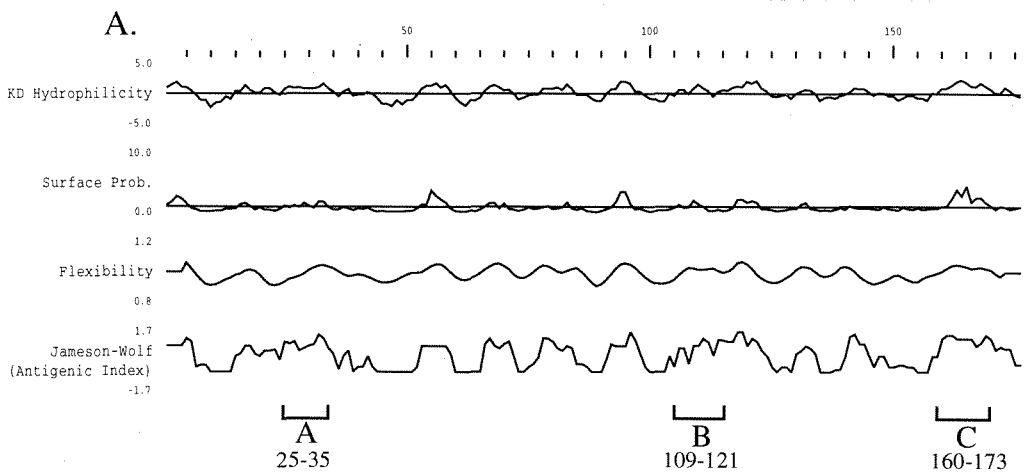
Figure IV-6

Figure IV-7: Anti-RNA Ligase Antibodies.

A. Antigenicity predictions for the RNA ligase amino acid sequence.

B. Western blot probing crude *E. coli* extract and partially purified *E. coli* RNA ligase with preimmune and immune sera.

C. Western probing extracts and ligase purification fractions with affinity-purified antibodies. Cr = crude extract, S-100 = S-100 supernatant, Dft = DEAE flow through, CP = cellulose phosphate, Hep = heparin hyper-D fractions.


A.
50
100
150
5.0
KD Hydrophilicity
-5.0
10.0
Surface Prob.
0.0
1.2
Flexibility
0.8
1.7
Jameson-Wolf
(Antigenic Index)
-1.7
A
25-35
B
109-121
C
160-173

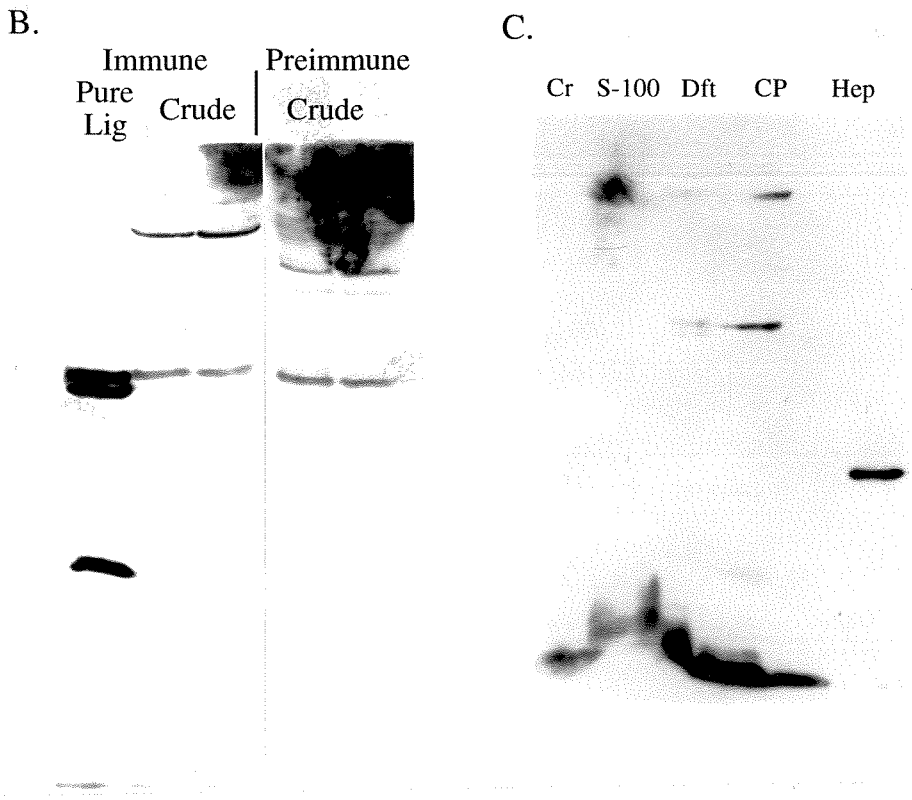

B.
Immune
Preimmune
Pure
Lig
Crude
Crude
C.
Cr
S-100
Dft
CP
Hep

Figure IV-7

demonstrated in Figure IV-7B. For purification of preparative amounts of tagged protein from yeast, partial purification on DEAE and phosphocellulose resins was employed previous to exposure to tag-specific affinity resins to circumvent problems caused by proteolytic enzymes in whole cell extracts. Effective concentrations of metal chelators could therefore be included in cell disruption buffers to inhibit metalloproteases. These chelators, which strip nickel ions from nickel-NTA-agarose, could then be removed from the buffer while the ligase was bound to one of the preliminary columns. These initial purification steps also serve to remove many other proteases which could destroy immobilized M2 anti-FLAG antibodies as well as RNA ligase protein itself.

To test the feasibility of large scale preparation of tagged *E. coli* ligase from *S. cerevisiae*, a 10L culture of YPH274 carrying the YpKRAP-ligase construct was grown in minimal media and induced by the addition of copper sulfate. Approximately 100g of cells were recovered from this culture and broken by agitation with glass beads. Tagged ligase was purified from this extract as detailed in Materials and Methods. Briefly, crude extract was cleared with an S-30 spin, and the supernatant was mixed with DEAE-sepharose. Ligase activity in the DEAE supernatant was further purified on cellulose phosphate and active fractions from this column were pooled, then mixed with nickel-NTA-agarose resin. After extensive salt washing, tagged protein was eluted with high concentrations of imidazole, and active elution fractions were mixed with M2-Mab resin. Tagged ligase was released from this resin in pure, but partially degraded, form by the addition of free FLAG peptide. Approximately 300μg of tagged, active RNA ligase was recovered.

Antibodies Recognizing *E. coli* RNA 2'-5' Ligase:

Another reagent which should prove invaluable for studies of ligase function is an avid, specific, polyclonal antibody directed against structural epitopes of the ligase protein. The Peptidestructure and Plotstructure programs of the GCG package were used

to predict potentially immunogenic regions within the *E. coli* RNA ligase amino acid sequence. Three regions, indicated in Figure IV-8A, were identified which were of sufficient hydrophilicity, surface probability, antigenic index and length to be chosen for use as antigens. Peptides corresponding to these three regions (residues 25-36, 109-122, and 160-175) of the ligase polypeptide were chemically synthesized and conjugated to KLH protein. The conjugated peptides were mixed in equimolar amounts and injected into two rabbits. After two booster injections, serum was collected from these animals regularly for several months and individual bleeds tested for the ability to recognize *E. coli* RNA ligase on a western blot. As can be seen in Figure IV-8B, antibodies recognizing RNA ligase in crude *E. coli* extract and ligase purification fractions were elicited, but immune serum also contained antibodies already present preimmunization which recognized a number of other *E. coli* protein epitopes.

In order to remove interfering antibodies and raise the titer of anti-ligase IgG, an affinity purification procedure was applied. Immune serum was applied to a column of pure tagged *E. coli* ligase (expressed in yeast) bound to nickel-NTA-agarose. Contaminating serum proteins were removed by high salt wash and specific antibodies released by the addition of high concentrations of magnesium. Purified antibody fractions were able to recognize the ligase protein in purified fractions, but still recognized several other bands in crude extracts (Figure IV-8C).

