STRUCTURE AND FUNCTION OF YEAST tRNA LIGASE

Thesis by
Shawn Kathleen Westaway

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1993
(Submitted December 17, 1992)

ACKNOWLEDGMENTS

I wish to acknowledge Dr. Jones, retired science teacher of North Hollywood High School, who piqued my scientific curiosity by carefully shaping my high school curriculum. I am grateful to C.S.U.N. biology professors Joyce Maxwell and Ken Jones, who revealed a new world of discovery for me in molecular biology and gave me a nudge in just the "right" direction. I owe a great deal to Jennifer Normanly and Marge Strobel for my initial training in molecular biology. Both were patient beyond sainthood with my continual questions. I would not be where I am today if not for Eric Phizicky, my closest collaborator, mentor, and friend during my graduate tenure. He exemplifies the scrupulous scientist and taught me that some procedures need more attention than others. I would also like to acknowledge Elizabeth Grayhack, Michael Bruist, and Michael Clark for many helpful discussions about my research and for supplementing my graduate student stipend with frequent dining out. Special thanks to labmates Kyle Tanner, Reinhard Rauhut, and David McPheeters for their individual contributions to my scientific growth, and to George Komatsoulis, dedicated to safeguarding my sense of humor and keeping me computer literate. I also wish to acknowledge Elliot Altman for sharing his experience as a graduate teaching assistant, Peggy Saks for her kind and timely critiques of my meager writing ability, and Jeff Sampson for his unflagging interest in my project. Nancy Gill, Patricia Bateman, Mary Marsh, and Jeannette Kalpakoff are to be commended for their efforts to keep my graduate career financially and bureaucratically in good order. It would be remiss not to thank Martha Fiallos, the kitchen staff, and Joyce Kato for keeping me in supplies and coordinating the technical details of advisor relations. I have been enriched for having had the opportunity to work with Barbara Apostol, Heather Belford (with whom I spent many a long week in the throes of DHFR-ligase protein purifications), and the rest of Chris Greer's laboratory. Chris Greer has an enlightened and curious scientific mind and I am indebted to him for his understanding, support, and friendship.

I would like to include my loving and supportive family in my acknowledgments - my brother Dave and aunt Margaret, and my parents, Jeanne and Dave, all who have made heroic attempts to understand just exactly what I was studying. Their love, moral and financial support have been a great strength to me. Most of all, my friend Eilien Van Patten deserves a medal for her selfless devotion, generosity, and unfailing patience during the rigors of my graduate stint.

Finally, I am most grateful to my advisor, John Abelson, for his energetic support and direction in my research. His unparalleled passion for science is contagious and his charisma and congeniality are equally delightful. It has been a pleasure working with him.

ABSTRACT

The gene for yeast tRNA ligase has been sequenced and its transcription start sites have been mapped. Three other open reading frames in the vicinity of the tRNA ligase gene were characterized. One open reading frame, ORF4, is the yeast *ARG3* gene. ORF1 is probably not transcribed or translated in yeast. ORF2 is an unidentified but essential gene in yeast.

A deletion of the central 200 amino acids has been engineered in the ligase protein. This deletion protein, designated DAC, was characterized in the *in vitro* tRNA splicing reaction with regard to the structure of the joined tRNA product. Cofactor requirements for tRNA joining activity and polynucleotide kinase activity were also determined. DAC possesses a GTP-dependent joining activity that is not manifested by wild-type ligase. In addition, both the wild-type and DAC proteins exhibit polynucleotide kinase activities that are more efficient with GTP than with ATP. Joining reactions with wild-type ligase indicate that joining of tRNA halves is more efficient in the presence of both GTP and ATP than with either cofactor alone. Wild-type tRNA ligase can incorporate the γ-phosphate of GTP into the splice junction of joined tRNA, but only when ATP is also provided. The ligase protein contains two distinct nucleotide triphosphate binding sites - one specific for GTP and one specific for ATP. A revised mechanism for tRNA splicing in yeast is presented.

TABLE OF CONTENTS

INTRODUCTION

tRNA Splicing in Yeast

INTRODUCTION

Transfer RNAs, or tRNAs, have long been studied as a means of understanding the basic translation mechanism for cellular proteins, as they are responsible for interacting with the ribosome and the messenger RNA to attach amino acids to the growing polypeptide chain. Goodman *et al.* (1) and Valenzuela *et al.* (2) independently discovered that genes coding for functional yeast tyrosine and phenylalanine tRNAs, respectively, contained a sequence within the DNA called an intron which was not present in the mature functional tRNA transcript. These introns would have to be removed from the precursor tRNA (pre-tRNA) in order to form the functional product; a process termed tRNA splicing (for a review, see 3). Many other steps are involved in the maturation of pre-tRNAs, including removal of the 5' leader and 3' trailing sequences (for a review, see 4), tRNA base modifications (5), and, in eukaryotes, addition of the CCA amino acid acceptor sequence to the 3' end of the tRNA (4). The subject of this discussion will be based primarily on tRNA splicing in yeast - the enzymes involved in removing the introns to form the mature tRNA, interactions between these enzymes and their tRNA substrates and cofactors, yeast mutants which have proven to have a direct effect on the splicing enzymes *in vivo*, and comparisons to tRNA splicing systems from other organisms.

It is estimated that there are approximately 360 tRNA genes in yeast (6). Of the predicted 46 different tRNA species, ten isoacceptors are now known to contain introns. These isoacceptors are represented by a documented minimum of 45 gene copies. If there is an average of eight gene copies per tRNA species (6), then approximately 80 tRNA genes, or about one-fifth of the genes in yeast, contain introns.

History of tRNA Splicing

A temperature sensitive mutant in *Saccharomyces cerevisiae*, *rna1*, accumulates precursor tRNAs as well as other RNA intermediates when shifted to the nonpermissive temperature (7). This mutation was instrumental in paving the way for future research in tRNA splicing. Knapp *et al.* (8) isolated intron-containing pre-tRNA^{Phe} and pre-tRNA^{Tyr} from *rna1* cells and used them to screen wild-type yeast extracts for an activity involved in tRNA splicing. The ribosomal wash fraction of cell-free yeast extracts, as well as yeast nuclear extracts, were found to contain activities which excised the intron and joined the two tRNA halves to form the mature tRNA (8, 9). The splicing reaction was shown to occur in two major stages, separable by the absence or presence of ATP in the *in vitro* splicing assay (10). Pre-tRNA treated with soluble yeast extract in the absence of added ATP accumulated as tRNA half-molecules and excised intron. The tRNA halves were purified and could subsequently be joined to form mature-length tRNA upon the addition of ATP. This result indicated that the joining step was not necessarily coupled to the cleavage step *in vitro* (10). Analysis of the structure of the splicing intermediates revealed the 5' half tRNA to be phosphorylated at its 3' end (11), and this structure was shown to be a 2',3'-cyclic phosphate (12). The presence of this phosphate moiety was required for ligating the two tRNA halves (11). The 3' half tRNA possessed a 5'-hydroxyl. The excised intron proved to be linear rather than circular, with a phosphate at its 3'-end and a hydroxyl at its 5'-end, identical to the cleavage termini of the tRNA halves (11). The structure of these intermediates added further support for a splicing pathway which proceeds by two separable enzymatic steps - cleavage of the precursor at the boundaries of the mature sequence to release the intron, followed by joining of the two tRNA halves (10).

Enzymes of Yeast tRNA Splicing

Yeast tRNA Splicing Endonuclease

The two steps of tRNA splicing could now be examined independently in yeast extracts in order to further purify the components for cleavage and joining. The endonuclease activity was separated from the ligase activity by nonionic detergent extraction of yeast membrane preparations, a method which indicated that the endonuclease behaves as an integral membrane protein during purification (12). This partially purified endonuclease fraction did not demonstrate any appreciable ligase activity, although a 2',3'-cyclic phosphodiesterase activity was detected (12). Though the endonuclease has previously proven refractory to purification, it has now been purified to homogeneity (13) and creates the same termini at the cleavage sites as described above. It does not possess any appreciable phosphodiesterase activity. The enzyme consists of three polypeptide subunits, α, β, and γ, of 31, 42, and 51 kDa, respectively. Glycerol gradient sedimentation and gel filtration analysis show the endonuclease exists as a heterotrimer of these three subunits.

The powerful methods of yeast genetics have been used to identify mutants defective in pre-tRNA cleavage. Two unlinked recessive mutations, *sen1-1* and *sen2-1*, were identified from a collection of temperature sensitive mutants (14) by testing yeast nuclear mini-extracts of strains grown at the permissive temperature for temperature sensitive *in vitro* pre-tRNA cleavage (15). The *sen1-1* mutation is temperature sensitive for growth, and exhibits defective endonuclease activity *in vitro* and accumulates uncleaved pre-tRNAs *in vivo* at both the permissive and nonpermissive temperatures. The *SEN1* gene has been isolated and shown to complement the *sen1-1* defect *in vivo* (16). *SEN1* is essential for growth in yeast and the sequence codes for a 239 kDa protein of the same size as the SEN1 protein detected by Western blot. The amino-terminal region of the *SEN1* gene corresponding to approximately 100 kDa of the protein can be deleted and is not essential in yeast. The *sen1-1* lesion has also been isolated and sequenced and the mutation

is a single nucleotide change resulting in substitution of an aspartic acid for a glycine residue. This lesion is located in the carboxy-terminal essential region of *SEN1* and changes a glycine residue which is conserved in a region of two other proteins with sequence similarity to *SEN1*, mouse Mov-10 protein and yeast *UPF1* protein (16). The functions of these two proteins are unknown, however *upf1* cells are involved in the messenger RNA degradation pathway in yeast (17).

The *sen2-1* mutation exhibits defective endonuclease activity *in vitro* at the nonpermissive temperature, however, it is not temperature sensitive for growth and does not accumulate pre-tRNAs at either the permissive or nonpermissive temperatures (15). Yeast extracts from a *sen1-1/sen2-1* heterozygous diploid show wild-type endonuclease activity in the *in vitro* assay at all temperatures tested, as would be expected for the complementation of two unlinked mutations. However, mixing of separate extracts from *sen1-1* and *sen2-1* strains grown at the permissive temperature do not complement the *sen2-1* thermosensitive *in vitro* phenotype. The defects in tRNA endonuclease activity of both *sen1-1* and *sen2-1* extracts cofractionate with endonuclease during partial purification, suggesting a direct enzymatic role in endonucleolytic cleavage of pre-tRNA for these two gene products.

A different collection of temperature sensitive yeast mutants (18) was screened for defective tRNA splicing by hybridizing RNA prepared from the strains grown at the nonpermissive temperature to an intron-specific probe (19). This assay revealed a strain, *tsa373*, which accumulates a splicing intermediate consisting of the 5'-half tRNA still joined to the intron, termed a pre-tRNA splicing "2/3" intermediate. The tRNA endonuclease defect of *tsa373* is recessive and does not cosegregate with the original temperature sensitive phenotype. However, the growth rate of *tsa373* is reduced at low temperatures (15°C) compared to wild-type, and this pseudo-cold sensitive phenotype is linked to the *in vivo* accumulation of "2/3" molecules.

A *tsa373/sen2-1* heterozygous diploid displays the *in vivo* accumulation of the "2/3" molecule as well as the *in vitro* thermosensitive splicing defect of the *sen2-1* allele. This result suggests that the two genes are closely linked and both mutations are presumed to be allelic. Thus *tsa373* has been designated *sen2-3* as the third identified allele of *SEN2*. The *sen2-3* strain contains a substantial amount of spliced $tRNA^{Tyr}$ as well as the "2/3" intermediate. It was concluded that the lesion probably does not completely block endonucleolytic cleavage at the 5'-splice site, and allows enough complete splicing activity to permit growth at normal temperatures (19).

Yeast tRNA splicing endonuclease was purified from the *sen2-3* strain and shown to have identical chromatographic properties as the wild-type enzyme (19). The *SEN2* gene has been isolated and complements the *sen2-3* defect *in vivo*, and immunological studies have indicated that the *SEN2* gene codes for the 42 kDa β-subunit of tRNA splicing endonuclease (unpublished results, cited in 13).

Since antibodies to the SEN1 protein do not crossreact with purified endonuclease, it is unlikely that *SEN1* codes for either the α- or γ-subunits of endonuclease. Current experiments are addressing the issue of whether the SEN1 protein interacts directly with the endonuclease in a splicing complex, or is a positive effector of other cellular processes affecting endonuclease production or localization (16).

Yeast tRNA Ligase

The separation and partial purification of the tRNA ligase responsible for the second step of tRNA splicing was accomplished by a high salt extraction of a soluble fraction of yeast extract followed by several chromatographic steps (20). A polynucleotide kinase activity, a cyclic phosphodiesterase activity, and the ligase activity all copurified with a protein which could be adenylylated with [α-^{32}P]ATP. Adenylylated protein releases its AMP under conditions where splicing is allowed to proceed (i.e., adding pre-tRNA half substrates). The ligation product contains a 2'-phosphomonoester, 3',5'-phosphodiester linkage at the

splice junction. The externally-derived γ-phosphate of ATP was incorporated at the 3'-5' phosphodiester bond of this splice junction (20).

A model was proposed for tRNA splicing which addresses the above enzyme activities, and considers the observed structure of the termini of the two tRNA halves created by endonuclease cleavage (20). The 2',3'-cyclic phosphate of the 5'-half tRNA is opened to a 2'-phosphate by the 2',3'-cyclic phosphodiesterase activity; the 5'-hydroxyl of the 3'-half tRNA is phosphorylated by the polynucleotide kinase activity; the protein is adenylylated and transfers its AMP residue to the 5'-phosphate of the 3'-half tRNA; and ligase joins the two tRNA halves with the concomitant release of AMP. Final tRNA products still retain a 2'-phosphate at the splice junction.

Yeast tRNA ligase was further purified to homogeneity using tRNA affinity elution from heparin-agarose, and antibody was prepared against the protein (21). The protein is a single polypeptide of approximately 90 kDa and exhibits all the activities proposed for ligase in the splicing model. An intermediate of the splicing reaction containing an AMP residue covalently linked to the 5'-phosphate of the 3'-half tRNA was also detected during assays for these activities. This was expected from the predicted splicing mechanism.

The amino-terminal sequence of the purified protein was obtained and the gene was isolated by screening a yeast DNA library for this sequence (21). The gene was placed under control of the *GAL10* promoter region of yeast, and a yeast strain harboring this plasmid overproduces tRNA ligase activity 25-fold over wild-type when grown on galactose. In addition, tRNA ligase can be overproduced in *E. coli* when expressed from a vector with an *E. coli* promoter. The gene for yeast tRNA ligase (*RLG1*) was sequenced and the *in vivo* transcription start sites were mapped (22). The gene is uninterrupted by intervening sequences and codes for an 827 amino acid basic protein of approximately 95 kDa.

Protease resistant fragments of the tRNA ligase polypeptide were detected during purification from *E. coli* (23). Three of these fragments were shown to contain only some

of the activities of intact tRNA ligase (24). These fragments, of 65, 40, and 25 kDa, were further purified and amino-terminal protein sequence data were obtained. Both the 65 kDa and 40 kDa fragments are from the amino-terminus of ligase, and the 25 kDa fragment is C-terminal. Only the 25 kDa fragment still retains 2',3'-cyclic phosphodiesterase activity. Both the 40 kDa and 65 kDa fragments can be adenylylated. Protease digestion of an adenylylated intact ligase polypeptide followed by fractionation of the peptides and mass spectrometry identified lysine-114 as the site of adenylylation (24). The region surrounding this lysine-114 is similar in sequence to the region surrounding the lysine-99 of T4 RNA ligase implicated in the phage T4 ligase adenylylation reaction (24-26).