RNA Binding by Immobilized *E. coli* RNA Ligase:

An experiment was performed to examine the feasibility of performing RNA binding experiments using immobilized, purified RNA ligase to sequester specific RNA species from total RNA populations *in vitro*. In order to help define the experimental parameters for RNA binding by the *E. coli* ligase, a cursory kinetic analysis was performed. Highly purified fractions of native ligase were used to ligate $tRNA^{Tyr}$ half-molecules at brief time periods (30 seconds) over a range of substrate concentrations, and

Figure IV-8: *in vitro* RNA Binding to Immobilized *E. coli* RNA Ligase.
A. pCp labeled knockout strain total RNA eluted from immobilized enzyme with a KCl gradient. T = labeled total RNA. M = markers, labeled rRNA and tRNA splicing products (and degradation products), S = RNA binding supernatant.
B. Eluted RNA from a ligase overexpressing strain. Lanes labeled as in A.

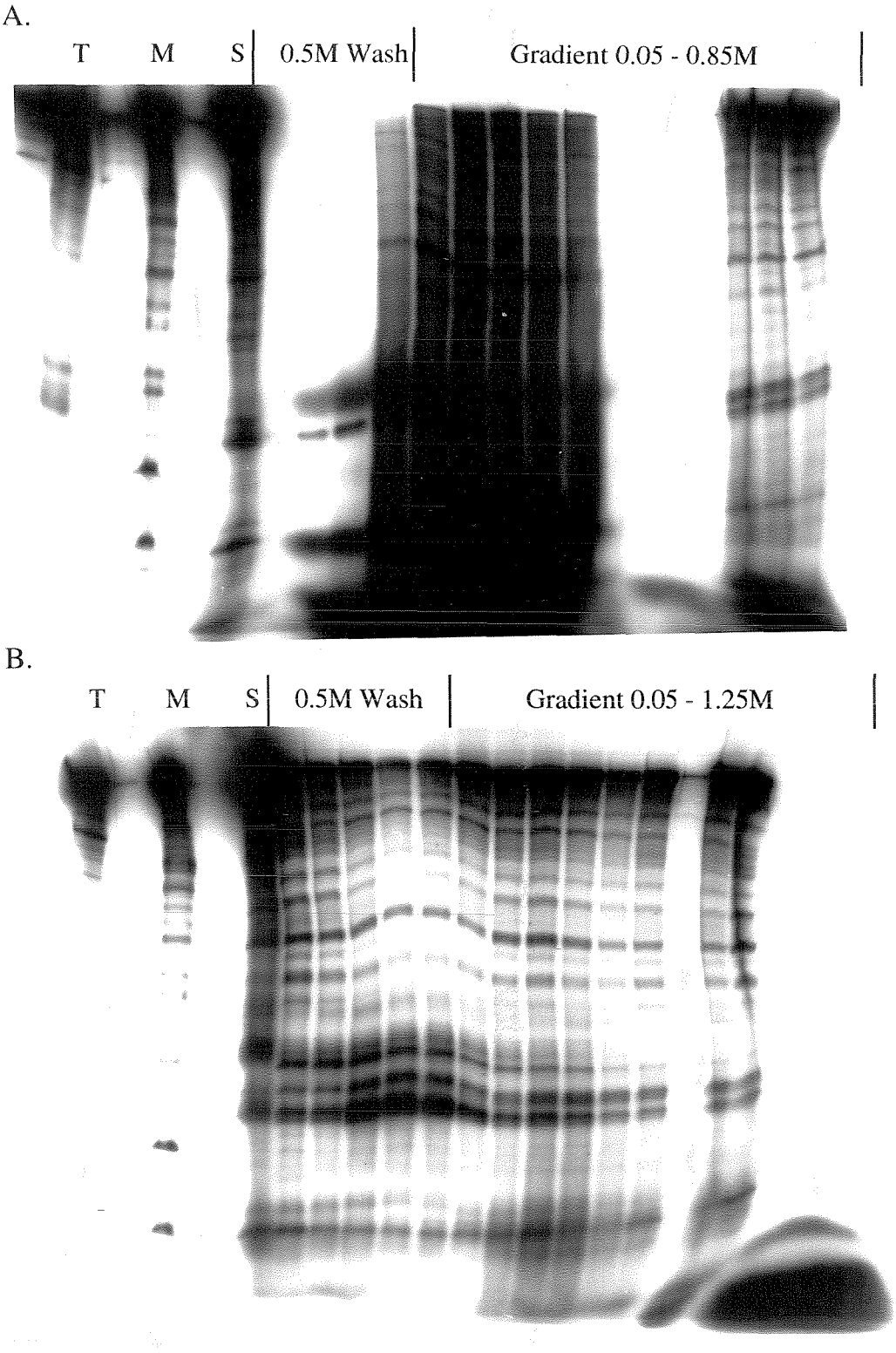

A.
T
M
S
0.5M Wash
Gradient 0.05 - 0.85M
B.
T
M
S
0.5M Wash
Gradient 0.05 - 1.25M

Figure IV-8

the various reaction velocities used to generate a Eadie-Hofstee plot (not shown). This plot indicates a Michaelis constant, which should approximate the dissociation constant of ligase for this substrate, of 66 nM. This was therefore taken as a estimation of the K_d for ligase and a true substrate. It was decided to try to favor specific binding by utilizing a high local concentration of immobilized protein in the presence of dilute total RNA where the concentration of any individual species will be significantly below the calculated K_d.

In order to provide a source of moderate amounts of easily purified *E. coli* RNA ligase which had not been exposed to RNA from a heterologous species, a tagged version of the enzyme was expressed in *E. coli*.. The coding sequence of the bifunctional affinity tag of the YpKRAP vector was subcloned into a pBS-RNA ligase clone at the 3' end of the ORF. This clone, whose expression was controlled by the unknown promoter elements apparently internal to the ligase ORF, was transformed into a ligase$^-$ strain where it supported the expression of ligase activity equal to that of untagged clones . This activity could be bound to and specifically eluted from both M2 anti-FLAG and nickel-NTA-agarose affinity resins. Sufficient tagged ligase for an RNA binding experiment was prepared from a one liter culture of this strain. Dual-affinity tagged protein was expressed in *E. coli* in a ligase$^-$ background, and purified on M2 anti-FLAG and nickel-NTA-agarose matrices.

While immobilized on the nickel resin, RNA ligase protein was exposed to dilute concentrations of total RNA . In order facilitate the identification of specific RNA ligands binding to *E. coli* RNA ligase, two types of total RNA populations were prepared. Total RNA isolated from a ligase$^-$ strain of *E. coli* which was exposed to 5% methanol during culture was expected to be enriched in ligase substrate(s). RNA prepared from a wild-type strain containing a pBS-RNA ligase overexpressing clone should contain little ligase substrate, but may have an increased proportion of the product of ligation. 500ml cultures of each of these were grown overnight and total RNA was harvested. Each total *E. coli* RNA preparation was diluted to a final concentration of about 100nM RNA,

leaving all individual species at concentrations below the estimated K_d of ligase for substrates. After allowing the RNA binding reaction time to come to equilibrium, the nickel matrix-ligase-RNA complex was recovered, poured as a column, and washed. RNA species of increasing affinity for the ligase were eluted with a gradient of KCl, and RNA in individual fractions was end-labeled and subjected to PAGE for visualization by autoradiography. The pattern of RNA species eluted from each total RNA preparation was complex, but clearly differed between the two.