These results suggest that the tRNA ligase protein consists of distinct enzymatic domains, each responsible for a specific activity involved in the joining step of tRNA splicing. A structural and functional analysis of tRNA ligase was undertaken to attempt to define these domains and the activities they exhibit. A set of ligase gene deletions was fused to *E. coli* dihydrofolate reductase to facilitate protein purification from *E. coli* (25). Detailed enzymatic analyses of these purified deletion polypeptides confirmed some previous results as well as revealing new information. 1) An approximately 43 kDa amino-terminal fragment of ligase is sufficient for adenylylation of the protein. 2) An approximately 28 kDa C-terminal fragment is sufficient for cyclic phosphodiesterase activity. 3) Pre-tRNA substrate binding could not be directly correlated with a single domain, but rather appeared to be associated with any two of the domains, implying that the substrate binding site might consist of more than one component along the linear peptide sequence. 4) Polynucleotide kinase activity could only be demonstrated for a large fragment containing the C-terminal 431 amino acids of ligase. 5) Surprisingly, a deletion construct missing the center 200 amino acids of ligase (designated "DAC") was able to join tRNA halves efficiently *in vitro* to form mature tRNA of the correct structure (25). This centrally-deleted ligase construct demonstrated the ability to join tRNA halves in the presence of GTP alone, a novel activity not found in wild-type ligase (S. K. Westaway,

H. G. Belford, B. Apostol, J. Abelson, and C. Greer, Journal of Biological Chemistry, in press). This result led to the discovery that both DAC and the full-length wild-type ligase fusion possessed a GTP-dependent polynucleotide kinase activity. Further experimentation demonstrated that the full-length DHFR/tRNA ligase protein incorporated radiolabel from [γ-^{32}P]GTP into the 3',5'-phosphodiester bond at the splice junction of tRNA only when incubated with ATP. A modified mechanism for tRNA splicing was proposed in which wild-type tRNA ligase would preferentially use GTP for the polynucleotide kinase step and reserve ATP for the subsequent adenylylation and ligation steps. This mechanism is supported by earlier results that the tRNA joining activity required two molar equivalents of ATP (Greer *et al.*, 1983), and by the results of nucleotide binding and competition assays. The competition assays demonstrated the presence of two distinct NTP binding sites in the ligase protein - one specific for GTP and one specific for ATP (H. G. Belford, S. K. Westaway, J. Abelson, and C. L. Greer, Journal of Biological Chemistry, in press).

The deletion construct described above (DAC) is capable of joining tRNA halves *in vitro*, but has not been shown to complement a ligase-deficient yeast strain *in vivo* (Mark Berlin, personal communication). This result is supported by previous results that either a large insertion (27) or a large deletion (28) in the ligase gene is lethal to the cell. Yeast tRNA ligase is a single copy gene and is essential in yeast, as would be expected for an enzyme directly involved in the splicing of essential tRNAs.

New genetic studies have uncovered two temperature sensitive yeast mutants in the tRNA ligase gene, designated *rlg1-4* and *rlg1-10* (28). Both mutant genes have been isolated and sequenced and code for a single amino acid substitution within the amino-terminal region of ligase. These lesions probably confer the temperature sensitive phenotype. As expected, neither of these temperature sensitive mutations are at the lysine-114 adenylylation site. Any amino acid substituted at lysine-114 (other than perhaps arginine) should result in a protein incapable of adenylylation and therefore unable to complete the tRNA joining process at any temperature. Phizicky *et al.* (28) discovered that

tRNA half molecules and introns accumulate not only in *rlg1-4* and *rlg1-10* when shifted to the nonpermissive temperature, but also in a strain conditionally expressing tRNA ligase when grown under nonexpressing conditions. This is expected for a mutation in tRNA ligase, as tRNA splicing endonuclease should still be fully functional in cleaving pre-tRNAs, and uncut pre-tRNAs should not accumulate.

The versatility of mutagenesis in the genes for tRNA ligase (*RLG1*) and the β-subunit of endonuclease (*SEN2*) encourages further examination of potential mutant proteins and their structural and enzymatic properties *in vitro*. Possession of both tRNA splicing endonuclease and tRNA ligase point mutants also permits a search for extragenic *in vivo* suppressors of these mutations, leading perhaps to the elucidation of other elements involved in the organization of nuclear tRNA splicing in yeast.

NAD-Dependent 2'-Phosphotransferase

As previously described, ligation of the two tRNA halves by purified yeast tRNA ligase results in a 2'-phosphomonoester, 3',5'-phosphodiester bond at the splice junction of joined tRNA. Removal of this 2'-phosphate must occur subsequent to tRNA joining, as mature yeast tRNAs do not contain a 2'-phosphate at the splice junction. An activity which specifically removes this internal 2'-phosphate was partially purified from yeast extracts by chromatography followed by reconstitution of specific flowthrough fractions with eluted fractions (29). This result indicated that the 2'-phosphatase activity was composed of two separate components of yeast extracts.

The 2'-phosphatase activity prefers an authentic tRNA substrate over an eight-nucleotide oligomer with an internal 2'-phosphate, and has no phosphatase activity on 3'- or 5'-phosphates (29). This result is opposite to that obtained with a nonspecific phosphatase. Experiments imply that the 2'-phosphatase, like the tRNA ligase and endonuclease, is specific for a subset of tRNAs, in this case those containing an internal 2'-phosphate. Further purification of component I and II of this phosphatase activity has

revealed that component I is most likely NAD^+ (30). Component I is highly heat stable and resistant to micrococcal nuclease, inferring that it is neither protein nor RNA. Gel filtration indicates that component I is a small molecule close in size to ATP, and experiments with the phosphatase reaction show it is not a divalent cation. Only NAD^+ can substitute effectively for component I in the 2'-phosphatase assay, and component I cochromatographs with NAD^+ on a C_{18} column.

Interestingly, the 2'-phosphate from the splice junction is not released as free phosphate, as might be predicted for a solitary protein carrying out a phosphate hydrolysis reaction. Instead, the 2'-phosphate is transferred to an acceptor molecule, whose chromatographic properties differ from $NADP^+$, a product which would be expected from a phosphotransfer reaction involving NAD^+. It has also been shown that the partially purified phosphatase activity does not have the properties of NAD^+ kinase. McCraith and Phizicky (30) speculate that the acceptor molecule may be NAD^+ with the phosphate transferred to a position other than the 2'-hydroxyl of NAD^+. Recently, a similar NAD^+-dependent 2'-phosphotransferase activity has been discovered in extracts from both HeLa cells and *Xenopus* oocytes (31, 32).

Removal of the 2'-phosphate from the ligated tRNA is the last step specific to the tRNA splicing pathway of intron-containing pre-tRNAs. Since the splice junction of mature tRNA is located only one base 3' to the anticodon in mature tRNA, the presence of a 2'-phosphate in an incompletely processed tRNA might sterically interfere with mature tRNA function *in vivo*. This interference might involve preventing essential post-splicing modifications of nucleotides in or near the tRNA anticodon, or inhibiting the association of the tRNA with either the cognate aminoacyl tRNA synthetase or the ribosomal translation machinery. For this reason, it is conceivable that genetic analysis of yeast mutants might reveal a mutation in component II of the 2'-phosphotransferase activity which would accumulate spliced tRNA still retaining the 2'-phosphate at the splice junction. It will be interesting to follow future discoveries of the identities of the phosphate-acceptor molecule

and of component II, and how these factors interact with the rest of the cellular splicing machinery.

Protein-Protein Interactions Among tRNA Splicing Enzymes

During the original assay for splicing activities, other important features of precursor tRNA processing began to emerge. Knapp *et al.* (8) showed that intron-containing pre-tRNAs which accumulated in the *rna1* strain at the nonpermissive temperature possessed some modified nucleosides, while other base modifications were absent. In addition, none of the pre-tRNAs still retained 5' leader sequences, and all had 3'-end CCA termini. Apparently, 5' and 3' end processing had already occurred as well as post-transcriptional addition of the uncoded CCA acceptor. This result suggested that there was a temporal order for processing of pre-tRNAs, and that splicing was a late step in the processing pathway.

Difficulty in fractionating nuclei from cytoplasm in the yeast system has complicated the unequivocal assignment of a cellular location for the three enzymes involved in tRNA splicing. However, early *in vivo* studies in *Xenopus* oocytes provided compelling evidence that the splicing machinery was located in the frog nucleus. By injecting yeast $tDNA^{Tyr}$ into both nuclear and cytoplasmic compartments of *Xenopus* oocytes, a generic order was defined for the steps in tRNA processing, and it was demonstrated that removal of introns occurs in the nucleus, at least in *Xenopus* (33). The consecutive order for processing nuclear tRNA transcripts in *Xenopus* appears to be: 1) 5' leader trimming accompanied by some base modifications; 2) removal of the 5' leader, the 3' trailer, and addition of the CCA terminus, accompanied by further base modifications; 3) removal of the intron and splicing of the two tRNA halves; and 4) transport of the mature $tRNA^{Tyr}$ from the nucleus to the cytoplasm, accompanied by final base modifications (33). Though the splicing step occurs late in the maturation pathway, these

results suggest it is still completed in the nucleus prior to transport of the tRNA to the cytoplasm.

The nuclei in *Xenopus* oocytes have been fractionated into nuclear envelope and nucleoplasm, and the *Xenopus* endonuclease activity is associated only with the soluble fraction (34). This nuclear fractionation has not been accomplished in yeast. However, in contrast to *Xenopus* fractions, yeast tRNA splicing endonuclease behaves as an integral membrane protein during purification (12).

The ability to separate the enzymatic steps of tRNA splicing *in vitro* and the observation that pre-tRNA splicing intermediates can be detected in normal yeast cells (35) might lead to the assumption that the splicing enzymes act independently *in vivo*. Evidence to the contrary emerged from *in vitro* competition assays with yeast tRNA splicing endonuclease, yeast tRNA ligase, T4 RNA ligase, and T4 polynucleotide kinase (36). The T4 enzymes are able to ligate tRNA halves produced by yeast tRNA splicing endonuclease, but the product tRNA has a different structure from that of the yeast ligase reaction (37). Analysis of product ratios in competition assays revealed that yeast ligase has preferential access to the products of endonuclease cleavage (36). This preference is evidence for the assembly of a splicing complex of ligase and endonuclease and/or pre-tRNA *in vitro*. This result is curious in that it suggests that other proteins may not be necessary *in vivo* for the assembly of a concerted splicing apparatus.

The purification of tRNA ligase and preparation of anti-tRNA ligase antibody have enabled cytological studies to determine the cellular location of tRNA ligase. Immune fluorescence and immune electron microscopy have localized tRNA ligase to the inner membrane of the nuclear envelope, possibly associated with nuclear pores (38). The ligase protein also appears to be positioned at a second location, 100-200 nm into the nucleoplasm. It has also been shown that tRNA ligase binds precursor tRNA *in vitro* (39). Yeast ligase can be crosslinked to pre-tRNA (40). The sites of one minor and two major crosslinks occur in the intron of the pre-tRNA, either near the anticodon-intron stem or

surrounding the single-stranded region of the 3'-splice site. Based on these results, a spatial model for splicing complex formation *in vivo* was proposed (38, 40; Chris Greer, personal communication). The tRNA ligase at its intranucleoplasmic location might bind 5'- and 3'-processed intron-containing pre-tRNA, move to the nuclear envelope acting as a substrate "shuttle," form a complex with the pre-tRNA, endonuclease, and other splicing machinery, and allow splicing to proceed. Upon completion of the ligation reaction, tRNA product would be released, presumably to be transported to the cytoplasm through the nuclear pore. Ligase would also be released from the complex to return to its intranuclear depot. This model should involve some sort of conformational transition in either tRNA ligase polypeptide or in the endonuclease subunit which may interact with ligase, in order to dissociate upon completion of the splicing reaction.

A yeast strain containing a large deletion (457 amino acids) of the tRNA ligase gene accumulates tRNA halves when production of the wild-type plasmid copy of ligase is conditionally repressed (28). If tRNA ligase were the only mechanism by which pre-tRNAs could be transported to the splicing apparatus at the nuclear envelope, then one might at first expect that such a large deletion in the protein would render the enzyme incapable of its proposed pre-tRNA shuttling function, as well as its ligase function. This should result in the accumulation of precursor tRNAs, and not tRNA halves. However, several different large deletions of the tRNA ligase protein are still competent at pre-tRNA binding *in vitro* (25). Therefore, even a large deletion of tRNA ligase may still be capable of functions other than the final ligation step *in vivo*.

A physical association between tRNA ligase and endonuclease has not yet been emphatically demonstrated. However, ATP-independent *in vitro* stimulation of endonuclease activity by the presence of tRNA ligase and preferential access of ligase over T4 enzymes to tRNA halves are good evidence for a complex of the enzymes *in vivo*. In this respect, it will be interesting to discover which of the three endonuclease subunits and what region of tRNA ligase may be involved in endonuclease-ligase interaction. The yeast

NAD^+-dependent 2'-phosphotransferase activity has been shown to be specific for internal 2'-phosphates, and would not be expected to act on the phosphate termini of unligated tRNA halves. This activity would interfere with the splicing reaction. Therefore it is conceivable that the dephosphorylating enzyme also resides near the nuclear membrane or nuclear pore in a complex with endonuclease or ligase, awaiting its final role in the tRNA splicing pathway (32).

tRNA Splicing Substrates and Interactions with the Splicing Enzymes

Ten isoacceptor species of pre-tRNA in yeast contain introns, comprising 45-80 gene copies. These tRNAs are listed in Table 1 (data compiled from 6, 41-43).

Table I. List of Yeast tRNA Precursors Containing Introns

tRNA[a]	Intron (nt)	# Genes
$tRNA^{Ser}_{GCU}$	19	3
$tRNA^{Tyr}$	14	8
$tRNA^{Phe}$	18, 19	10
$tRNA^{Lys}_{UUU}$	23	N.D.[b]
$tRNA^{Pro}_{UGG}$	28, 31	5
$tRNA^{Trp}$	34	≥4
$tRNA^{Ser}_{CGA}$	19	1
$tRNA^{Leu}_{UAG}$	19	N.D.[b]
$tRNA^{Leu}_{CAA}$	32, 33	≥10
$tRNA^{Ile}_{UAU}$	60	2

[a]tRNAs are identified by the unmodified sequence of their anticodons.
[b]N.D., not determined.

It is evident that since many species of yeast precursor tRNAs do not contain introns, the presence of an intron cannot be required for other tRNA processing steps. However, the observation that yeast tRNA ligase is an essential gene in yeast is supported by the fact that there has been only one isoaccepting species found for each of $tRNA^{Phe}$, $tRNA^{Trp}$, and $tRNA^{Tyr}$, and all genes for these tRNAs contain introns. In addition, a minor but nonetheless unique species of serine-accepting tRNA ($tRNA^{Ser}_{CGA}$) is single copy in yeast and contains an intron.

It is clear from Table I that pre-tRNA introns differ in size, however, lengths and nucleotide sequences are very similar among tRNAs within a single family. The sequences of introns show no similarity between families, but it is important to note that they are all positioned alike within the precursor (i.e., one base downstream of the anticodon (41)).

The next logical step was to examine how a subset of precursor tRNAs with introns might interact with splicing enzymes. These enzymes should normally be outside the pathway of pre-tRNA processing for intronless tRNAs. Experiments addressing this issue have largely centered on testing the fidelity of splicing using altered tRNA substrates. It is beyond the scope of this discussion to detail all the substrates which have been tested (for a more comprehensive review, see 44). However, general points can be made to present an emerging picture of the recognition signals required by the substrate for accurate splicing.

Introns in pre-tRNAs are not conserved between families, and only a limited set of base positions in the mature domain of tRNAs are conserved among the different isoacceptors (45). This is in contrast to the consensus sequences found in messenger RNA introns, which are necessary for recognition by the mRNA splicing machinery (for a review, see 3). Previous evidence has suggested that the intron plays little or no part, beyond its mere presence, in the recognition by the splicing enzymes (46). An unstructured poly(U) intron variant of $pre\text{-}tRNA^{Phe}$ is accurately spliced *in vitro* (47). The common theme from pre-tRNA splicing studies is that the secondary and tertiary structure

of the mature domain of the precursor tRNA is recognized by the splicing enzymes. Some evidence supporting this model and some exceptions to the rule are noted below.

Cleavage by endonuclease at the 5'-splice site of the intron has been shown to be dependent upon the presence of a purine residue 5' to the cleavage site (47). Cleavage by endonuclease at the 3'-splice site may be more dependent on structure than absolute sequence. A cut at this site requires the region to be in a single-stranded accessible conformation, and not base paired in a helix (48, 49). This result is supported by the conservation of this single-stranded structure in the predicted secondary conformations of all intron-containing pre-tRNAs (41, 43).

A single base change at an important site in the mature domain of precursor tRNA can block splicing completely (see below). Until recently, such small changes in the intron have had little effect, though there are some exceptions (49). New experiments suggest that most normal intron-containing pre-tRNAs possess a tertiary base pair interaction between a pyrimidine in the anticodon loop and a purine in the intron (50). This tertiary interaction is required for accurate cleavage by the *Xenopus* splicing endonuclease.

One might expect that pre-tRNAs altered in their introns would only affect endonuclease cleavage, as the resulting tRNA halves should be normal ligase and 2'-phosphotransferase substrates. This assumption may be premature if the ligase plays a role in recognition and delivery of uncut pre-tRNAs to the endonuclease. As with the *rna1* phenotype *in vivo*, there are several yeast mutants which accumulate all or a subset of intron-containing precursor tRNAs under nonpermissive conditions, and these mutants are not all direct protein catalysts of the splicing steps (for a review, see 44).

Certain alterations in the mature domain of tRNA have deleterious effects on endonucleolytic cleavage or ligation. Disruption of predicted tertiary interactions (51), single nucleotide changes in primary sequence (47, 52), lack of a proper D-stem (53, 54), disruption of the anticodon-intron helix (55-57), or even improper conformation of a precursor (58) have all been shown to affect splicing. Mutations in the mature domain

which lengthen the anticodon stem are inaccurately spliced (47, 59). This result coupled with the invariant location of the intron in all pre-tRNAs led to the proposal of a model where the endonuclease chooses its splice sites by a "ruler" mechanism, binding the precursor tRNA and measuring a specified distance to where the cuts should be made (47, 52).

Structure of the mature domain of pre-tRNAs containing introns is similar to mature tRNA structure (42, 60). The secondary and tertiary structure surrounding the splice sites and the presence of an anticodon-intron helix presents a means for the splicing enzymes to distinguish between intron-containing pre-tRNA substrates, mature spliced counterparts, and intronless pre-tRNAs, while still retaining the ability to recognize a tRNA structure common to all ten intron-containing tRNA species.

tRNA Splicing in Related Systems

Precursor tRNA splicing has been studied in a variety of organisms besides yeast. The best characterized systems are wheat, *Xenopus*, HeLa cells, and phage T4. The mechanism of nuclear tRNA splicing differs radically from other known classes of RNA splicing. Recent reviews discuss tRNA splicing among different organisms and implications on the evolutionary origin of this unique reaction (44, E. M. Phizicky and C. L. Greer, Trends in Biochemical Sciences, in press). One could almost wish for the complication of a system in which mitochondrial tRNAs might contain introns in need of a splicing apparatus. Such a system might have provided other clues to the evolutionary origin of tRNA splicing in diverse systems. As yet, no intron-containing mitochondrial tRNAs have been found. However, there are two enzymes in phage T4 capable of tRNA splicing, though its bacterial host does not have tRNA genes containing introns.

The ligation of yeast tRNA half molecules *in vitro* requires both T4 polynucleotide kinase and T4 RNA ligase to substitute for the single yeast ligase protein. The mechanism

of joining by the T4 enzymes is slightly different. The phosphatase activity of T4 polynucleotide kinase removes the phosphate of the 5'-half tRNA prior to ligation, resulting in a spliced tRNA lacking the 2'-phosphate at the splice junction (37). Otherwise, the mechanisms of the two ligases are remarkably similar. Both require adenylylation of the protein at a lysine residue and transfer of the AMP to the 3'-half tRNA. Both the T4 polynucleotide kinase and T4 RNA ligase genes have limited sequence similarities to yeast tRNA ligase (24, 25). It is not clear what functions the T4 enzymes have *in vivo*, or what evolutionary relationship, if any, they have with the yeast tRNA ligase splicing mechanism.

Both wheat germ extracts (61, 62) and *Chlamydomonas* extracts (63) also contain tRNA splicing activities, and the mechanisms are the same used by the yeast system. Wheat germ RNA ligase has a much broader substrate range than yeast ligase, which prefers cognate tRNA halves in the *in vitro* splicing assay (21).

In higher eukaryotes a different tRNA splicing mechanism was initially discovered. Both *Xenopus laevis* and HeLa cell extracts possess tRNA splicing activity with yeast pre-tRNA substrates. However, the major tRNA product of these reactions did not contain a 2'-phosphate at the splice junction. Moreover, the phosphate moiety of the 5'-half tRNA was retained as the 3',5'-splice junction phosphate in tRNA product (64-66). This is in contrast to the yeast ligase which incorporates the γ-phosphate of ATP at the splice junction. Filipowicz and Vicente (67) have characterized a 3'-phosphate cyclase from HeLa cells but its *in vivo* function is still unknown.

The differences in ligation mechanisms between the yeast/wheat germ and *Xenopus*/human systems leads one to speculate that the mechanism of tRNA splicing is not conserved among lower and higher eukaryotes. More recent evidence indicates that the mechanism common to yeast and wheat germ exists in human cells as well. Zillmann *et al.* (31) have discovered a ligase activity in HeLa cell extracts that can incorporate the γ-phosphate of ATP into the 3',5'-splice junction of yeast $tRNA^{Phe}$. The splice junction is RNase T1-resistant due to the presence of a 2'-phosphate. This product is more than

reminiscent of the yeast ligase mechanism, though it has previously gone undetected in similar experiments. Furthermore, a yeast-like NAD^+-dependent 2'-phosphotransferase has been discovered in both HeLa cells and *Xenopus* oocytes. This activity carries out removal of the internal 2'-phosphate at the junction of spliced tRNA the same as the yeast activity (32). Admittedly, only a minority of tRNA spliced in the HeLa cell extracts ends up as the yeast-like product. However, it is suggested that the conservation of the yeast-like ligation pathway in higher eukaryotes implies an essential function in at least some part of pre-tRNA splicing (32). Two different pathways capable of *in vitro* tRNA splicing with heterologous tRNA substrates awaits further clarification regarding their true functions *in vivo*.

An important lesson can be learned from the past 15 years of research in tRNA splicing. Purification of the splicing enzymes and *in vitro* biochemical characterization led to elucidation of the two-stage pathway of cleavage followed by ligation. However, results of *in vitro* assays with artificial substrates can be misleading and should be interpreted with care. The advantages of yeast genetics cannot be ignored. In turn, analysis of genetic results in yeast can be quite complicated compared to simpler bacterial systems. An unexpected genetic result in yeast, however, can often illuminate a new process or condition that is associated *in vivo* with the problem under study. The yeast system is unique in that models can be designed based on *in vitro* results and then tested *in vivo*, and vice versa.

REFERENCES

1. Goodman, H. M., Olson, M. V., and Hall, B. D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5453-5457
2. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 190-194
3. Abelson, J. (1991) *Harvey Lect.* **85**, 1-42
4. Deutscher, M. P. (1984) *CRC Critical Rev. in Biochem.* **17**, 45-71
5. Bjork, G. R., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jönsson, Y. H., and Wikström, P. M. (1987) *Ann. Rev. Biochem.* **56**, 263-287
6. Guthrie, C., and Abelson, J. (1982) *in* Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, Strathern, J. N., Jones, E. W., and Broach, J. R. Eds., p. 487-528, Cold Spring Harbor Laboratory, New York
7. Hopper, A. K., Banks, F., and Evangelidis, V. (1978) *Cell* **14**, 211-219
8. Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J. (1978) *Cell* **14**, 221-236
9. O'Farrell, P. Z., Cordell, B., Valenzuela, P., Rutter, W. J., and Goodman, H. M. (1978) *Nature* **274**, 438-445
10. Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J. (1979) *Cell* **18**, 27-35
11. Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979) *Cell* **18**, 37-45
12. Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536
13. Rauhut, R., Green, P. R., and Abelson, J. (1990) *J. Biol. Chem.* **265**, 18180-18184
14. Hartwell, L. H. (1967) *J. Bacteriol.* **93**, 1662-1670
15. Winey, M. and Culbertson, M. R. (1988) *Genetics* **118**, 609-617
16. DeMarini, D. J., Winey, M., Ursic, D., Webb, F., and Culbertson, M. R. (1992) *Mol. Cell. Biol.* **12**, 2154-2164

17. Leeds, P., Peltz, S. W., Jacobson, A., and Culbertson, M. R. (1991) *Genes Dev.* **12**, 2303-2314

18. Vijayraghavan, U., Company, M., and Abelson, J. (1989) *Genes Dev.* **3**, 1206-1216

19. Ho, C. K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990) *EMBO J.* **9**, 1245-1252

20. Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

21. Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986) *J. Biol. Chem.* **261**, 2978-2986

22. Westaway, S. K., Phizicky, E. M., and Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171-3176

23. Xu, Q., Phizicky, E. M., Greer, C. L., and Abelson, J. N. (1990) *Methods Enzymol.* **181**, 463-471

24. Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990) *Biochemistry* **29**, 6132-6138

25. Apostol, B. L., Westaway, S. K., Abelson, J., and Greer, C. L. (1991) *J. Biol. Chem.* **266**, 7445-7455

26. Thøgersen, H. C., Morris, H. R., Rand, K. N., and Gait, M. J. (1985) *Eur. J. Biochem.* **147**, 325-329

27. Abelson, J. N., Brody, E. N., Cheng, S.-C., Clark, M. W., Green, P. R., Dalbadie-McFarland, G., Lin, R.-J., Newman, A. J., Phizicky, E. M., and Vijayraghavan, U. (1986) *Chemica Scripta* **26B**, 127-137

28. Phizicky, E. M., Consaul, S. A., Nehrke, K. W., and Abelson, J. (1992) *J. Biol. Chem.* **267**, 4577-4582

29. McCraith, S. M., and Phizicky, E. M. (1990) *Mol. Cell. Biol.* **10**, 1049-1055

30. McCraith, S. M., and Phizicky, E. M. (1991) *J. Biol. Chem.* **266**, 11986-11992

31. Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1991) *Mol. Cell. Biol.* **11**, 5410-5416

32. Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1992) *J. Biol. Chem.* **267**, 10289-10294

33. Melton, D. A., De Robertis, E. M., and Cortese, R. (1980) *Nature* **284**, 143-148

34. DeRobertis, E. M., Black, P., and Nishikura, K. (1981) *Cell* **23**, 89-93

35. O'Connor, J. P., and Peebles, C. L. (1991) *Mol. Cell. Biol.* **11**, 425-439

36. Greer, C. L. (1986) *Mol. Cell. Biol.* **6**, 635-644

37. Greer, C. L., Javor, B., and Abelson, J. (1983) *Cell* **33**, 899-906

38. Clark, M. W., and Abelson, J. (1987) *J. Cell Biol.* **105**, 1515-1526

39. Apostol, B. L. and Greer, C. L. (1991) *Nucleic Acids Res.* **19**, 1853-1860

40. Tanner, N. K., Hanna, M. M., and Abelson, J. (1988) *Biochemistry* **27**, 8852-8861

41. Ogden, R. C., Lee, M.-C., and Knapp, G. (1984) *Nucleic Acids Res.* **12**, 9367-9382

42. Lee, M.-C., and Knapp, G. (1985) *J. Biol. Chem.* **260**, 3108-3115

43. Stucka, R., and Feldman, H. (1988) *Nucl. Acids Res.* **16**, 3583

44. Culbertson, M. R., and Winey, M. (1989) *Yeast* **5**, 405-427

45. Sprinzl, M., Hartmann, T., Meissner, F., Moll, J., and Vorderwülbecke, T. (1987) *Nucleic Acids Res. Suppl.* **15**, r53-r188

46. Strobel, M. C. and Abelson, J. (1986) *Mol. Cell. Biol.* **6**, 2663-2673

47. Reyes, V. M. and Abelson, J. (1988) *Cell* **55**, 719-730

48. Strobel, M. C. and Abelson, J. (1986) *Mol. Cell. Biol.* **6**, 2674-2683

49. Szekely, E., Belford, H. G., and Greer, C. L. (1988) *J. Biol. Chem.* **263**, 13839-13847

50. Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G. P. (1992) *Science* **255**, 1404-1408

51. Pearson, D., Willis, I., Hottinger, H., Bell, J., Kumar, A., Leupold, U., and Söll, D. (1985) *Mol. Cell. Biol.* **5**, 808-815

52. Greer, C. L., Söll, D., and Willis, I. (1987) *Mol. Cell. Biol.* **7**, 76-84

53. Gandini-Attardi, D., Margarit, I., and Tocchini-Valentini, G. P. (1985) *EMBO J.* **4**, 3289-3297

54. Baldi, M. I., Mattoccia, E., and Tocchini-Valentini, G. P. (1983) *Cell* **35**, 109-115

55. Willis, I., Hottinger, H., Pearson, D., Chisholm, V., Leupold, U., and Söll, D. (1984) *EMBO J.* **3**, 1573-1580

56. Shapero, M. H. and Greer, C. L. (1992) *Biochemistry* **31**, 2359-2367

57. Mathison, L., Winey, M., Soref, C., Culbertson, M. R., and Knapp, G. (1989) *Mol. Cell. Biol.* **9**, 4220-4228

58. Shapero, M. H., and Greer, C. L. (1991) *Biochemistry* **30**, 6465-6475

59. Mattoccia, E., Baldi, I. M., Gandini-Attardi, D., Ciafrè, S., and Tocchini-Valentini, G. P. (1988) *Cell* **55**, 731-738

60. Swerdlow, H., and Guthrie, C. (1984) *J. Biol. Chem.* **259**, 5197-5207

61. Schwartz, R. C., Greer, C. L., Gegenheimer, P., and Abelson, J. (1983) *J. Biol. Chem.* **258**, 8374-8383

62. Gegenheimer, P., Gabius, H.-J., Peebles, C. L., and Abelson, J. (1983) *J. Biol. Chem.* **258**, 8365-8373

63. Tyc, K., Kikuchi, Y., Konarska, M., Filipowicz, W., and Gross, H. J. (1983) *EMBO J.* **2**, 605-610

64. Laski, F. A., Fire, A. Z., RajBhandary, U. L., and Sharp, P. A. (1983) *J. Biol. Chem.* **258**, 11974-11980

65. Filipowicz, W., and Shatkin, A. J. (1983) *Cell* **32**, 547-557

66. Nishikura, K., and DeRobertis, E. M. (1981) *J. Mol. Biol.* **145**, 405-420

67. Filipowicz, W., and Vicente, O. (1990) *Meth. Enzymol.* **181**, 499-510

CHAPTER I

Structure and Function of the Yeast tRNA Ligase Gene

Shawn K. Westaway, Eric M. Phizicky, and John Abelson

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

[J. Biol. Chem. (1988) 263:3171-3176]

ABSTRACT

We report here the DNA sequence of the entire coding region of the *Saccharomyces cerevisiae* tRNA ligase gene. tRNA ligase is one of two enzymes required for tRNA splicing in yeast, and the enzyme is likely a single polypeptide with multiple activities. We find that tRNA ligase is a basic protein of 827 amino acids corresponding to a molecular weight of approximately 95,400. The inferred amino acid sequence for tRNA ligase is not significantly homologous to that of other known proteins of similar activity. In addition to the tRNA ligase reading frame and several other unidentified open reading frames, we have found two open reading frames, ORF1 and ORF2, near the 5'-end of the ligase structural gene. One of these, ORF2, produces a divergent transcript which initiates only 125 nucleotides upstream of the tRNA ligase transcript, and is present in approximately the same relative abundance as the transcript for tRNA ligase.