Figure IV-9 shows that a wide size range of RNA molecules were recovered in all gradient fractions from both binding reactions. The pattern of eluted RNA in the two gradients does however show some differences which may represent the presence of different ligands of *E. coli* RNA ligase. The bulk of rRNA-sized molecules appears in the unbound supernatant from both reactions, but a significant amount is recovered in the highest salt fractions of RNA derived from a ligase$^-$ strain. In comparison, the RNA recovered from a ligase-overproducing wild type strain appears to be enriched in an RNA species of approximately 150-200N which elutes in lower salt wash fractions. Whether or not such differences can be used to identify actual *in vivo* substrates of the *E. coli* ligase will depend on whether they are reproducible as the protocol is improved for greater specificity. The results of this experiment do predict that the *in vitro* RNA binding approach can be successful for identifying ligands of this enzyme.

Discussion:

Phenotypes of Altered *E. coli* RNA Ligase Expression *in vivo*:

The ability to abolish expression of the *E. coli* 2'-5' RNA ligase *in vivo* and to restore expression at an elevated level has allowed observation of some of the various phenotypes caused by titration of ligase activity. RNA ligase function is not absolutely required for viability under standard laboratory culture conditions. This suggests that

although the ligase is expressed constitutively in all strains of *E. coli* tested, its function might be a subtle regulatory one or one that is required only under certain physiological conditions. This theory is supported by the conditional induction of ligase activity upon exposure of cells to environmental ethanol or methanol. It is further supported by the abnormal growth curve of ligase-overexpressing cultures which display a lower stationary culture density. The entrance into stationary phase due to depletion of nutrients at high cell density is a stressor, and it is marked by a global change in the pattern of gene expression including induction of over 30 stress response proteins, most in a regulon recognized by a stress-specific sigma factor σ^S (64). One particular stress-induced *E. coli* protein, PspA of the phage shock operon, shows a functional requirement similar to that of the *E. coli* RNA ligase. PspA function, which is induced by stressors including ethanol and the onset of stationary phase, is dispensable during logarithmic growth but required for survival during stationary phase (118). It should be enlightening to observe how endogenous *E. coli* RNA ligase expression might vary throughout the growth curve.

The effects of overexpression of RNA ligase on growth also point to a role for this enzyme in tRNA metabolism. At 43°C, all overexpressing strains are inviable. This may be due to an increased intracellular concentration of ligase protein sequestering it's substrate or product, or simply non-specifically binding other RNA(s) which are required for metabolism.

It is difficult to reconcile the catalytic activity and apparent substrate range of *E. coli* RNA ligase with the phenotypes resulting from altered expression levels of this enzyme. One possibility is that the ligase is necessary to modify a specific tRNA species so that it might be efficiently utilized for translation of stress response proteins necessary for growth during stationary phase. Alternatively, a heretofore unknown nuclease activity might be activated upon the transition to stationary phase and cleave tRNA

molecules analogously to the anticodon nuclease. *E. coli* RNA ligase-mediated repair of these tRNAs might then be necessary for translation and survival.

RNA Populations

The alteration of steady-state RNA levels seen in RNA extracted from *E. coli* strains expressing various levels of 2'-5' RNA ligase provide additional evidence of a direct role for this enzyme in RNA metabolism *in vivo*. The accumulation of tRNA in knockouts and depletion of tRNA in overexpressing strains, as well as the appearance of low molecular weight RNA fragments in overexpressors, point towards a possible role for cleavage by the *E. coli* ligase in some sort of tRNA processing. The possibility of this enzyme acting as a nuclease is discussed below.

Equilibrium of the RNA Ligase Reaction:

The reaction catalyzed by the *E. coli* ligase enzyme was shown to be fully reversible with an apparent equilibrium constant near unity, but slightly favoring cleavage of 2'-5' bonds. The tendency towards cleavage can be explained by the thermodynamics of the reaction, which have been investigated in the hammerhead ribozyme system. The internal equilibrium of the hammerhead ribozyme favors cleavage of 3'-5' phosphodiesters to 2',3' cyclic phosphate and 5'-hydroxyl termini despite the unfavorable enthalpy associated with cyclic phosphate formation due to the favorable entropy of bond cleavage (41). Increase of entropy in what is essentially a unimolecular reaction (given the tight structural association of tRNA halves) may be due to the additional degrees of freedom available to released termini and to disruption of water ordering in and about the closed, structured anticodon loop.

This observed equilibrium of cleavage and ligation explains the maximum extent of ligation observed in *in vitro* activity assays. It also begs the question of whether the function of the enzyme is to catalyze ligation or cleavage *in vivo*. Despite the fact that the

equilibrium observed *in vitro* favors cleavage, the direction of the equilibrium *in vivo* will depend on the effective concentrations of substrates available for each reaction. If ligated tRNA products are utilized for translation, and thereby removed from the pool of substrates available to the ligase, then the ligation reaction will be favored. If, however the cleavage products are removed by some process such as ribonucleolytic degradation, then the equilibrium will favor cleavage. To propose a cleavage function for the ligase enzyme in vivo, however, it is necessary to posit a source of substrates with 2'-5' bonds, presumably in the context of a tRNA. Although a variety of endoribonuclease are known which could theoretically produce substrates for ligation by the *E. coli* RNA ligase, and in fact an activity capable of doing so has been observed in *E. coli* extracts, no other *E. coli* enzyme is known or proposed which might form internal 2'-5' linkages in a tRNA structure. Thus the available evidence, while circumstantial, favors a ligation function for this enzyme *in vivo*.

E. coli 2'-5' RNA Ligase Transcriptional Promoter:

The results of attempts to overexpress 2'-5' RNA ligase in *E. coli* using several different expression systems imply the existence of a very unusual promoter for the ligase gene. An identical pattern of expression was seen when the ORF encoding the ligase was cloned in four different vectors under the control of three different inducible promoters: a consistent level of expression of approximately 10 times that of the chromosomal gene in the absence of inducible promoter activation with no significant increase in expression upon induction. The location of the promoter controlling expression would appear to be internal to the coding sequence as clones including 200bp of upstream sequence in front of the coding region did not express at consistently different levels than clones which included only the predicted ORF with five base pairs of heterologous upstream sequence (bearing no similarity to promoter consensus sequences) added to provide a restriction site for cloning. This promoter may in fact be the major element controlling expression of

the chromosomal version of the gene, with overexpression caused simply by the increased copy number of the plasmid-borne copy. This would imply that this internal promoter is highly regulated to control RNA ligase expression levels, which is consistent with its ability to interfere with strong upstream promoting elements. It is unlikely that sequences at the 5' end of the ORF as cloned are actually directing transcription from some internal start site for translation of a truncated ligase protein as the next methionine in the sequence is 90 amino acids downstream. Translation from this residue would produce an 86 residue peptide of approximately 9kD which is not observed in extracts of overexpressing strains. It seems unlikely that a fragment that small, representing less than half of the native protein, would be able to reconstitute enzymatic activity. This would indicate that the RNA ligase exhibits a very unusual gene structure for *E. coli* , where a promoter internal to one gene may commonly regulate a downstream gene but promoters internal to the genes they regulate are seen only in tRNA genes.

The presence of an *E. coli* promoter within an open reading frame appears somewhat paradoxical given the recognition by *E. coli* RNA polymerase of consensus sequences at base pairs -10 and -35 relative to the transcriptional start site. The recognition of internal promoter sequences is known to occur during transcription by eukaryotic RNA polymerase III, but has not been described for a bacterial polymerase. Elucidation of the mechanism of recognition and delineation of sequence elements promoting transcription in these clones must therefore await further experimentation, such as RNAse protection studies to determine the start site of transcription and DNase footprinting to locate protein binding sites. Subcloning of ligase gene sequences into a promoter probe vector may also be used to locate those elements capable of activating transcription.