INTRODUCTION

tRNA processing in the yeast, *Saccharomyces cerevisiae*, as well as in other eukaryotes, often involves a splicing step. The substrates for tRNA splicing are a subset of tRNA precursors which contain introns ((1, 2) see Ref. 3). Unlike the case of introns in mRNA precursors, the introns in precursor tRNAs do not have obvious conserved sequences at their splice junctions or within the intron; the major common feature these introns share is their location, which is one base removed from the 3'-end of the anticodon (4). In contrast to the splicing of mRNA precursors, which involves several different ribonucleoproteins in complexes (for a review, see Ref. 5), the removal of introns from yeast tRNA precursors has been shown to be a relatively simple process involving two enzymes (6) (Fig. 1). After an endonuclease excises the intron to leave 5'-hydroxyl and 2',3'-cyclic phosphodiester ends, tRNA ligase joins the cognate 5' and 3' tRNA half-molecules (7, 8).

tRNA ligase protein is a single 90-kDa polypeptide which is likely to contain three activities required for the tRNA splicing reaction (9). tRNA ligase phosphorylates the 5' terminus of the 3' half-tRNA in the presence of ATP; opens the 2',3'-cyclic phosphodiester bond of the 5' half-tRNA, leaving a 2'-phosphomonoester; and, in a second ATP-dependent reaction, ligates the two tRNA halves (7) (see Fig. 1). Ligation occurs through an adenylylated enzyme intermediate, presumably at a lysine residue as it is in T4 RNA ligase (7, 10). The AMP moiety of this intermediate is transferred to the 5'-phosphate of the 3' half-tRNA to form a 5',5'-phosphoanhydride bond and is then released upon ligation of the two tRNA halves (9). tRNA ligase has been purified to near homogeneity, and the three activities described above cosediment with this protein in the final glycerol gradient step (9). (Another enzyme presumably removes the 2'-phosphate which remains at the splice junction after the action of tRNA ligase.)

In order to study the functional domains of the tRNA ligase protein, we have cloned the gene from *S. cerevisiae* (9). We present here its DNA sequence, which encodes a protein of 827 amino acids (95.4 kDa). The ligase amino acid sequence does not resemble that of other proteins with similar activities.

The transcript for tRNA ligase starts at two major sites, 104 and 125 nucleotides upstream from the initiation codon. Near the 5'-end of tRNA ligase, there are two other open reading frames, ORF1[1] and ORF2. ORF1 begins upstream from tRNA ligase and overlaps the beginning of the ligase gene. We discuss experiments suggesting that its translation is not essential. The ORF2 reading frame begins 342 nucleotides upstream from the ligase initiation codon and proceeds in the opposite direction. This gene is transcribed to produce a 2.1-kilobase RNA that begins 125 nucleotides upstream from the tRNA ligase transcript.

EXPERIMENTAL PROCEDURES

Strains and Plasmids. *Escherichia coli* JM101 (11) was used for preparing M13 phage for sequencing and restriction analysis. Yeast strain TSY6-11Bα (*MATα prc1-Δ2 pep4-3 ura3-52 leu2-3,112 his4*) was used as the host for analysis of transcription products and transcript mapping and was obtained from Dr. S. Emr (California Institute of Technology). Yeast strain SWY497 (our nomenclature) has an ochre mutation within ORF1 constructed by site-directed oligonucleotide mutagenesis of single-stranded M13 phage (12, 13) followed by standard yeast gene replacements methods (14, 15) using yIP5 (15) as the integrating vector. EMPY20 is the parent yeast diploid for SWY497 and was constructed by mating SS328 (*MATα his3Δ200 lys2-801^a ade2-101° ura3-52 GAL suc2)* to SS330 (*MATa his3Δ200 tyr1 ade2-101° ura3-52 GAL suc2*). SS328 and SS330 were obtained from Dr. S. Scherer (University of Minnesota). Construction of plasmid pUC12-RLG was described previously (9). Plasmid pEP99 (our nomenclature) is a derivative of yEP24 which contains the entire tRNA ligase gene as well as 5 kilobases of upstream DNA. This plasmid was obtained by screening an *E. coli* library of yeast genomic DNA (16) for inserts that hybridize to the ligase DNA (the library was provided by Dr. J. Campbell). Screening was performed as described by Maniatis *et al.* (17). Transformation of this plasmid and its parent, yEP24 (18), into strain TSY6-11Bα was done as described by Ito *et al.* (19).

M13 Template Construction and Sequencing. Four unique restriction sites were used to subclone the 4-kilobase *Eco*RI insert of plasmid pUC12-RLG into phage M13 for sequencing (see Table I). Phage were purified by CsCl density gradient centrifugation prior to sequencing (20). These templates were sequenced by the Sanger dideoxy sequencing method (21) by first using the M13 sequencing primer and then by priming with synthetic oligonucleotides as needed (22, 23). These oligonucleotide primers were

synthesized on an Applied Biosystems 380A DNA synthesizer (deoxyoligonucleotides were provided by Dr. S. H. Horvath, California Institute of Technology).

Preparation of RNA. Yeast cells were grown in SM minimal minus uracil media (24) to 2 X 10^7 cells/ml. After harvest, RNA was extracted by disruption of the cells with glass beads in the presence of cold phenol (25). The polyadenylated RNA was separated from total RNA on oligo(dT)-cellulose (17).

RNA Analysis. 3 μg of polyadenylated RNA was resolved on a 1.5% formaldehyde-agarose gel and transferred to nitrocellulose (17). Single-stranded DNA probes used in the Northern blot were made by primer extension with [α-^{32}P]dCTP across the ligase insert of the M13 template followed by restriction cutting, denaturation of the duplex DNA, and isolation of the labeled single-stranded DNA. The primers and templates were selected from those listed in Table I and Fig. 2. Prehybridization, hybridization, and washing of the Northern blot were carried out according to Thomas (26).

Transcript mapping by primer extension using 20-40 units of reverse transcriptase was carried out as described previously (27) with 3 μg of polyadenylated RNA. S1 nuclease analysis was performed as described (28, 29) with 3 μg of polyadenylated RNA and the appropriate kinased and primer-extended oligonucleotides. The oligonucleotide used as a primer for detecting the tRNA ligase transcript is complementary to the first 30 nucleotides of the tRNA ligase coding region.

Sequence Analysis. Sequence data was assembled, translated, and analyzed using Versions 4.2 and 4.5 of the Intelligenetics GENED, GEL, SEQ, and PEP programs of the BIONET™ National Computer Resource for Molecular Biology. All protein sequences used in homology searches were obtained using the Intelligenetics QUEST, IFIND, and XFASTP programs with the National Institutes of Health sequence and National

Biomedical Research Foundation protein data bases in BIONET™, unless otherwise referenced.

Materials. M13 sequence primer was from New England Biolabs. [α-^{32}P]dNTPs were purchased from Amersham Corp. dNTPs, dideoxy-NTPs, and S1 nuclease were from Pharmacia LKB Biotechnology Inc. Klenow fragment of DNA polymerase was purchased from Bethesda Research Laboratories. Reverse transcriptase was from Boehringer Mannheim.

RESULTS

tRNA Ligase Sequence. The 4-kilobase *Eco*RI fragment containing the yeast tRNA ligase gene was sequenced according to the strategy outlined in Fig. 2 and described under "Experimental Procedures." Five unique restriction fragments were cloned in both orientations into phage M13 (Table I). M13 sequencing primer was used to obtain initial sequence information. Oligonucleotides corresponding to unique regions of nucleotide sequence were then used as primers in directing continued synthesis along the same strand of each clone (22, 23). In this way, we obtained the sequence of both strands of each restriction fragment. To confirm the sequence at the restriction site junctions, separate clones were sequenced which span the junctions (Table I). The entire sequence is shown in Fig. 3.

The ligase reading frame begins at nucleotide 551 and extends to nucleotide 3031. This sequence encodes a protein of 827 amino acids, with a molecular mass of 95.4 kDa, comparable to that estimated (90 kDa) from the migration of purified tRNA ligase on sodium dodecyl sulfate-polyacrylamide gels (9). We have shown previously that the amino-terminal sequence of purified tRNA ligase begins at the amino acid residue immediately following the methionine codon at nucleotide 551, and we presumed that this

codon was the initiation codon (9). This AUG is indeed the initiation codon. It is the first AUG in the open reading frame corresponding to ligase, and, as we show below, it is the first AUG in any reading frame following the ligase transcription initiation site.

Other Open Reading Frames. There are two other open reading frames, ORF1 and ORF2, near the 5'-end of the ligase structural gene. These are diagrammed in Fig. 4 together with all open reading frames that contain a start codon and are greater than 150 amino acids in length. ORF1 precedes the ligase reading frame and overlaps the beginning of the ligase protein-coding region by eight nucleotides. We have sequenced this region of overlap in both directions from each of two independent ligase clones isolated from different yeast libraries (pUC12-RLG and pEP99) and have obtained identical nucleotide sequences. We are thus confident that ligase and ORF1 are out-of-frame with respect to one another and that ligase is translated beginning at nucleotide 551 of this sequence (see also below).

If ORF1 is translated in yeast, its translation product is not essential for cell viability. By standard methods (see "Experimental Procedures"), we have constructed a diploid strain (SWY497) in which ORF1 is interrupted by an ochre mutation before the region of overlap with the ligase gene. This diploid sporulated to yield four healthy viable spores in 18 of 20 tetrads dissected, comparable to the frequency of four-spore tetrads in the parent strain. Southern hybridization of genomic DNA from three of these tetrads confirms that in each case two of the four spores contained the ochre mutation. In addition, the homozygous diploid sporulates normally (data not shown). These results demonstrate that ORF1 is not translated to yield an essential gene product, either independently or as a frameshift fusion protein, as has been seen for the *Ty* elements in yeast (30, 31). We cannot detect an ORF1 transcript or a transcript fusing ORF1 to the full-length tRNA ligase gene (see below). In addition, we detect no polypeptide using immune precipitation or

immune blot analysis with antibodies raised to a synthetic oligopeptide corresponding to the amino-terminal sequence of ORF1.[2]

The second open reading frame, ORF2, begins 342 nucleotides upstream from the tRNA ligase coding region and proceeds in the opposite direction. We have obtained evidence, which will be described elsewhere, that ORF2 encodes an essential gene product.[3] The other two open reading frames depicted in Fig. 3, ORF3 and ORF4, are as yet uncharacterized.

Transcript Analysis. Because of the proximity of ORF1 and ORF2 to tRNA ligase, we have fully analyzed the transcripts upstream of and within the coding region of tRNA ligase. We isolated polyadenylated RNA from two yeast strains. One strain carried the yeast plasmid yEP24 and the other strain carried a derivative of yEP24 (pEP99) containing all of the tRNA ligase gene, as well as all of the ORF1 and ORF2 DNA sequences. The polyadenylated RNA from these two strains was resolved on gels, transferred to nitrocellulose, and analyzed by hybridization to various probes. One such example is shown in Fig. 5.

A ligase-specific probe hybridizes to a major 2.8-kilobase transcript (Fig. 5*A*), which is of sufficient size to accommodate the entire tRNA ligase coding region. This transcript is detected with each of the probes we used (*Eco*RI to *Hpa*I, *Hpa*I to *Bgl*II, *Bgl*II to *Hin*dIII, and *Hin*dIII to *Xba*I, see Fig. 5*C*). In addition, there is a minor transcript of 2.2 kilobases which is about 50-fold less abundant than the major ligase transcript. This minor transcript is only faintly detected with a probe to the 3'-end of the ligase gene (*Hin*dIII to *Xba*I, data not shown). We did not detect small transcripts which could correspond to the ORF1-polyadenylated RNA nor a larger 3.1-kilobase transcript which could correspond to a transcript of ORF1 fused to tRNA ligase.

The ORF2-specific probe hybridizes to a 2.1-kilobase transcript (Fig. 5*B*), of comparable abundance to the major tRNA ligase transcript. This ORF2 transcript is not

detected using probes within the tRNA ligase gene. As expected, after longer exposure of the autoradiogram, all of the transcripts described above are detected in polyadenylated RNA isolated from the wild-type strain. Fig. 5*C* summarizes the transcripts, their direction, and their relative abundance for this region.

We have mapped the 5'-ends of the ligase and ORF2 transcripts more thoroughly, using both S1 nuclease and reverse transcriptase. The results are shown in Fig. 6. As is the case for some other yeast mRNA transcripts (32-37), the ligase transcript has multiple transcription initiation sites (Fig. 6*A*). In each of two independent experiments, two major transcription initiation sites were detected for the tRNA ligase transcript with both reverse transcriptase and S1 nuclease, together with several minor transcription initiation sites. (Some of these minor sites are seen preferentially with one or the other method.) The two major transcription initiation sites for tRNA ligase are diagrammed in Fig. 6*C*, relative to the coding region of ligase. Both of these sites, as well as all of the minor initiation sites, begin upstream of the initiation codon at nucleotide 551, and this AUG is the first in any reading frame following any of the transcription initiation sites. This AUG, therefore, represents the initiation codon for tRNA ligase.

The ORF2 transcript begins at about seven different positions (Fig. 6*B*) within the 40-base pair region from nucleotides 263 to 303 (numbers refer to the sequence shown in Fig. 3). The four most intense initiation signals are depicted in Fig. 6*C*. Two of these start sites are followed immediately by an AUG codon at nucleotides 291-289, which has an in-frame stop codon two amino acids downstream. The next AUG occurs at nucleotide 209 and represents the beginning of the ORF2 open reading frame. We do not know how ORF2 translation is affected. Either ORF2 is only translated from the two shorter transcripts, or the false AUG at nucleotide 291 plays some role in the regulation of ORF2 expression. Such a regulatory role has been ascribed to the multiple AUG codons at the 5'-end of the *GCN4* gene in yeast (38).

DISCUSSION

tRNA Ligase Sequence Analysis. We have sequenced the yeast tRNA ligase gene and mapped the 5'-end of its transcript. Purified ligase protein begins at the amino acid residue following the methionine codon at nucleotide 551 (9). The data presented here provide further evidence that this methionine residue is the correct amino terminus of the tRNA ligase gene. (The methionine residue itself is presumably removed after translation.)

tRNA ligase is 827 amino acids long and has a molecular mass of 95.4 kDa. As might be expected for a protein involved in tRNA metabolism, tRNA ligase is a basic protein. Greater than 10% of all the amino acid residues are lysine, and the net charge of the protein is calculated to be +12.5 if histidine is counted as +0.5.

We have previously reported that yeast tRNA ligase is a relatively rare protein present in approximately 400 copies/yeast cell (9). A comparison of the codon usage of tRNA ligase to other yeast proteins (32, 39-42) indicates that tRNA ligase uses a large percentage of rarely used codons (data not shown). These observations fit well with the hypothesis proposed by Bennetzen and Hall (39) that less abundant proteins do not require the extreme codon bias toward "preferred codons" as is seen for highly abundant proteins. Since yeast tRNA ligase is directly involved in the processing of intron-containing tRNAs, its expression might be controlled by the levels of that subset of tRNAs. We, therefore, examined the codon usage of tRNA ligase relative to the codons in yeast which are specifically translated by tRNAs whose precursors contain introns (see Ref. 3). We were unable to detect any such correlation. There is no codon bias either for or against the use of spliced tRNAs to translate tRNA ligase.

The mechanism by which yeast tRNA ligase ligates RNA is very similar to that used by T4 RNA ligase and T4 DNA ligase. All three enzymes proceed through an adenylylated enzyme intermediate whose adenyl moiety is used to activate their substrates (for reviews, see Refs. 43 and 44). Available evidence indicates that both yeast tRNA

ligase and T4 RNA ligase are adenylylated at a lysine residue (7, 10). Given these similarities in mechanism, it was of obvious interest to compare the sequence of tRNA ligase with that of other ligases. We were unable to detect any obvious sequence homologies between yeast tRNA ligase and either of the T4 ligases. Another T4 enzyme, polynucleotide kinase, phosphorylates 5'-hydroxy termini of RNA molecules and possesses a 3'-phosphatase activity (45, 46) resembling the kinase and 2',3'-cyclic phosphodiesterase activities catalyzed by the tRNA ligase protein (47). There is, however, no obvious sequence homology between tRNA ligase and T4 polynucleotide kinase. In fact, a screen of the entire protein data base (SNBRF, BIONET) for homologies also proved negative.

Studies in progress indicate that tRNA ligase does have separable domains responsible for some of the different activities observed during tRNA splicing.[4] Results from this type of analysis should pinpoint the location of these domains within the tRNA ligase protein. This, in turn, will allow us to search for more subtle similarities between tRNA ligase and other proteins of similar function.

Transcription Analysis. Upstream of ligase there are two other open reading frames, ORF1 and ORF2. As described above, the evidence indicates that ORF1 does not encode an essential protein. By contrast, ORF2 is transcribed, and our preliminary results indicate that it is essential. The divergent ORF2 and tRNA ligase transcripts do not overlap, but they are separated by only 125 nucleotides at their closest points. Although closely situated divergent transcripts are relatively common in yeast (48-50), to our knowledge, tRNA ligase and ORF2 are closer than most transcripts. Though we have no evidence that bears directly on this, it is possible that tRNA ligase and ORF2 share regulatory elements. In this connection, the region between the tRNA ligase and ORF2 transcript initiation sites contains a 17-base pair GC-rich region (nucleotides 349-365, see Fig. 3), bounded on the ORF2 side by two possible "TATA"-like sequences (nucleotides 338-334 and 310-306),

and on the ligase side by a "TATA"-like sequence (nucleotides 413-416), 11 nucleotides upstream from the first major transcription initiation site. The significance of these sequences is unknown.

The 3'-end of the tRNA ligase transcript has not been mapped; however, both tRNA ligase and ORF2 mRNAs are clearly polyadenylated (see Fig. 5). tRNA ligase contains the sequence "AAUAA" 111 nucleotides downstream of the termination codon (see Fig. 3), which is a close match to the polyadenylation signal consensus of AAUAAA, proposed to be important in polyadenylation of messenger RNA in other eukaryotic genes (51). Several other yeast genes including enolase gene 8, enolase gene 46, and glyceraldehyde-3-phosphate dehydrogenase gene 63 (52, 53) also contain this "AAUAA" sequence in their 3'-nontranslated regions.

Yeast tRNA ligase protein possesses several different properties which make it unique. tRNA ligase catalyzes three separate enzymatic activities responsible for the splicing of precursor tRNAs; there is evidence indicating that tRNA ligase forms a tRNA splicing complex *in vitro* in the presence of endonuclease and tRNA precursors (54); tRNA ligase has been localized to a specific site at or near the inner nuclear membrane (55), so therefore it ought to contain a nuclear localization signal of some sort and/or a site which recognizes the nuclear component to which it binds; finally, tRNA ligase must be able to recognize all of the different pre-tRNA substrates for splicing. Studies in this laboratory are aimed at determining how these many different aspects are embodied in this single polypeptide.

ACKNOWLEDGMENTS

We thank Suzanna Horvath for preparing the oligopeptide and oligonucleotides, Russell Doolittle for assistance in sequence comparisons, and Ryn Miake-Lye and Michael Clark for critical analysis of the manuscript. Special thanks to Marjorie Strobel, Jennifer Normanly, and Elizabeth Grayhack for help and advice on various aspects of the work.

FOOTNOTES

[1]The abbreviation used: ORF, open reading frame.

[2]P. Green, M. Clark, and J. Abelson, unpublished data.

[3]S. Westaway and J. Abelson, unpublished results.

REFERENCES

1. Goodman, H. M., Olson, M. V., and Hall, B. D. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5353-5457
2. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75**, 190-194
3. Ogden, R. C., Lee, M.-C., and Knapp, G. (1984) *Nucleic Acids Res.* **12**, 9367-9382
4. Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979) *Cell* **18**, 37-45
5. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) *Ann. Rev. Biochem.* **55**, 1119-1150
6. Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J. (1979) *Cell* **18**, 27-35
7. Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

8. Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536

9. Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1983) *J. Biol. Chem.* **261**, 2978-2986

10. Tho/gersen, H. C., Morris, H. R., Rand, K. N., and Gait, M. J. (1985) *Eur. J. Biochem.* **147**, 325-329

11. Messing, J. (1979) *Recomb. DNA Tech. Bull.* **2**, 43-48

12. Newman, A. J., Ogden, R. C., and Abelson, J. (1983) *Cell* **35**, 117-125

13. Zoller, M. J., and Smith, M. (1982) *Nucleic Acids Res.* **10**, 6487-6500

14. Boeke, J. D., LaCroute, F., and Fink, G. R. (1984) *Mol. Gen. Genet.* **197**, 345-346

15. Struhl, K., Stinchcomb, D. T., Scherer, S., and Davis, R. W. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1035-1039

16. Carlson, M., and Botstein, D. (1982) *Cell* **28**, 145-154

17. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

18. Botstein, D., Falco, S. C., Stewart, S. E., Brennan, M., Scherer, S., Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979) *Gene (Amst.)* **8**, 17-24

19. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* **153**, 163-168

20. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, G. (1970) *Virology* **40**, 734-744

21. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467

22. Hong, G. F. (1982) *Biosci. Rep.* **2**, 907-912

23. Sanchez-Pescador, R., and Urdea, M. S. (1984) *DNA* **3**, 339-343

24. Sherman, F., Fink, G. R., and Hicks, J. B. (1982) in *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

25. Elder, R. T., Loh, E. Y., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2432-2436

26. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 5201-5205

27. Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E., and Abelson, J. (1984) *Cell* **39**, 611-621

28. Favaloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* **65**, 718-749

29. Treisman, R., and Kamen, R. (1981) *J. Mol. Biol.* **148**, 273-301

30. Clare, J., and Farabaugh, P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2829-2833

31. Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J. (1985) *Nature* **313**, 243-246

32. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257**, 3018-3025

33. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., and Smith, M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2258-2262

34. Guarente, L. (1984) *Cell* **36**, 799-800

35. Hinnebusch, A. G., and Fink, G. R. (1983) *J. Biol. Chem.* **258**, 5238-5247

36. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 961-981

37. Zalkin, H., and Yanofsky, C. (1982) *J. Biol. Chem.* **257**, 1491-1500

38. Hinnebusch, A. G. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6442-6446

39. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257**, 3026-3031

40. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) *Nucleic Acids Res.* **9**, r43-r74

41. Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. (1980) *Nucleic Acids Res.* **8**, r49-r62

42. Ikemura, T. (1985) *Mol. Biol. Evol.* **2**, 13-34

43. Engler, M. J., and Richardson, C. C. (1982) in *The Enzymes* (Boyer, P. D., ed.) Vol. 15, pp. 1-29, Academic Press, Orlando, FL

44. Uhlenbeck, O. C., and Gumport, R. I. (1982) in *The Enzymes* (Boyer, P. D., ed.) Vol. 15, pp. 31-58, Academic Press, Orlando, FL

45. Cameron, V., and Uhlenbeck, O. C. (1977) *Biochemistry* **16**, 5120-5126

46. Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. U. S. A.* **54**, 158-165

47. Uhlenbeck, O. C. (1983) *Trends Biochem. Sci.* **8**, 94-96

48. Johnston, M., and Davis, R. W. (1984) *Mol. Cell. Biol.* **4**, 1440-1448

49. Osley, M. A., Gould, J., Kim, S., Kane, M., and Hereford, L. (1986) *Cell* **45**, 537-544

50. Struhl, K. (1985) *Nucleic Acids Res.* **13**, 8587-8601

51. Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263**, 211-214

52. Holland, J. P., and Holland, M. J. (1980) *J. Biol. Chem.* **255**, 2596-2605

53. Holland, M. J., Holland, J. P., Thill, G. P., and Jackson, K. A. (1981) *J. Biol. Chem.* **256**, 1385-1395

54. Greer, C. L. (1986) *Mol. Cell. Biol.* **6**, 635-644

55. Clark, M. W., and Abelson, J. (1987) *J. Cell Biol.* **105**, 1515-1526

FIGURE LEGENDS

Figure 1. **The mechanism of tRNA splicing in yeast, as originally described by Greer *et al.* (7).** An endonuclease from yeast is responsible for the first step, cleaving the precursor tRNA into two tRNA half-molecules. Yeast tRNA ligase likely carries out the remaining steps of the process, leaving a tRNA with a 2'-phosphate at the splice junction. See text for details. *IVS*, intervening sequence.

Figure 2. **Restriction map of the 4.2-kilobase insert in pUC12-RLG.** Sequencing directions and distances are shown above and below the restriction map. *Open boxes* indicate positions of primed synthesis using M13 sequence primer. *Closed boxes* indicate positions of synthetic oligonucleotide primers. *R(A), Eco*RI (site A); *Hp, Hpa*I; *Sc, Sca*I; *Bg, Bgl*II; *Hd, Hin*dIII; *K, Kpn*I; *RV, Eco*RV; *X, Xba*I; *R(B), Eco*RI (site B). *nt*, nucleotide.

Figure 3. **Sequence of the 4168-nucleotide *Eco*RI fragment of pUC12-RLG.** The tRNA ligase coding region is shown translated. A putative initiation codon and the direction of transcription are shown for ORF2. The start and stop positions for ORF1 are also indicated.

Figure 4. **Diagram showing open reading frames (containing an AUG codon) found within the sequence of the *Eco*RI fragment of pUC12-RLG.** The two ORFs downstream of tRNA ligase, ORF3 and ORF4, are uncharacterized. Other possible open reading frames of less than 150 amino acids have been omitted for clarity. *aa*, amino acids; other abbreviations are same as defined in the legend to Fig. 2.

Figure 5. **Transcripts of tRNA ligase and ORF2.** Polyadenylated RNA from a wild-type strain (*WT*) and a strain carrying an overproducer of tRNA ligase and ORF2 (*OP*) was resolved on gels, transferred to nitrocellulose, and hybridized to various probes as described under "Experimental Procedures." *MW*, molecular weight markers of single-stranded DNA. *A*, RNA was hybridized with a single-stranded DNA probe complementary to tRNA ligase and spanning the region between *Hpa*I and *Bgl*II (see *C*). *B*, RNA was hybridized with a single-stranded DNA probe complementary to ORF2 and spanning the region between *Eco*RI (site A) and *Hpa*I (see *C*). *C*, diagrammatic representation of the positions and relative abundance of the major transcripts for ORF2 and tRNA ligase. *R, Eco*RI; *H, Hpa*I; *B, Bgl*II; *Hd, Hin*dIII; *X, Xba*I. Numbers correspond to the transcripts shown in *A* and *B*.

Figure 6. **Transcription initiation sites for tRNA ligase and ORF2.** *WT*, analysis of a wild-type strain; *OP*, analysis of a strain overproducing tRNA ligase and ORF2. (Transcripts are detected from wild-type cells upon overexposure of the autoradiograms.) *A*, S1 nuclease and reverse transcriptase primer extension analysis for tRNA ligase. S1 nuclease *lanes (S1)* are on the left, reverse transcriptase *lanes (RT)* are in the middle, and Sanger dideoxy sequencing ladders are on the right. The *arrows* indicate positions of major transcription initiation sites as diagrammed in *C*. *B*, analysis of the ORF2 transcription initiation sites; *symbols* are the same as in *A*. *C*, diagram of the *Eco*RI to *Hpa*I region depicting the major transcription initiation sites for both tRNA ligase and ORF2. Numbers correspond to those used in Fig. 3. The translation start sites (ATG) are shown for each gene.

TABLE I

Cloning strategy for sequencing the yeast tRNA ligase gene

Restriction Fragment[a]	M13 vector	Vector sites	Size
			bp
*Eco*RI(A) - *Hpa*I	mp10 and mp11	*Eco*RI - *Sma*I	595
*Hpa*I - *Bgl*II	mp10 and mp11	*Sma*I - *Bam*HI	966
*Bgl*II - *Hin*dIII	mp10 and mp11	*Bam*HI - *Hin*dIII	726
*Hin*dIII - *Xba*I	mp10 and mp11	*Hin*dIII - *Xba*I	1458
*Xba*I - *Eco*RI(B)	mp10 and mp11	*Xba*I - *Eco*RI	423
*Eco*RI(A) - *Sca*I	mp11	*Eco*RI - *Sma*I	1379
*Sca*I - *Kpn*I	mp18	*Sma*I - *Kpn*I	928
*Eco*RV - *Eco*RI(B)	mp11	*Sma*I - *Eco*RI	905

[a]See restriction map, Fig. 2.