Expression of Ligase in *S. cerevisiae* and Production of Anti-RNA Ligase Antibodies:

The development of a heterologous expression system and the production and purification of antibodies recognizing the ligase protein do not in themselves reveal much information about function of *E. coli* RNA ligase. Both of these processes have been successful though, in creating biochemical reagents which will be invaluable for future studies of this enzyme. Heterologous expression in *S. cerevisiae* is a reasonable method for producing large amounts of *E. coli* RNA ligase for use in biochemical experiments. This protein can not only be utilized for standard enzymatic analyses, but can also be easily immobilized via the peptide tag for use in RNA binding or protein-protein interaction experiments and affinity purification of anti-ligase antibodies. If the efficiency of purification can be improved, for example at the M2 resin binding step, considerably higher yields might be achieved. In fact, if a large enough concentration and variety of protease inhibitors can be added, it may be possible to bind tagged protein directly to M2 matrix in the S-30 supernatant without inactivation of the antibody resin. It is conceivable that large enough amounts of ligase could even be isolated from yeast extracts to crystallize for structural determination, especially if a specific protease site is added to facilitate removal of the peptide tag after purification.

Trial preparations indicate that it will be possible to produce milligram amounts of pure affinity-tagged 2'-5' RNA ligase from small fermentations of *S. cerevisiae*, particularly if the parameters of expression and purification conditions are optimized further. Such large preparative amounts of protein may be used for kinetic analyses, ligand binding studies, affinity purification of antibodies, and structural studies of the enzyme. Antibodies recognizing the *E. coli* RNA ligase protein have been raised in rabbits and partially purified by affinity purification. Although this purification was not completely successful in isolating ligase-specific antibodies at high titer, there remains a very good chance of being able to do so. It may be possible to load enough immune serum onto a nickel-bound RNA ligase column that when the binding reaction is allowed

to proceed to equilibrium, all available ligase epitopes will be occupied by specific anti-ligase antibodies. This would have the effects of increasing both the purity and concentration of the desired antibodies. A potential pitfall of this technique is the possibility that the lack of specificity seen on westerns is due not to the impurity of specific antibodies, but to inherent non-specific binding. It may however also be possible to separately purify antibodies recognizing individual peptides used as antigens by employing affinity columns displaying the individual synthetic peptides. This could allow testing of the specificity and avidity of individual epitypes for recognition of the entire RNA ligase protein to determine their utility for western probing, immunoprecipitation, or activity interference experiments. If they can be purified and concentrated more efficiently using a refined version of the protocol developed for this purpose, purified anti-ligase polyclonal antibodies should be useful for a variety of biochemical experiments. These may include identification of native ligands of the enzyme by immunoprecipitation from crude extracts, for tracking ligase protein expression in vivo through the cell cycle or growth curve by western, or for interference studies to localize the active site within the enzyme's primary structure.

in vitro RNA Binding:

If the results of this preliminary experiment in *in vitro* RNA selection by affinity to *E. coli* 2'-5' RNA ligase can be replicated in a more stringent binding assay, they may reveal the general metabolic function of this enzyme. This indicates that the specificity of binding could be improved, possibly by using even more dilute RNA solutions with a higher salt concentration for binding and by adding extensive washing steps before RNA elution. Subfractionated RNA populations, such as tRNA or rRNA alone may provide a starting material more enriched in specific substrates for binding. Additionally, a useful control experiment would be the inclusion of an assay of the binding of RNA to the affinity resins in the absence of RNA ligase (utilizing a ligase- strain extract for the initial

M2 binding). This would reveal any non-ligase mediated retention of RNA. It appears from this data that abundant high molecular weight transcripts in total RNA isolated from a ligase$^-$ strain are bound avidly by the enzyme, and that a low molecular weight species is enriched in total RNA from a ligase overexpressing strain but released from immobilized ligase in low salt. One possible explanation of this might be a role for the 2'-5' ligase in the processing of multicistronic transcripts of stable RNAs. 2'-5' linkages introduced into tRNA or tRNA-like portions of these precursor transcripts may serve as recognition elements for specific processing nucleases.

This experiment serves as a pilot study of the feasibility of identifying high affinity RNA ligands of the *E. coli* RNA ligase by binding to recombinant ligase protein *in vitro*.

Although this procedure did demonstrate that *in vitro* RNA binding to purified, immobilized ligase and differential elution of ligands of varying affinity is possible, the protocol will need to be refined (and additional controls performed) in order to unambiguously isolate high-affinity RNAs. Once individual RNA ligands can be clearly identified, they can be amplified by ligation-mediated reverse transcription and PCR, and then cloned for sequence analysis. This method shows great promise for the identification of in vivo substrates and corresponding products of the *E. coli* RNA ligase.

Materials and Methods:

Growth Curves: Cultures were grown in LB plus selective antibiotics (Tet for P1 transducants, Kan for strains with genomic disruptions, and Amp in all strains carrying expression plasmids). 1µl aliquots were withdrawn at various time points, diluted 1:10^4 in LB, and aliquots plated on selective media for growth at 37°C.

Overexpression plasmids - pET 11 and 19 vectors were from Novagen. pATH vectors used were as described by Koerner (56). pKRAP vector was constructed as described. pKRAP was constructed in the yeast multicopy vector pRS424 (18)by the addition of the CUP1 promoter amplified by PCR from the *S. cerevisiae* metallothionein gene and the transcriptional terminator of the *S. cerevisiae* ADH1 gene, also amplified from genomic (10, 34). The sequence of the dual affinity tag is as follows: MSH_8ADYKDDDDKDP, plus several additional c-terminal residues encoded by the vector, which varied according to the cloning scheme utilized.

Purification of tagged RNA ligase from yeast: Frozen cells were thawed and extracted in the presence of yeast extraction buffer (40mM HEPES pH 7.8, 10% glycerol, 5mM βME) containing protease inhibitors and 150mM KCl. Soluble extract was collected after centrifugation at 30,000xg for one hour and flowed through a 400ml bed of DEAE-sepharose. The DEAE outflow was loaded onto a 150ml column of cellulose phosphate which was then washed with buffer containing 300mM KCl. Ligase activity was eluted with a gradient of 300mM-1.2M KCl. Some ligase activity was detected in the cellulose phosphate flow-through, but the majority was present in the 300mM and early gradient fractions. The 300mM wash and active gradient fractions were combined, brought to 500mM KCl and 5mM imidazole (a competitor for histidine binding to nickel) and mixed batchwise with 12 ml of nickel-NTA-agarose for 3 hours at 4°C. The nickel resin was then poured as a column and washed with 700mM KCl and 7mM imidazole. Approximately 90% of ligase activity was retained on the column. Specifically bound proteins were eluted with 250mM imidazole in the presence of 700mM KCl. Elution fractions containing RNA ligase were pooled, diluted to 200mM KCl, and slowly loaded onto a 5ml bed of M2 anti-FLAG Mab-agarose. At this step only an estimated 50% of ligase activity remained bound to the column. After washing at the same concentration of

KCl, FLAG-tagged protein was specifically eluted by addition of 700μM free FLAG peptide.

Antibodies - Synthetic peptides corresponding to regions of the *E. coli* RNA ligase were synthesized and conjugated to KLH by the Caltech Microchemical Synthesis Facility. They were injected in an equimolar mixture into rabbits on a regular schedule by Cocalico Biologicals, who also provided periodic serum samples. Purification- 70 micrograms of Purified, tagged *E. coli* RNA ligase expressed in yeast was bound batchwise to 1ml of nickel-NTA-agarose which was then poured as a column and washed. 0.5ml of immune serum was loaded onto the column and allowed to bind, then washed extensively with 2M KCl to remove nonspecifically bound serum proteins. Ligase-specific antibodies were eluted by the addition of 4M $MgCl_2$. Mg^{2+} elution fractions were pooled, concentrated by dialysis against PEG 20K, and dialyzed into PBS

in vitro RNA binding : An overnight one liter culture of cells expressing tagged *E. coli* 2'-5' RNA ligase over a genomic ligase knockout yielded 2.6 grams of cells. These cells were disrupted by sonication on ice in pH 7.4 extraction buffer containing 150mM KCl and a battery of protease inhibitors, and soluble extract was recovered after centrifugation at 8,000xg, 4°C, for 10 minutes. All subsequent steps were also performed at 4°C. Tagged ligase in the extract was bound batchwise to 1 ml of M2 resin for one hour. This resin was then recovered by centrifugation and washed extensively in buffer containing 200mM KCl but lacking metal chelators. FLAG-tagged molecules were eluted with sequential aliquots of 300μM FLAG peptide in a pH 7.8 buffer containing 200mM KCl, for a recovery of approximately 350μg of protein. M2 elution fractions were pooled, brought to 500mM KCl and 8mM imidazole, and bound batchwise to 10 ml of nickel-NTA-agarose for several hours. The nickel resin was recovered, washed, and eluted with 250mM imidazole. Imidazole eluate was diluted to 50mM KCl and bound batchwise to 0.5 ml M2 resin for use in RNA binding.