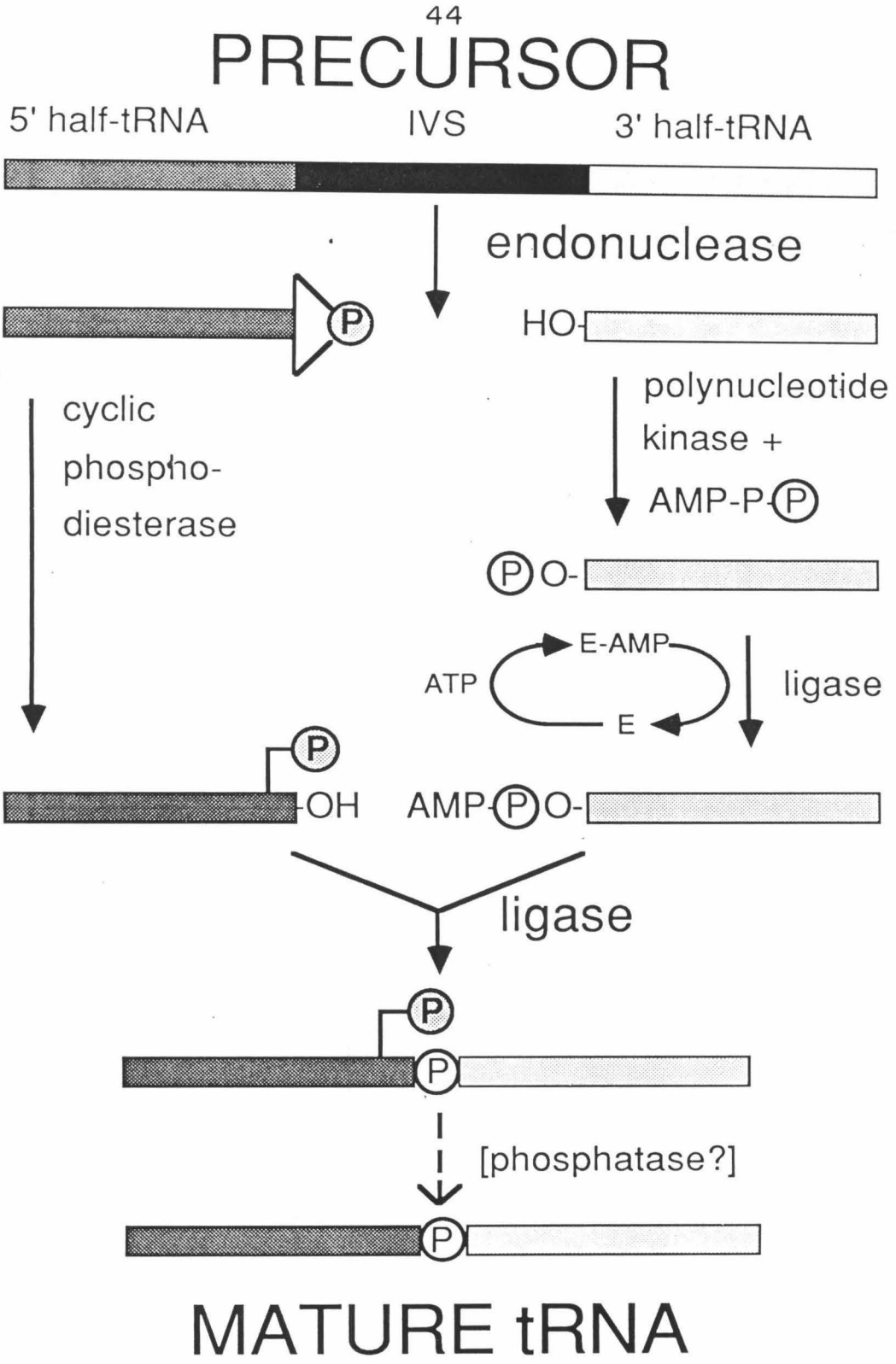

PRECURSOR
5' half-tRNA
IVS
3' half-tRNA
endonuclease
HO-
P
cyclic
phospho-
diesterase
polynucleotide
kinase +
AMP-P-P
P O-
E-AMP
ATP
E
ligase
P
OH
AMP-P O-
ligase
P
P
[phosphatase?]
P
MATURE tRNA

Figure 1

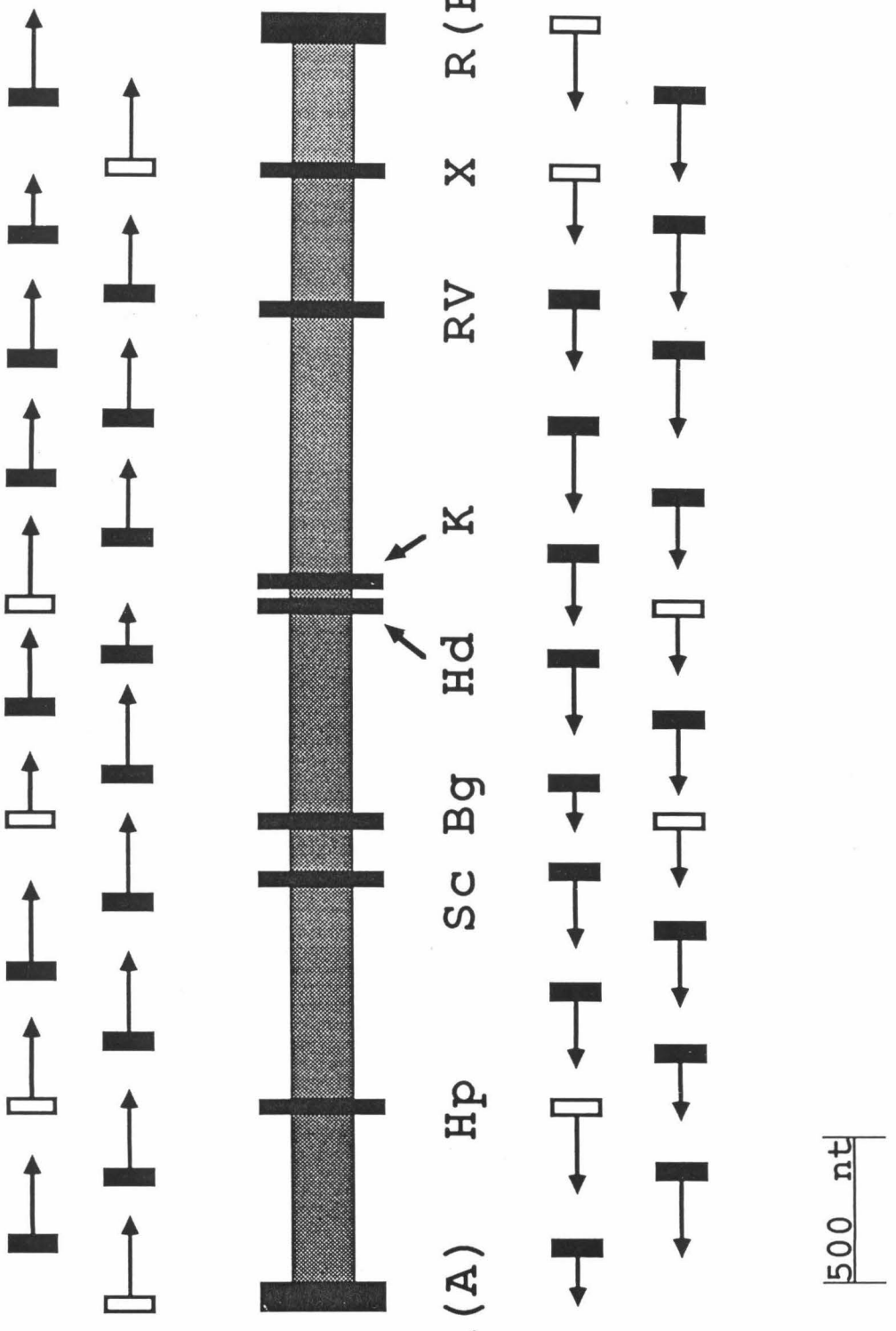


Figure 2

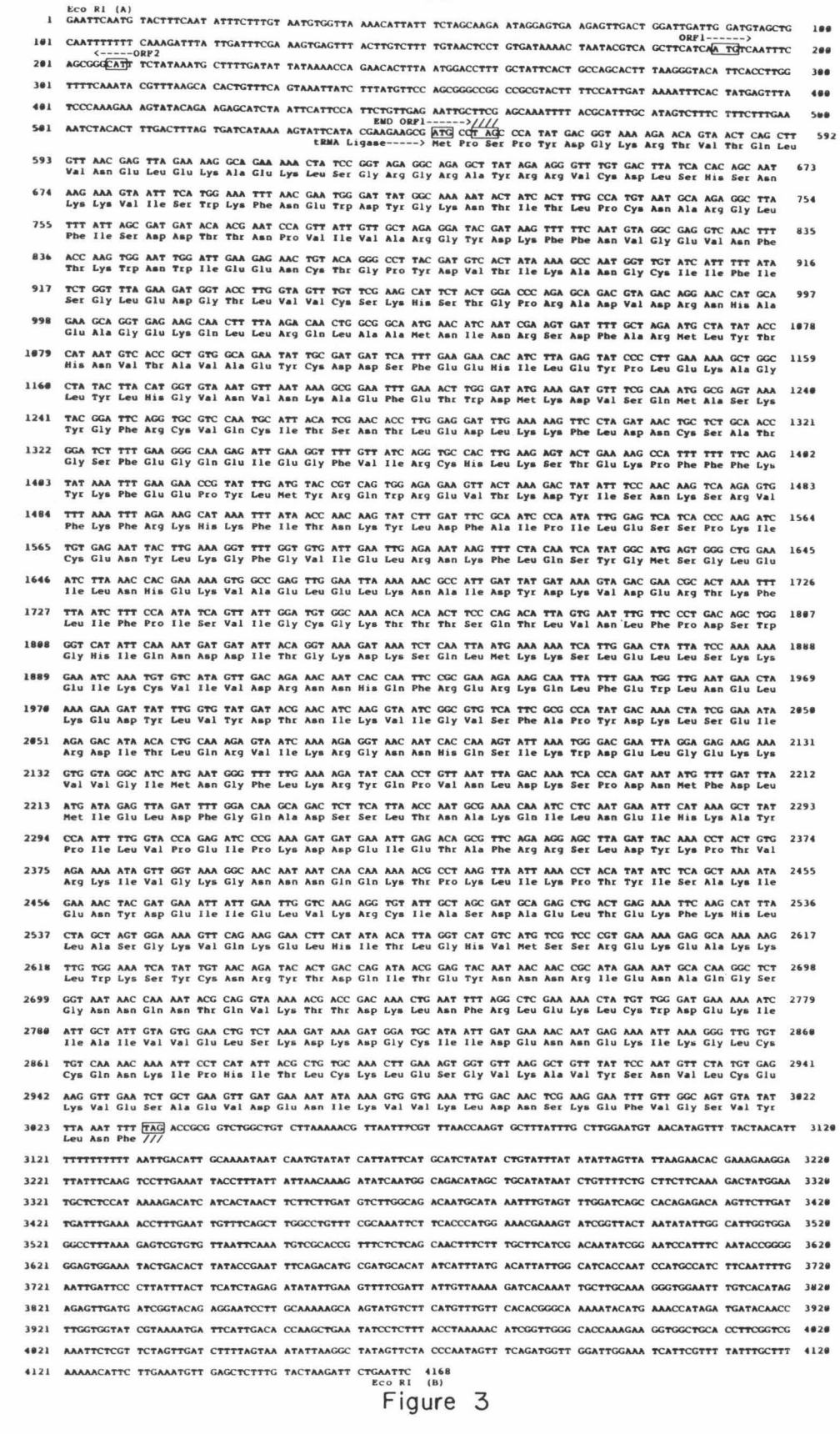

Eco RI (A)
ORF1------>
<-----ORF2
END ORF1------>////
tRNA Ligase-----> Met Pro Ser Pro Tyr Asp Gly Lys Arg Thr Val Thr Gln Leu
Val Asn Glu Leu Glu Lys Ala Glu Lys Leu Ser Gly Arg Gly Arg Ala Tyr Arg Arg Val Cys Asp Leu Ser His Ser Asn
Lys Lys Val Ile Ser Trp Lys Phe Asn Glu Trp Asp Tyr Gly Lys Asn Thr Ile Thr Leu Pro Cys Asn Ala Arg Gly Leu
Phe Ile Ser Asp Asp Thr Thr Asn Pro Val Ile Val Ala Arg Gly Tyr Asp Lys Phe Phe Asn Val Gly Glu Val Asn Phe
Thr Lys Trp Asn Trp Ile Glu Glu Asn Cys Thr Gly Pro Tyr Asp Val Thr Ile Lys Ala Asn Gly Cys Ile Ile Phe Ile
Ser Gly Leu Glu Asp Gly Thr Leu Val Val Cys Ser Lys His Ser Thr Gly Pro Arg Ala Asp Val Asp Arg Asn His Ala
Glu Ala Gly Glu Lys Gln Leu Leu Arg Gln Leu Ala Ala Met Asn Ile Asn Arg Ser Asp Phe Ala Arg Met Leu Tyr Thr
His Asn Val Thr Ala Val Ala Glu Tyr Cys Asp Asp Ser Phe Glu Glu His Ile Leu Glu Tyr Pro Leu Glu Lys Ala Gly
Leu Tyr Leu His Gly Val Asn Val Asn Lys Ala Glu Phe Glu Thr Trp Asp Met Lys Asp Val Ser Gln Met Ala Ser Lys
Tyr Gly Phe Arg Cys Val Gln Cys Ile Thr Ser Asn Thr Leu Glu Asp Leu Lys Lys Phe Leu Asp Asn Cys Ser Ala Thr
Gly Ser Phe Glu Gly Gln Glu Ile Glu Gly Phe Val Ile Arg Cys His Leu Lys Ser Thr Glu Lys Pro Phe Phe Phe Lys
Tyr Lys Phe Glu Glu Pro Tyr Leu Met Tyr Arg Gln Trp Arg Glu Val Thr Lys Asp Tyr Ile Ser Asn Lys Ser Arg Val
Phe Lys Phe Arg Lys His Lys Phe Ile Thr Asn Lys Tyr Leu Asp Phe Ala Ile Pro Ile Leu Glu Ser Ser Pro Lys Ile
Cys Glu Asn Tyr Leu Lys Gly Phe Gly Val Ile Glu Leu Arg Asn Lys Phe Leu Gln Ser Tyr Gly Met Ser Gly Leu Glu
Ile Leu Asn His Glu Lys Val Ala Glu Leu Glu Leu Lys Asn Ala Ile Asp Tyr Asp Lys Val Asp Glu Arg Thr Lys Phe
Leu Ile Phe Pro Ile Ser Val Ile Gly Cys Gly Lys Thr Thr Thr Ser Gln Thr Leu Val Asn Leu Phe Pro Asp Ser Trp
Gly His Ile Gln Asn Asp Asp Ile Thr Gly Lys Asp Lys Ser Gln Leu Met Lys Lys Ser Leu Glu Leu Leu Ser Lys Lys
Glu Ile Lys Cys Val Ile Val Asp Arg Asn Asn His Gln Phe Arg Glu Arg Lys Gln Leu Phe Glu Trp Leu Asn Glu Leu
Lys Glu Asp Tyr Leu Val Tyr Asp Thr Asn Ile Lys Val Ile Gly Val Ser Phe Ala Pro Tyr Asp Lys Leu Ser Glu Ile
Arg Asp Ile Thr Leu Gln Arg Val Ile Lys Arg Gly Asn Asn His Gln Ser Ile Lys Trp Asp Glu Leu Gly Glu Lys Lys
Val Val Gly Ile Met Asn Gly Phe Leu Lys Arg Tyr Gln Pro Val Asn Leu Asp Lys Ser Pro Asp Asn Met Phe Asp Leu
Met Ile Glu Leu Asp Phe Gly Gln Ala Asp Ser Ser Leu Thr Asn Ala Lys Gln Ile Leu Asn Glu Ile His Lys Ala Tyr
Pro Ile Leu Val Pro Glu Ile Pro Lys Asp Asp Glu Ile Glu Thr Ala Phe Arg Arg Ser Leu Asp Tyr Lys Pro Thr Val
Arg Lys Ile Val Gly Lys Gly Asn Asn Asn Gln Gln Lys Thr Pro Lys Leu Ile Lys Pro Thr Tyr Ile Ser Ala Lys Ile
Glu Asn Tyr Asp Glu Ile Ile Glu Leu Val Lys Arg Cys Ile Ala Ser Asp Ala Glu Leu Thr Glu Lys Phe Lys His Leu
Leu Ala Ser Gly Lys Val Gln Lys Glu Leu His Ile Thr Leu Gly His Val Met Ser Ser Arg Glu Lys Glu Ala Lys Lys
Leu Trp Lys Ser Tyr Cys Asn Arg Tyr Thr Asp Gln Ile Thr Glu Tyr Asn Asn Asn Arg Ile Glu Asn Ala Gln Gly Ser
Gly Asn Asn Gln Asn Thr Gln Val Lys Thr Thr Asp Lys Leu Asn Phe Arg Leu Glu Lys Leu Cys Trp Asp Glu Lys Ile
Ile Ala Ile Val Val Glu Leu Ser Lys Asp Lys Asp Gly Cys Ile Ile Asp Glu Asn Asn Glu Lys Ile Lys Gly Leu Cys
Cys Gln Asn Lys Ile Pro His Ile Thr Leu Cys Lys Leu Glu Ser Gly Val Lys Ala Val Tyr Ser Asn Val Leu Cys Glu
Lys Val Glu Ser Ala Glu Val Asp Glu Asn Ile Lys Val Val Lys Leu Asp Asn Ser Lys Glu Phe Val Gly Ser Val Tyr
Leu Asn Phe ///
Eco RI (B)

Figure 3

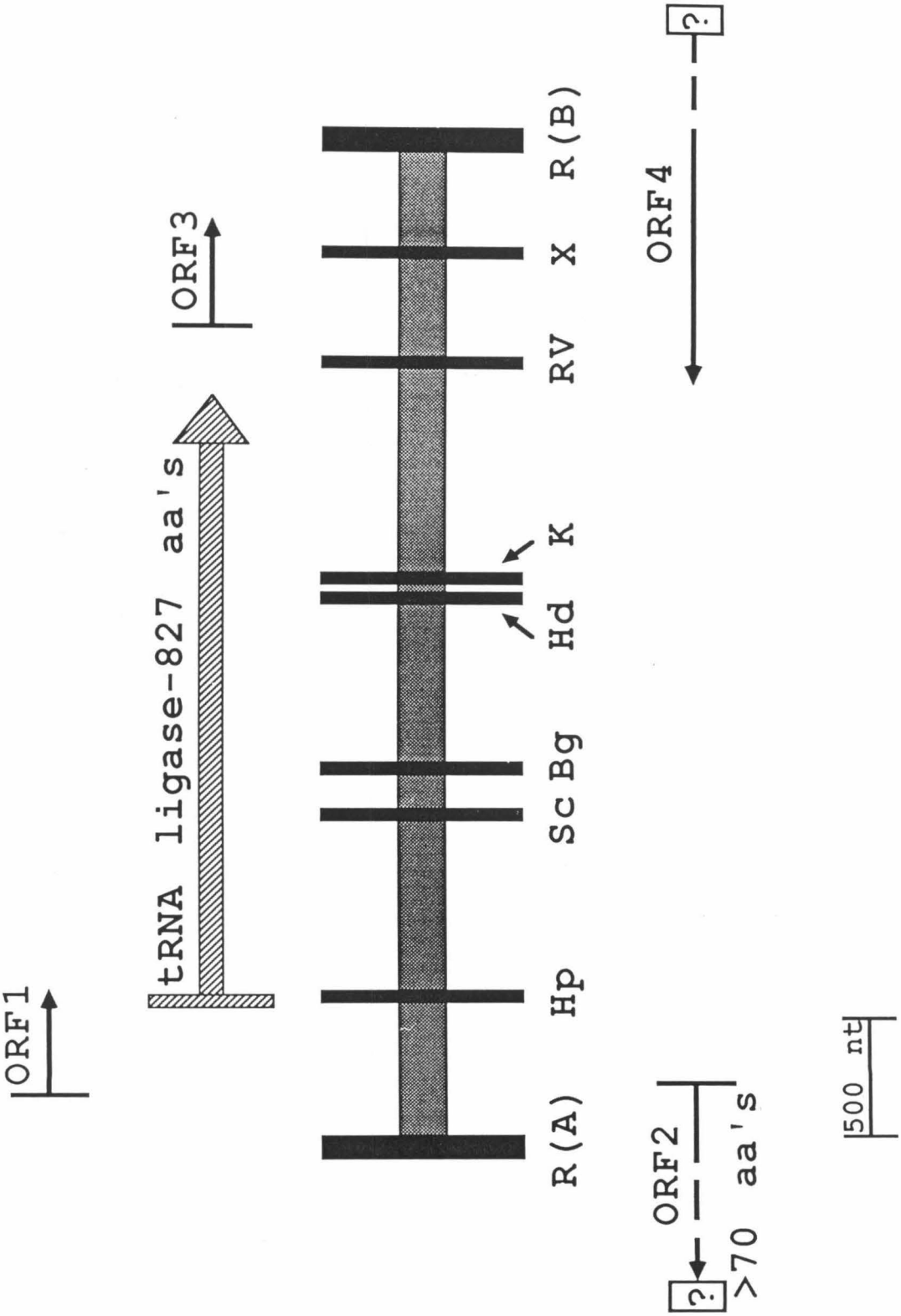
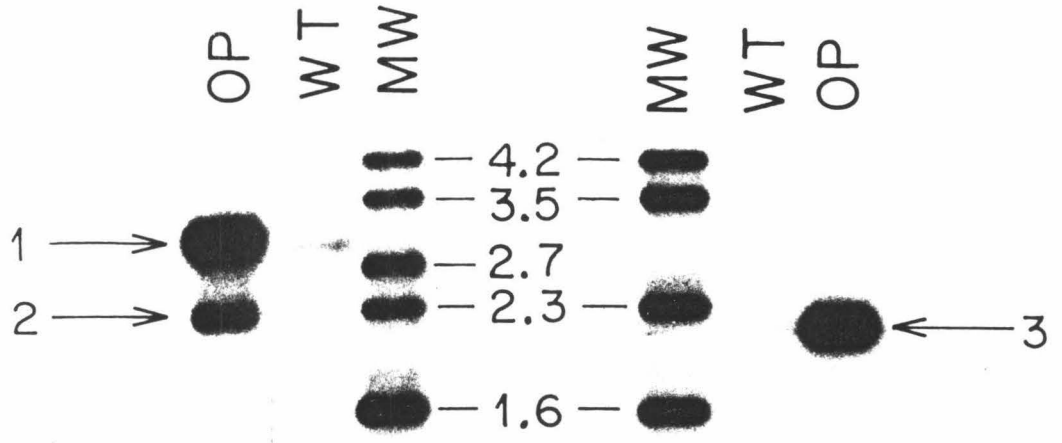

ORF1
tRNA ligase-827 aa's
ORF3
R(A) Hp Sc Bg Hd K RV X R(B)
ORF4
?
ORF2
?
>70 aa's
500 nt

Figure 4



A. tRNA Ligase B. ORF2

C. Transcript Map

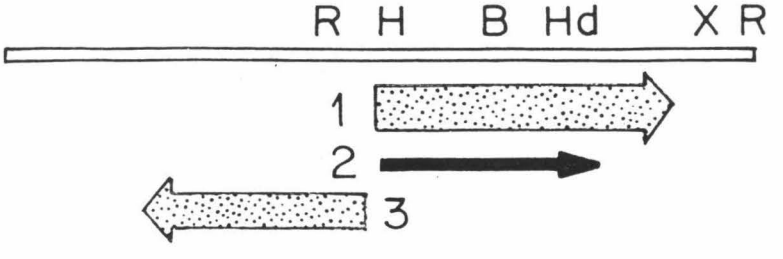


Figure 5

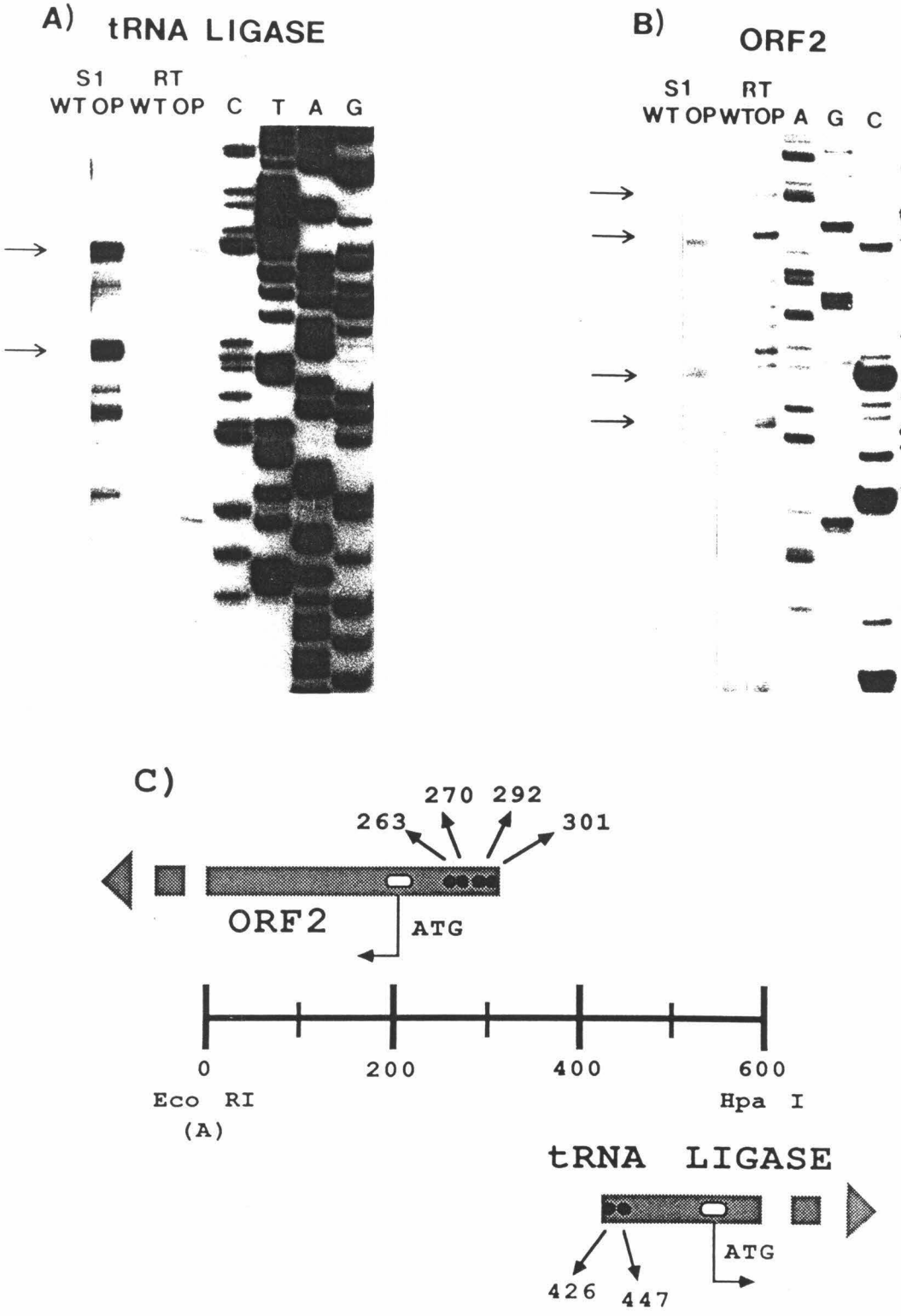

A) tRNA LIGASE
S1 RT
WT OP WT OP C T A G
B) ORF2
S1 RT
WT OP WT OP A G C T
C)
263
270
292
301
ORF2
ATG
0
200
400
600
Eco RI
(A)
Hpa I
tRNA LIGASE
ATG
426
447

Figure 6

CHAPTER II

Open Reading Frames Near the tRNA Ligase Gene

RATIONALE

As previously reported, there are four open reading frames (ORF1, ORF2, ORF3, and ORF4) in proximity to the yeast tRNA ligase open reading frame (Westaway *et al.*, 1988). Detailed studies of the characteristics of these open reading frames were undertaken to further elucidate the requirements for tRNA ligase activity. Figure 1 is a diagram depicting their locations with respect to the tRNA ligase coding region, as well as their directions and predicted length in amino acids.

Comparison of the amino acid sequence deduced from ORF4 with that of the yeast *ARG3* gene (Huygen *et al.*, 1987) revealed that these genes are identical. The termination codons for tRNA ligase and ORF4 (*ARG3*) are separated by 186 nucleotides (Huygen *et al.*, 1987; Westaway *et al.*, 1988) and the genes do not overlap. This 186 nucleotide region at the 3'-ends of both genes is A:T rich. The previous determination that the gene for *ARG3* was localized to yeast chromosome X (Hilger *et al.*, 1982) and its close association with tRNA ligase now permit the latter gene to be assigned a chromosomal location.

ORF3 has not as yet been characterized, however, it is unlikely to be transcribed or translated in yeast. This open reading frame is on the opposite strand from ORF4 (*ARG3*). Two complementary strands of DNA do not often code for two divergent genes in yeast, as they do in *E. coli*. Therefore ORF3 was not characterized further. Instead, we have concentrated our efforts on investigating the two open reading frames 5' to the ligase gene, ORF1 and ORF2.

ORF1. ORF1 was of interest due to its proximity to the 5'-end of the tRNA ligase coding region. This ORF is 122 amino acids in length (see Figure 1), and its start codon is on the same strand approximately 360 nucleotides upstream from the start codon of tRNA ligase. Its termination codon lies 8 nucleotides downstream of the ligase start codon. Although the

two open reading frames overlap, ORF1 is out-of-frame with respect to the tRNA ligase reading frame. An expanded diagram of these two overlapping reading frames is shown in Figure 2. This overlap between the two open reading frames suggested that their activities might be related.

At approximately the same time as ORF1 was discovered, Clark and Abelson (1987) showed that the tRNA ligase protein is localized to the yeast cell nucleus. Potential nuclear localization signals were not found within the amino acid sequence of the tRNA ligase gene itself. It was initially thought that ORF1 might provide this signal. (However, more recent sequence comparisons suggest there may be two different potential nuclear localization signals in tRNA ligase, one very near the beginning of the coding region but not within ORF1, and these are currently under investigation (Mark Berlin, personal communication).)

Another possibility was that ORF1 was translated as a frameshift fusion protein to tRNA ligase. Frameshift translations are not unknown in yeast, as was demonstrated for the *Ty* elements in yeast (Clare and Farabaugh, 1985; Mellor *et al.*, 1985). As a frameshift fusion, ORF1 could be part of a signal to localize tRNA ligase to the yeast cell nucleus. At that time, these signals included short sequences within the SV40 large T antigen (Kalderon *et al.*, 1984), the yeast alpha-2 protein (Hall *et al.*, 1984), and the yeast histone H2B (Moreland *et al.*, 1987). Therefore the predicted amino acid sequence of ORF1 was searched for similarities to these nuclear localization signals. However, no similarities with these sequences were discovered in ORF1.

Yeast tRNA ligase protein behaves during purification from yeast cells as a soluble or peripheral membrane protein (Peebles *et al.*, 1983; Greer *et al.*, 1983). Therefore, there was also the possibility that ORF1, translated as a frameshift fusion to tRNA ligase, was acting as a membrane transfer signal which could be cleaved post-translationally after delivery of ligase to the nucleus, much like the signal leader sequences delivering newly translated polypeptides to the endoplasmic reticulum. Alternatively, ORF1 could be a

transmembrane domain fused to ligase to anchor it near the nuclear membrane, possibly being lost in the purification procedure of ligase from yeast cells. Two transplantation antigens of the murine major histocompatibility complex, mouse H-2L^d (Goodenow *et al.*, 1982; Moore *et al.*, 1982) and mouse H-2K^b (Uehara *et al.*, 1980a; Uehara *et al.*, 1980b) of the mouse histocompatibility complex, contain transmembrane domains. Therefore, the predicted ORF1 amino acid sequence was analyzed for its hydrophobicity (using Versions 4.2 and 4.5 of the Intelligenetics PEP Hydropathicity programs of the BIONET™ National Computer Resource for Molecular Biology), and the profile was compared to that generated for the transmembrane domains of the mouse proteins. The plots of the mouse antigens were very similar to the hydrophobicity profile of the carboxyl-terminus of ORF1, which is near the overlap of the tRNA ligase gene. A comparison of these hydrophobicity profiles is shown in Figure 3. Because of these intriguing characteristics of the predicted ORF1 sequence, experiments were conducted to determine if ORF1 coded for RNA or protein *in vivo* (see *Results, ORF1*).