Total RNA was prepared by protoplasting with lysozyme, lysis of protoplasts in the presence of the ribonuclease inhibitor DEPC, and subsequent precipitation of chromosomal DNA and proteins with SDS and NaCl. Crude RNA was recovered from the supernatant by ethanol precipitation, then digested with RQ1 DNAse and proteinase K, extracted with phenol and chloroform, and ethanol precipitated a second time. Approximately 10mg of pure RNA was recovered from each prep. Each total *E. coli* RNA preparation was diluted with 500ml of buffer containing 50mM KCl to result in a final concentration of about 100nM RNA, leaving all individual species at concentrations below the estimated K_D of ligase for substrates. 0.25ml of ligase-M2 resin complex was added to each, and binding was allowed to occur over several hours with gentle agitation. The M2 resin in each was then recovered and eluted with 300μM FLAG peptide. M2 eluates were then onto 2ml columns of nickel-NTA-agarose. RNA was eluted from these columns with a gradient from 50mM to 1M KCl, followed by a 2M KCl wash, and a 200mM imidazole wash to release bound ligase. RNAs in each gradient fraction were labeled by yeast poly-A polymerase with α-^{32}P-cordycepin. Labeled RNAs were separated by PAGE on 4-10% gradient gels and detected by autoradiography.

References

1. **Amitsur, M., Levitz, L, and Kaufmann, G.** 1987. Bacteriophage T4 Anticodon Nuclease, Polynucleotide Kinase, and RNA ligase Reprocess the Host Lysine tRNA. EMBO J **6:**2499-2503.
2. **Amitsur, M., Morad, I, Chapman-Shimshoni, D, and Kaufmann, G.** 1992. HSD Restriction-Modification Proteins Partake in Latent Anticodon Nuclease. EMBO J **11:**3129-3134.
3. **Andachi, Y., Yamao, F, Muto, A, and Osawa, S.** 1989. Codon Recognition Patterns as Deduced from Sequences of the Complete Set of Transfer RNA Species in Mycoplasma capricolum - Resemblance to Mitochondria. J Mol Biol **209:**37-54.
4. **Apostol, B., and Greer, CL.** 1991. Preferential Binding of Yeast tRNA Ligase to pre-tRNA Substrates. Nuc. Acids Res. **19:**1853-1860.
5. **Arnez, J., Steitz, TA.** 1994. Crystal-Structure of Unmodified tRNA(GLN) Complexed With Glutaminyl-Transfer-RNA Synthetase and ATP Suggests a Possible Role for Pseudo-Uridines in Stabilization of RNA Structure. Biochem. **33:**7560-7567.
6. **Bakalara, N., Simpson, AM, and Simpson, L.** 1989. The Leishmania Kinetoplastid-Mitochondrion Contains Terminal Uridylyltransferase and RNA Ligase Activities. J. Biol. Chem. **264:**18679-18686.
7. **Barrio, J., Barrio, MCG, Leonard, NJ, England, TE, and Uhlenbeck, OC.** 1978. Synthesis of Modified Nucleoside 3',5'-Bisphosphates and Their Incorporation into Oligoribonucleotides with T4 RNA Ligase. Biochem. **17:**2077-2081.
8. **Baymiller, J., Jennings, S, Kienzle, B, Gorman, JA, Kelly, R, and McCullough, JE.** 1994. Isolation and Sequence of the tRNA Ligase-Encoding Gene of Candida albicans. Gene **142:**129-134.
9. **Belford, H., Westaway, SK, Abelson, J, and Greer, CL.** 1993. Multiple Nucleotide Cofactor Use by Yeast Ligase in tRNA Splicing. J Biol. Chem. **268:**2444-2450.
10. **Bennetzen, J., and Hall, BD.** 1982. The Primary Structure of the Saccharomyces cerevisiae Gene for Alcohol Dehydrogenase I. J Biol. Chem. **257:**3018-3225.
11. **Bickle, T., and Kruger, DH.** 1993. Biology of DNA Restriction. Microb. Rev. **57:**434-450.
12. **Biniszkiewicz, D., Cesnaviciene, E, and Shub, DA.** 1994. Self-splicing Group-I Intron in Cyanobacterial Initiator Methionine Transfer-RNA - Evidence for Lateral Transfer of Introns in Bacteria. EMBO J **13:**4629-4635.
13. **Bjork, G.** 1987. Modification of Stable RNA, p. 719-731. *In* F. Neidhardt (ed.), Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology., vol. Vol. 1. American Society for Microbiology, Washington, DC.
14. **Blum, B., Sturm, NR, Simpson, AM, and Simpson, L.** 1991. Chimeric gRNA-mRNA Molecules with Oligo(U) Tails Covalently Linked at Sites of RNA Editing Suggest That U Addition Occurs by Transesterification. Cell **65:**543-550.
15. **Branch, A., Robertson, HD, Greer, C, Gegenheimer, P, Peebles, C, and Abelson, J.** 1982. Cell-Free Circularization of Viroid Progeny RNA by an RNA Ligase from Wheat Germ. Science **217:**1147-1149.
16. **Bruce, A., and Uhlenbeck, OC.** 1978. Reactions at the Termini of tRNA With T4 RNA Ligase. Nuc. Acids Res. **5:**3665-3677.
17. **Bryant, F., and Benkovic, SJ.** 1982. Phosphorothioate Substrates for T4 RNA Ligase. Bioch. **21:**5877-5885.

18. **Christianson, T., Sikorski, RS, Dante, M, Shero, JH, and Hieter, P.** 1992. Multifunctional Yeast High-Copy-Number Shuttle Vectors. Gene **110:**119-122.
19. **Clark, M., and Abelson, J.** 1987. The Subnuclear Localization of tRNA Ligase in Yeast. J. Cell Biol. **105:**1515-1526.
20. **Cranston, J., Silber, R, Malathi, VG, and Hurwitz, J.** 1974. Studies on Ribonucleic Acid Ligase. J Biol. Chem. **249:**7447-7456.
21. **Culver, G., Consaul, SA, Tycowski, KT, Filipowicz, W, and Phizicky, EM.** 1994. tRNA Splicing in Yeast and Wheat Germ. A Cyclic Phosphodiesterase Implicated in the Metabolism of ADP-Ribose 1",2"-cyclic Phosphate. J Biol. Chem. **269:**24928-24934.
22. **Depew, R., and Cozzarelli, NR.** 1974. Genetics and Physiology of Bacteriophage T4 3'-Phosphatase: Evidence for Involvement of the Enzyme in T4 DNA Metabolism. J Virol. **13:**888-897.
23. **Ekland, E., Szostak, JW, and Bartel, DP.** 1995. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. Science **269:**364-370.
24. **England, T., and Uhlenbeck, OC.** 1978. Enzymatic Oligoribonucleotide Synthesis With T4 RNA Ligase. Biochem. **17:**2069-2076.
25. **Engler, M., and Richardson, CC.** 1982. DNA Ligases, p. 3-30. *In* P. Boyer (ed.), The Enzymes, 3rd ed, vol. 15B. Academic Press.
26. **Evans, C., and Engelke, DR.** 1990. Yeast Extracts for Transfer RNA Gene Transcription and Processing, p. 439-450. *In* J. Dahlberg, and Abelson, JN (ed.), Methods in Enzymology, vol. 181. Academic Press.
27. **Fersht, A.** 1977. Enzyme Structure and Mechanism. W.H. Freeman and Company, New York.
28. **Filipowicz, W., and Shatkin, AJ.** 1983. Origin of Splice Junction Phosphate in tRNAs Processed by HeLa Cell Extract. Cell **32:**547-557.
29. **Filipowicz, W., Konarska, M, Gross, HJ, and Shatkin, AJ.** 1983. RNA 3'-Terminal Phosphate Cyclase Activity and RNA Ligation in HeLa Cell Extract. Nuc. Acids Res. **11:**1405-1418.
30. **Findlay, D., Herries, DG, Mathias, AP, Rabin, BR, and Ross, CA.** 1961. The Active Site and Mechanism of Action of Bovine Pancreatic Ribonuclease. Nature **190:**781-784.
31. **Fitt, P., Sharma, N, and Ball, D.** 1992. Induction of Error-Free DNA Repair in Escherichia coli by Nonmutagenic Stress. Curr. Microb. **24:**251-256.
32. **Gegenheimer, P., Gabius, H-J, Peebles, CL, and Abelson, J.** 1983. An RNA Ligase from Wheat Germ Which Participates in Transfer RNA Splicing in vitro. J Biol. Chem. **258:**8365-8373.
33. **Gerlt, J., Westheimer, FH, and Sturtevant, JH.** 1975. The Enthalpies of Hydrolysis of Acyclic, Monocyclic, and Glycoside Cyclic Phosphate Diesters. J Biol. Chem. **250:**5059-5067.
34. **Gorman, J., Clark, PE, Lee, MC, Debouck, C, and Rosenberg, M.** 1986. Regulation of the Yeast Metallothionein Gene. Gene **48:**13-22.
35. **Greer, C., Peebles, CL, Gegenheimer, P, and Abelson, J.** 1983. Mechanism of Action of a Yeast RNA Ligase in tRNA Splicing. Cell **32:**537-546.
36. **Greer, C., Javor, B, and Abelson, J.** 1983. RNA Ligase in Bacteria: Formation of a 2',5' Linkage by an E. coli Extract. Cell **33:**899-906.
37. **Greer, C., Soll, D, and Willis, I.** 1987. Substrate Recognition and Identification of Splice Sites by the tRNA-Splicing Endonuclease and Ligase from Saccharomyces cerevisiae. Mol. Cell. Biol. **7:**76-84.
38. **Hajduk, S., Harris, ME, and Pollard, VW.** 1993. RNA Editing in Kinetoplastid Mitochondria. FASEB J **7:**55-63.

39. **Hamilton, C., Aldea, M, Washburn, BK, Babitzke, P, and Kushner, SR.** 1989. New Method for Generating Deletions and Gene Replacements in Escherichia coli. J Bact. **171:**4617-4622.
40. **Heaphy, S., Singh, M, Gait, MJ.** 1987. Effect of Single Amino Acid Changes in the Region of the Adenylylation Site of T4 RNA Ligase. Bioch. **26:**1688-1696.
41. **Hertel, K., and Uhlenbeck, OC.** 1995. The Internal Equilibrium of the Hammerhead Ribozyme Reaction. Biochem. **34:**1744-1749.
42. **Hottinger, T., Furst, P, Pohlig, G, and Heim, J.** 1994. Physiological Characterization of the Yeast Metallothionein (CUP1) Promoter, and Consequences of Overexpressing its Transcriptional Activator, ACE1. Yeast **10:**283-296.
43. **Huang, J., and Van der Ploeg, LHT.** 1988. A 5'-Exoribonuclease and RNA Ligase of T. bruceii. Nuc. Acids Res. **16:**9737-9759.
44. **Johnson, P., and Abelson, J.** 1983. The Yeast tRNATyr Gene Intron is Essential for Correct Modification of its tRNA Product. Nature **302:**681-687.
45. **Juodka, B., Marksutskas, AY, Snechkute, MA, Zilinskiene, VJ, and Drigin, YF.** 1980. On the Structure of the RNA Ligase-Adenylate Complex. Bioorg. Khim. **6:**1733-1734.
46. **Juodka, B., Labeikite, DY, and Sasnauskene, SI.** 1991. Substrate Specificity of T4 RNA Ligase. The Role of the Phosphate Residue of the Nucleotide in the Formation of a Covalent AMP-RNA Ligase Complex. Biochem. Mos. **56:**540-545.
47. **Juodka, B., Labeikite, DJ, and Sasnauskiene, SI.** 1993. Substrate Specificity of T4 RNA-Ligase: Role of the Purine Base in Formation of the Covalent AMP-RNA Ligase Complex. Bioch. Mos. **58:**857-865.
48. **Kaufmann, G., and Kallenbach, NR.** 1975. Determination of Recognition Sites of T4 RNA Ligase on the 3'-OH and 5'-P Termini of Polyribonucleotide Chains. Nature **254:**452-454.
49. **Kawamukai, M., Utsumi, R, Takeda, K, Higashi, A, Matsuda, H, Choi, Y-L, and Komano, T.** 1991. Nucleotide Sequence and Characterization of the sfs1 Gene: sfs1 is Involved in CRP*-dependent mal Gene Expression in Escherichia coli. J Bact. **173:**2644-2648.
50. **Kiberstis, P., Haselhoff, J, and Zimmern, D.** 1985. 2' Phosphomonoester, 3'-5' Phosphodiester Bond at a Unique Site in a Circular Viral RNA. EMBO J **4:**817-827.
51. **Kikuchi, Y., Tyc, K, Filipowicz, W, Sanger, HL, and Gross, HJ.** 1982. Circularization of Linear Viroid RNA via 2'-Phosphomonoester, 3',5'-Phosphodiester Bonds by a Novel Type of RNA Ligase from Wheat Germ and Chlamydomonas. Nuc. Acids Res. **10:**7521-7529.
52. **Kjems, J., Jensen, J, Olesen, T, and Garrett, RA.** 1989. Comparison of Transfer RNA and Ribosomal RNA Intron Splicing in the Extreme Thermophile and Archeabacterium Desulfurococcus mobilis. Can. J. Microbiol. **35:**210-214.
53. **Kjems, J., and Garrett, RA.** 1988. Novel Mechanism of Splicing for the Ribosomal RNA Intron in the Archaebacterium Desulfurococcus mobilis. Cell **54:**693-703.
54. **Knapp, G., Beckmann, JS, Johnson, PF, Fuhrman, SA, and Abelson, J.** 1978. Transcription and Processing of Intervening Sequences in Yeast tRNA Genes. Cell **14:**221-236.
55. **Kodama, K.-I., Barnes, DE, and Lindahl, T.** 1991. In vitro Mutagenesis and Functional Expression in Escherichia coli of a cDNA encoding the Catalytic Domain of Human DNA Ligase I. Nuc. Acids Res. **19:**6093-6099.
56. **Koerner, T., Hill, JE, Myers, AM, and Tzagoloff, A.** 1991. High-Expression Vectors with Multiple Cloning Sites for Construction of trpE Fusion Genes: pATH Vectors, p. 477-490, Methods in Enzymology, vol. 194. Academic Press.

57. **Komine, Y., Adachi, T, Inokuchi, H, and Ozeki, H.** 1990. Genomic Organization and Physical Mapping of the Transfer-RNA Genes in Escherichia coli K12. J Mol. Biol. **212:**579-598.
58. **Konarska, M., Filipowicz, W, Domdey, H, and Gross, HJ.** 1981. Formation of a 2'-phosphomonoester, 3',5'-phosphodiester Linkage by a Novel RNA Ligase in Wheat Germ. Nature **293:**112-116.
59. **Koonin, E., and Gorbalenya, AE.** 1990. Related Domains in Yeast tRNA Ligase, Bacteriophage T4 Polynucleotide Kinase and RNA Ligase, and Mammalian 2',3'-cyclic Nucleotide Phosphohydrolase Revealed by Amino Acid Sequence Comparison. FEBS **268:**231-234.
60. **Kossel, H., and Seliger, H.** 1975. Recent Advances in Polynucleotide Synthesis, p. 298-508. *In* W. Herz, Grisebach, H, and Kirby, GW (ed.), Progress in the Chemistry of Organic Natural Products, vol. 32.
61. **Krug, M., and Uhlenbeck, OC.** 1982. Reversal of T4 RNA Ligase. Bioch. **21:**1858-1864.
62. **Levitz, R., Chapman, D, Amitsur, M, Green, R, Snyder, L, and G. Kaufmann.** 1990. The Optional E. coli prr Locus Encodes A Latent Form ofPhage T4-Induced Anticodon Nuclease. EMBO J **9:**1383-1389.
63. **Liong, E., and Ferenci, T.** 1994. Molecular Cloning of a Maltose Transport Gene from Bacillus stearothermophilus and its Expression in Escherichia coli K12. Mol. Gen. Genetics **243:**343-352.
64. **Loewen, P., and Hengge-Aronis, R.** 1994. The Role of the Sigma Factor σs (KatF) in Bacterial Global Regulation. Ann. Rev. Microb. **48:**53-80.
65. **McCraith, S., and Phizicky, EM.** 1991. An Enzyme from Saccharomyces cerevisiae uses NAD+ to Transfer the Splice Junction 2'-Phosphate from Ligated tRNA to an Acceptor Molecule. J Biol. Chem. **266:**11986-11992.
66. **McLaughlin, L., Piel, N, and Graeser, E.** 1985. Donor Activation in the T4 RNA Ligase Reaction. Biochem. **24:**267-273.
67. **Meinnel, T., Mechulam, Y, Blanquet, S.** 1993. Methionine as Translation Start Signal - A Review of the Enzymes of the Pathway in Escherichia coli. Biochimie **75:**1061-1075.
68. **Moseman McCoy, M., and Gumport, RI.** 1980. T4 Ribonucleic Acid Ligase Joins Single-Strand Oligo(deoxyribonucleotide). Biochem. **19:**635-642.
69. **Nishikura, K., and DeRobertis, EM.** 1981. RNA Processing in Microinjected Xenopus Oocytes: Sequential Addition of Base Modifications in a Spliced Transfer. J Mol. Biol. **145:**405-420.
70. **Ost, K. a. D., MP.** 1990. RNase PH Catalyzes a Synthetic Reaction, the Addition of Nucleotides to the 3' end of RNA. Biochimie **72:**813-818.
71. **Parthasrathy, R., Malik, M, and Fridey, SM.** 1982. X-ray Structure of a Dinucleoside Monophosphate A2'p5'C that Contains a 2'-5' Link Found in (2'-5')oligo(A)s Induced by Interferons: SIngle-stranded Helical Conformation of 2'-5'-linked Oligonucleotides. Proc. Natl. Acad. Sci. USA **79:**7292-7296.
72. **Peebles, C., Gegenheimer, P, and Abelson, J.** 1983. Precise Excision of Intervening Sequences from Precursor tRNAs by a Membrane-Associated Yeast Endonuclease. Cell **32:**525-536.
73. **Perkins, K., Furneaux, H, and Hurwitz, J.** 1985. Isolation and Characterization of an RNA Ligase from HeLa Cells. Proc. Natl. Acad. Sci. USA **82:**684-688.
74. **Perret, V., Garcia, A, Grosjean, H, Ebel, J-P, Florentz, C, and Giege, R.** 1990. Relaxation of a Transfer RNA Specificity by Removal of Modified Nucleotides. Nature **344:**787-789.

75. **Phizicky, E., Schwartz, RC, and Abelson, J.** 1986. Saccharomyces cerevisiae tRNA Ligase. Purification of the Protein and Isolation of the Structural Gene. J Biol. Chem. **261:**2978-2986.
76. **Phizicky, E., Consaul, SA, Nehrke, KW, and Abelson, J.** 1992. Yeast tRNA Ligase Mutants are Nonviable and Accumulate tRNA Splicing Intermediates. J Biol. Chem. **267:**4577-4582.
77. **Pick, L., Furneaux, HM, and Hurwitz, J.** 1990. Purification and Characterization of Wheat Germ RNA Ligase and Associated Activities, p. 480-499. *In* J. Dahlberg, and Abelson, JN (ed.), Methods in Enzymology, vol. 181.
78. **Pick, L., and Hurwitz, J.** 1986. Purification of Wheat Germ RNA Ligase I. Characterization of a Ligase-Associated 5'-Hydroxyl Polynucleotide Kinase Activity. J Biol. Chem. **261:**6684-6693.
79. **Pick, L., Furneaux, H, and Hurwitz, J.** 1986. Purification of Wheat Germ RNA Ligase II. Mechanism of Action of Wheat Germ RNA Ligase. J Biol. Chem. **261:**6694-6704.
80. **Piller, K., Decker, CJ, Rusche, LN, and Sollner-Webb, B.** 1995. Trypanosoma brucei Mitochondrial Guide RNA-mRNA Chimera-Forming Activity Cofractionates with an Editing-Domain-Specific Endonuclease and RNA Ligase and is Mimicked by Hetereologous Nuclease and RNA Ligase. Mol. Cell. Biol. **15:**2925-2932.
81. **Podder, S.** 1970. Synthetic Action of Ribonuclease T1. Bioch. Biop. Acta **209:**455-462.
82. **Przykorska, A.** 1995. Influence of Modified Nucleosides on tRNA Structure as Probed by Two Plant Nucleases. Biochimie **77:**109-112.
83. **Rand, K., and Gait, MJ.** 1984. Sequence and Cloning of Bacteriophage T4 gene 63 Encoding RNA Ligase and Tail Fibre Attachment Activities. EMBO J **3:**397-402.
84. **Rauhut, R., Green, PR, and Abelson, J.** 1990. Yeast tRNA-Splicing Endonuclease is a Heterotrimeric Enzyme. J Biol. Chem. **265:**18180-18184.
85. **Reinholdhurek, B., and Shub, DA.** 1992. Self-splicing Introns in Transfer-RNA Genes of Widely Divergent Bacteria. Nature **357:**173-176.
86. **Reyes, V., and Abelson, J.** 1987. A Synthetic Substrate for tRNA Splicing. Analyt. Bioch. **166:**90-106.
87. **Runnels, J., Soltis, D, Hey, T, and Snyder, L.** 1982. Genetic and Physiological Studies of the Role of RNA Ligase of Bacteriophage T4. J Mol. Biol. **154:**273-286.
88. **Rusche, L., Piller, KJ, and Sollner-Webb, B.** 1995. Guide RNA Chimeras, Which are Potential RNA Editing Intermediates, Are Formed by Endonuclease and RNA Ligase in a Trypanosome Mitochondrial Extract. Mol. Cell. Biol. **15:**2933-2941.
89. **Sabalyauskene, V., Kiss, L, and Juodka, B.** 1991. The Role of the Hydroxyl Groups of the Tyrosine Residues of RNA Ligase. Biochem. Mos. **55:**883-890.
90. **Sabatini, R., and Hajduk, SL.** 1995. RNA Ligase and its Involvement in Guide RNA/mRNA Chimera Formation - Evidence for a Cleavage-Ligation Mechanism of Trypanosoma brucei Messenger RNA Editing. J Biol. Chem. **270:**7233-7240.
91. **Sampson, J., and Saks, ME.** 1993. Contributions of Discrete tRNASer Domains to Amino-acylation by E. coli Seryl-tRNA Synthetase: a Kinetic Analysis Using Model RNA Substrates. Nuc. Acids Res. **21:**4467-4475.
92. **Schagger, H., Link, TA, Engel, WD, and von Jagow, G.** 1986. Isolation of the Eleven Protein Subunits of the bc1 Complex from Beef Heart, p. 224-237. *In* S. Fleischer, and Fleischer, B (ed.), Methods in Enzymology, vol. 126. Academic Press.

93. **Schwartz, R., Greer, CL, Gegenheimer, P, and Abelson, J.** 1983. Enzymatic Mechanism of an RNA Ligase from Wheat Germ. J Biol. Chem. **258:**8374-8383.

94. **Shapero, M., and Greer, CL.** 1991. Conformational Transition Required for Efficient Splicing of Transcripts from Hybrid λ Promoter Yeast tRNA Gene Fusion. Biochem. **30:**6465-6474.

95. **Shuman, S., Liu, Y, and Schwer, B.** 1994. Covalent Catalysis in Nucleotidyl Transfer Reactions: Essential Motifs in Saccharomyces cerevisiae RNA Capping Enzyme are Conserved in Schizosaccharomyces Pombe and Viral Capping Enzymes and Among Polynucleotide Ligases. Proc. Natl. Acad. Sci. USA **91:**12046-12050.

96. **Shuman, S., and Hurwitz, J.** 1981. Mechanism of mRNA Capping by Vaccinia Virus Guanylyltransferase: Characterization of an Enzyme-Guanylate Intermediate. Proc. Natl. Acad. Sci. USA **78:**187-191.

97. **Shuman, S., and Ru, X-M.** 1995. Mutational Analysis of Vaccinia DNA Ligase Defines Residues Essential for Covalent Catalysis. Virology **211:**73-83.

98. **Silber, R., Malathi, VG, and Hurwitz, J.** 1972. Purification and Properties of Bacteriophage T4 RNA Ligase. Proc. Natl. Acad. Sci. USA **69:**3009-3013.

99. **Snopek, T., Wood, WB, Conley, MP, Chen, P, and Cozzarelli, NR.** 1977. Bacteriophage T4 RNA Ligase is gene 63 Product, the Protein that Promotes Tail Fiber Attachment to the Baseplate. Proc. Natl. Acad. Sci. USA **74:**3355-3359.

100. **Snopek, T., Sugino, A, Agarwal, K, and Cozzarelli, NR.** 1976. Catalysis of DNA Joining by Bacteriophage T4 RNA Ligase. Bioch. Bioph. Res. Comm. **68:**417-424.

101. **Steitz, T., and Steitz, JA.** 1993. A General 2-Metal-Ion Mechanism for Catalytic RNA. Proc. Natl. Acad. Sci. USA **90:**6498-6502.

102. **Sugimoto, N., Matsumura, A, Hasegawa, K, and Sasaki, M.** 1991. Effect of Unpaired Nucleotides of Substrate RNAs on the RNA Ligation with T4 RNA Ligase. Bull. Chem. Soc. Jpn. **64:**2978-2982.

103. **Sugino, A., Snopek, TJ, and Cozzarelli, NR.** 1977. Bacteriophage T4 Ligase: Reaction Intermediate and Interaction of Substrates. J Biol. Chem. **252:**1732-1738.

104. **Sugira, A., Hirokawa, K, Nakashima, K, and Mizuno, T.** 1994. Signal-sensing Mechanisms of the Putative Osmosensor KdpD in Escherichia coli. Mol. Microb. **14:**929-938.

105. **Tanner, N., Hanna, MM, and Abelson, J.** 1988. Binding Interactions Between Yeast tRNA Ligase and a Precursor Transfer Ribonucleic Acid Containing Two Photoreactive Uridine Analogs. Bioch. **27:**8852-8861.

106. **Thogersen, H., Morris, HR, Rand, KN, and Gait, MJ.** 1985. Location of the Adenylylation Site in T4 RNA Ligase. Eur. J. Biochem. **147:**325-329.

107. **Thompson, L., Brandon, LD, Nieuwlandt, DT, and Daniels, CJ.** 1989. Transfer RNA Intron Processing in the Halophilic Archaebacteria. Can. J Microb. **35:**36-42.

108. **Trujillo, M., Roux, D, Fueri, JP, Cailla, HL, and Rickenberg, HV.** 1987. The Occurance of 2'-5' Oligoadenylates in Escherichia coli. Eur. J Biochem. **169:**167-173.

109. **Tsagris, M., Tabler, M, and Sanger, HL.** 1991. Ribonuclease T1 Generates Circular RNA Molecules From Viroid-Specific RNA Transcripts by Cleavage and Intermolecular Ligation. Nuc. Acids Res. **19:**1605-1611.

110. **Tyc, K., Kikuchi, Y, Konarska, M, Filipowicz, W,and Gross, HJ.** 1983. Ligation of Endogenous tRNA 3' Half Molecules to Their Corresponding 5'

Halves via 2'-Phosphomonoester, 3',5'-Phosphodiester Bonds in Extracts of Chlamydomonas. EMBO J **2:**605-610.
111. **Uhlenbeck, O., and Gumport, RI.** 1982. T4 RNA Ligase, p. 31-60. *In* P. Boyer (ed.), The Enzymes, 3rd ed, vol. 15B. Academic Press.
112. **van Tol, H., Stange, N, Gross, HJ, and Beier, H.** 1987. A Human and a Plant Intron-containing tRNATyr Gene are Both Transcribed in a HeLa Cell Extract but Spliced Along Different Pathways. EMBO J **6:**35-41.
113. **VanBogelen, R., Kelley, PM, and Neidhardt, FC.** 1987. Differential Induction of Heat Shock, SOS, and Oxidation Stress Regulons and Accumulation of Nucleotides in Escherichia coli. J Bact. **169:**26-32.
114. **Walker, G., Uhlenbeck, OC, Bedows, E, and Gumport, RI.** 1975. T4-Induced RNA Ligase Joins Single-Stranded Oligoribonucleotides. Proc. Natl. Acad. Sci. USA **72:**122-126.
115. **Wallace, R., Johnson, PF, Tanaka, S, Schold, M, Itakura, K, and Abelson, J.** 1980. Directed Deletion of a Yeast Transfer RNA Intervening Sequence. Science **209:**1396-1400.
116. **Ward, S., Luftig, RB, Wilson, JH, Eddleman, H, Lyle, H, and Wood, WB.** 1970. Assembly of Bacteriophage Tail Fibers II. Isolation and Characterization of Tail Fiber Precursors. J Mol. Biol. **54:**15-31.
117. **Webber, T.** 1985. U. Illinois, Champaign-Urbana.
118. **Weiner, L., and Model, P.** 1994. Role of an Escherichia coli Stress-Response Operon in Stationary-Phase Survival. Proc. Natl. Acad. Sci. USA **91:**2191-2195.
119. **Westaway, S., Phizicky, EM, Abelson, J.** 1988. Structure and Function of the Yeast tRNA Ligase Gene. J Biol. Chem. **263:**3171-3176.
120. **White, T., and Borst, P.** 1987. RNA End-labeling and RNA Ligase Activities can Produce a Circular rRNA in Whole Cell Extracts from Trypanosomes. Nuc. Acids Res. **15:**3275-3290.
121. **Winicov, I., and Button, JD.** 1981. Nuclear Ligation of 5'-OH Kinase Products in tRNA. Mol. Cell. Biol. **2:**241-249.
122. **Wray, W., Boulikas, T, Wray, VP, and Hancock, R.** 1981. Silver Staining of Proteins in Polyacrylamide Gels. Anal. Biochem. **118:**197-203.
123. **Xu, Q., Teplow, D, Lee, TD, and Abelson, J.** 1990. Domain Structure in Yeast tRNA Ligase. Bioch. **29:**6132-6138.
124. **Zagrebel'nii, S., and Zernov, YP.** 1987. Mechanism of the Reaction Catalyzed by RNA Ligase of Bacteriophage T4. Mol. Biol. Mos. **21:**334-340.
125. **Zillman, M., Gorovsky, MA, and Phizicky, EM.** 1991. Conserved Mechanism of tRNA Splicing in Eukaryotes. Mol. Cell. Biol. **11:**5410-5416.