ORF2. ORF2 begins 342 nucleotides upstream from the beginning of the tRNA ligase coding region and proceeds in the opposite direction (Westaway *et al.*, 1988; see Figure 1). This ORF produces a polyadenylated RNA transcript of approximately 2.1 kilobases that begins only 125 nucleotides away from the tRNA ligase transcription initiation site (Westaway *et al.*, 1988). This proximity of the two transcription initiation sites suggested that tRNA ligase and ORF2 might share transcriptional regulatory signals. Very few other divergently transcribed genes in yeast are as closely situated with respect to their transcription initiation sites as are ORF2 and tRNA ligase. The possibility that ligase and ORF2 shared regulatory signals led to the speculation that the cellular function of ORF2 might somehow be related to tRNA ligase, i.e., ORF2 might have some function in tRNA metabolism.

The proximity of ORF2 to tRNA ligase and the possibility that they shared

regulatory signals or functions in the cell led to experiments which would determine: 1) Is ORF2 an essential yeast gene?; 2) What are the *in vivo* genetic complementation characteristics of ORF2 and tRNA ligase?; 3) Does ORF2 perform a function in the yeast cell related to tRNA processing?; and 4) What is the nucleotide sequence and predicted amino acid sequence of ORF2 and does it possess similarity to any other genes? These questions are addressed in *Results, ORF2*.

EXPERIMENTAL PROCEDURES

Plasmids. Plasmid pYSW11, containing the *E. coli* beta-galactosidase gene fused to the first fifteen amino acids of yeast tRNA ligase was constructed as follows. Plasmid pSEY8 was constructed by Scott Emr and is a derivative of yIP5 (Struhl *et al.*, 1979) containing the yeast 2 micron fragment and the polylinker from pUC8 (Vieira and Messing, 1982). Plasmid pEP9B was constructed by Eric Phizicky and contains the 4.2 kilobase *Eco*RI fragment of pUC-RLG (Phizicky *et al.*, 1986) inserted in the *Eco*RI site of pSEY8. The vector portion of pEP9B digested with *Bam*HI and *Nhe*I was ligated to the *Bam*HI-*Nhe*I gene fragment containing the tRNA ligase and ORF2 genes from pEP108 to create pYSW9. Plasmid pEP108 was constructed by Eric Phizicky and is an approximately 7-kilobase *Xba*I-*Xba*I fragment of pEP99 (described in Westaway *et al.*, 1988) inserted into the *Xba*I site of pUC12 (Vieira and Messing, 1982). Plasmid pSEY101, provided by Scott Emr, has a 1015-amino acid fragment of *E. coli* beta-galactosidase (missing the first 9 amino acids, including the initiation codon, replaced with an *Eco*RI-*Sma*I-*Bam*HI polylinker (Casadaban *et al.*, 1980) inserted in pSEY8. The blunt *Sma*I - *Nru*I fragment of beta-galactosidase from pSEY101 was inserted into pYSW9 in frame at a unique *Hpa*I site fifteen amino acids downstream of the start codon of tRNA ligase to create pYSW11.

Figure 4 presents a schematic of the cloning procedure to create the ochre mutation in the predicted open reading frame of ORF1. Plasmid pSW30d was constructed by

cleaving pUC12-RLG (Phizicky *et al.*, 1986) with *Xba*I and religating to remove the multiple cloning site downstream of tRNA ligase. M13 phage mpSW162 contains the ochre mutation at amino acid 76 of the ORF1 sequence, and is derived from site-directed oligonucleotide mutagenesis of M13 mpSW116, and M13mp11 derivative (Messing, 1983) containing the 595 base pair *Eco*RI(A) - *Hpa*I fragment of the 4.2 kilobase *Eco*RI fragment containing ORF1 and the tRNA ligase gene (described in Westaway *et al.*, 1988). Plasmid pEP108 is described above.

Plasmid pYSW18 was created by inserting the *Xba*I fragment of pEP99 (Westaway *et al.*, 1988) into the *Nhe*I site of yIP5 (Struhl *et al.*, 1979). Plasmid pYSW20 is the plasmid containing the ORF2/*Xho*I disruption and was created by cleaving pYSW18 at a unique *Xho*I site within ORF2, filling in the restriction site overhang, ligating, and transforming to *E. coli* JM101 (Messing, 1979).

Plasmids created by Nuclease BAL31 digestion, diagrammed in the middle portion of Figure 10, were created by cleaving pEP108 (described above) at a unique *Sal*I site downstream of the ORF2 gene, treating with Nuclease BAL31 for various times, cleaving at the unique *Xba*I site downstream of the tRNA ligase gene, and inserting the resulting different-sized fragments into the *Xba*I and blunt *Sma*I sites of pSEYC58. Plasmid pSEYC58 was constructed by Scott Emr and is a derivative of yCP50 (Johnston and Davis, 1984) containing the pUC8 polylinker, CENIV, ARS1, and URA3 yeast genes.

Plasmids created by Nuclease BAL31 digestion, diagrammed in the bottom portion of Figure 10, were first prepared by cleaving pEP108 (described above) at a unique *Sma*I site downstream of the tRNA ligase gene, treating with Nuclease BAL31 for various times, cleaving at the *Sal*I site downstream of ORF2, and inserting the resulting different-sized fragments into the vector pSEY8. Plasmid pSEY8 was first prepared to accept the blunt-ended Nuclease BAL31 ends by cutting with *Bam*HI, filling in the restriction site overhang, and cleaving with *Sal*I. The resulting set of plasmids (pYSW21-24) were

digested with *Sma*I and *Bam*HI and the different fragments were inserted into the *Sma*I and *Bam*HI sites of pSEYC58.

Plasmid pYSW93, which contains the intact ORF2 gene and the tRNA ligase gene missing the *Kpn*I fragment, was constructed by inserting the *Bam*HI - *Xba*I fragment of p298-1 (Phizicky *et al.*, 1992) in the identical sites of pSEYC58.

Plasmid pYSW101A was constructed as follows. pYSW22 is one of the Nuclease BAL31-generated plasmids described above which contains approximately 680 base pairs of DNA upstream of the first *Eco*RI site in ORF2 (see Figure 10). The *Sma*I - *Bam*HI fragment of pYSW22 was inserted in the identical sites of pUC18 (Yanisch-Perron *et al.*, 1985) to create pSW82. A *Hin*dIII - *Sal*I fragment from pBM150 (Johnston and Davis, 1984) containing the yeast *GAL10* promoter was filled in and inserted at a unique *Nae*I site of pSW82 to create pSW86. The *Sac*I - *Bam*HI fragment of pSW86 containing the *GAL10* promoter fused 58 nucleotides upstream of the first ORF2 transcription start site, along with the entire ORF2 gene, was inserted at the *Sac*I and *Bam*HI sites of pSEYC68 to create pYSW101A. Plasmid pSEYC68 is a derivative of pSEYC58 containing the pUC18 polylinker, and was provided by Scott Emr.

Strains. EMPY20 (*his3-Δ200 ade2-101° ura3-52 GAL⁺ suc2*) is the diploid yeast strain constructed by Eric Phizicky by mating SS328 (*MATα his3-Δ200 lys2-801ª ade2-101° ura3-52 GAL⁺ suc2*) to SS330 (*MATa his3-Δ200 tyr1 ade2-101° ura3-52 GAL⁺ suc2*) (Westaway *et al.*, 1988). Strains SS328 and SS330 were obtained from S. Scherer. SWY484A is EMPY20 with plasmid pYSW100 integrated at the ORF1 locus (see Figure 6). SWY497 has been described previously (Westaway *et al.*, 1988). SWY506A-D, 507A-D, and 508A-D are the haploid yeast strains resulting from sporulation of SWY497. SWY333 is EMPY20 *ORF2⁺/orf2⁻-∇XhoI* and is represented as "*3B8*" in Figure 8. SWY515 (*MATα his3-Δ200 lys2-801ª ade2-101° ura3-52 GAL⁺ suc2 orf2⁻-∇XhoI*

(pYSW101A)) is a haploid yeast strain resulting from the sporulation of SWY333 transformed with pYSW101A.

Bacterial strain *E. coli* JM101 (Messing, 1979) was used for all recombinant plasmid and M13 phage manipulations.

Bacterial and DNA Manipulations. Restriction enzyme digests were performed according to the manufacturer's instructions. Nuclease BAL31 digestions, bacterial transformations, plasmid ligations, filling in of restriction site overhangs, and plasmid preparations from bacterial cells were performed according to Maniatis *et al.* (1982). DNA sequencing was by the method of Sanger *et al.* (1977).

Mutagenic oligonucleotides were provided by S. Horvath (California Institute of Technology), and were synthesized on an Applied Biosystems 380A DNA Synthesizer. Oligonucleotides were purified on 20% polyacrylamide 7 M urea gels, eluted in 1% phenol at 37°C, and ethanol precipitated. Mutagenesis, detection of mutant plaques, and isolation of double-stranded and single-stranded M13 phage were performed essentially as described by Maniatis *et al.* (1982). Briefly, the mutagenic oligonucleotide primer was phosphorylated with T4 polynucleotide kinase and hybridized to mpSW116. Klenow DNA polymerase and deoxyribonucleotides were used to synthesize the complement strand, and T4 DNA ligase and ATP for ligation. Tris-acetate agarose gels were performed to test the conversion of single-stranded M13 phage to the double-stranded species. Plaque lifts were performed on *E. coli* JM101 cells transformed with this construct. The viral DNA was denatured, neutralized, and bound to the filter by baking. The filters were probed with the oligonucleotide containing the mutation at increasing temperatures to identify the positive plaques containing the ochre mutation. Probe was prepared by phosphorylation with T4 polynucleotide kinase and [γ-^{32}P]ATP, followed by separation of the ^{32}P-labeled oligonucleotide from unincorporated label on a DE52 column. Positive plaques were plaque purified and retested as above, and double-stranded and single-

stranded phage were prepared. Restriction site analysis of the double-stranded DNA and sequencing of single-stranded DNA confirmed the presence of the mutation.

Yeast Methods. Transformations of competent yeast cells were essentially by the method of Ito *et al.* (1983). *In vivo* gene replacement by transformation and treatment with 5-fluoroorotic acid was by the method of Boeke *et al.* (1984). Yeast cell growth and sporulation media were as described in Sherman *et al.* (1982). When yeast strains were grown in 2% galactose, 2% glycerol and 2% ethanol were added to supplement the carbon source, and glucose-free galactose was used (Phizicky *et al.*, 1992).

Small scale yeast extracts were prepared according to Phizicky *et al.* (1992) and used to test splicing activity of yeast tRNA ligase. Splicing assays were performed according to Greer *et al.* (1983). The preparation of rabbit anti-tRNA ligase antibody has been described previously (Phizicky *et al.*, 1986). Antibody against ORF1 was prepared from a New Zealand White rabbit after injection of a synthetic peptide corresponding to the first fifteen amino acids of the predicted ORF1 sequence (Phizicky *et al.*, 1986). Spheroplasts were prepared by resuspending yeast cell pellets in 0.1 M phosphate citrate buffer (pH 5.8), 1.2 M Sorbitol, and 1 mg/ml Zymolyase. Spheroplasts were stored at -70°C. SDS-polyacrylamide gels were performed according to Laemmli (1970). Proteins were visualized by silver-staining (Wray *et al.*, 1981) or were transferred to nitrocellulose and probed with rabbit anti-ORF1 or rabbit anti-ligase antibody, or anti-beta-galactosidase antibody, conjugated to horseradish peroxidase, using BioRad Laboratories antibody staining kits.

Genomic yeast DNA was prepared from yeast spheroplasts by the method of Davis *et al.* (1980). Southerns were performed according to Maniatis *et al.* (1982) using GeneScreen™ Hybridization Transfer Membranes (NEN Research Products, Boston, Mass.) according to the manufacturer's directions. Southern probes were prepared from double-stranded DNA fragments eluted from low melting point agarose Tris-acetate gels

after treatment of the appropriate plasmid with restriction enzymes. Probes were labeled by the method of Feinberg and Vogelstein (1984) using mixed sequence hexadeoxynucleotides $pd(N)_6$ (Pharmacia), deoxyribonucleotides, [α-^{32}P]dCTP, and Klenow DNA polymerase. Tetrad dissections were performed using a 1:30 dilution of Glusulase (NEN).

^{32}P-Orthophosphate Labeling of Yeast Cells and RNA Extraction. SWY515 and *rna1* cells were labeled with -orthophosphate essentially as described by Ogden *et al.* (1979). Briefly, SWY515 cells were grown in galactose-glycerol-ethanol, harvested, washed, and diluted into low phosphate media containing glucose or galactose, and allowed to multiply for one generation (~2.5 hours) before labeling with 4 mCi ^{32}Pi. Yeast strain *rna1* cells were grown at 23°C in glucose, harvested, washed, resuspended in low phosphate media containing glucose and allowed to grow for one generation (~3 hours). Half of the *rna1* cells were equilibrated to 37°C (nonpermissive temperature) and then both the 23°C and 37°C cultures were labeled with ^{32}Pi. Labeled RNAs were extracted from cells using the hot phenol method, separated on 10% polyacrylamide 4 M urea gels, and autoradiographed (Knapp *et al.*, 1978). The small RNA region of this first dimension gel corresponding approximately to 4-5.8S RNA was cut out and reseparated in the second dimension on 20% polyacrylamide 7 M urea gels and autoradiographed (Knapp *et al.*, 1978).

Materials. [α-^{32}P]dNTPs and [γ-^{32}P]ATP were from Amersham Corp. Deoxyribonucleotides and mixed sequence hexadeoxynucleotides $pd(N)_6$ were from Pharmacia LKB Biotechnology, Inc. Klenow fragment of DNA Polymerase I and T4 polynucleotide kinase were from Bethesda Research Laboratories. Antibody staining kits were from BioRad Laboratories. Restriction enzymes were from New England Biolabs or Boehringer Mannheim. ^{32}P-orthophosphate was from New England Nuclear. T4 DNA ligase was from Boehringer Mannheim.

RESULTS AND DISCUSSION

ORF1. As previously reported (Westaway *et al.*, 1988), Northern hybridization did not detect any polyadenylated RNA from yeast which could correspond to ORF1, either alone or as a fusion to the yeast tRNA ligase transcript. This did not rule out the possibility that ORF1 was transcribed alone and was not polyadenylated. In addition, no polypeptide of the correct size representing a translational frameshift fusion of ORF1 to the yeast tRNA ligase protein was detected using anti-tRNA ligase antibodies to protein purified from yeast (Phizicky *et al.*, 1986). This also did not rule out the possibility that ORF1 was translated from a nonpolyadenylated RNA to produce a protein which contained only the ORF1 122 amino acids.

In collaboration with Michael Clark, a synthetic polypeptide corresponding to amino-terminus of ORF1 was synthesized and used to generate polyclonal antibodies in rabbits. Spheroplasts were prepared from SS328 yeast cells harboring pEP99, a 2 micron plasmid overproducing the tRNA ligase gene (Westaway *et al.*, 1988) and loaded on SDS-polyacrylamide gels. Proteins were transferred to nitrocellulose and probed with the anti-ORF1 antibody. No ORF1 protein was detected using this procedure (data not shown). This negative result could have been due to a very low and undetectable level of ORF1 protein *in vivo*, or the amino acids of ORF1 which would be recognized by the antibody may have been hidden in the predicted intact protein. Therefore we fused ORF1 upstream of the *E. coli* beta-galactosidase gene to enable the use of a more effective antibody. Plasmid pYSW11 was constructed containing all of the upstream ORF2 gene, all of ORF1, and the first fifteen amino acids of tRNA ligase fused to beta-galactosidase. Yeast strain SS328 was transformed with this plasmid, spheroplasts were prepared and proteins were separated on SDS-polyacrylamide gels. Proteins were transferred to nitrocellulose and probed with anti-beta-galactosidase antibody. Only two bands were detected, one corresponding to the size of beta-galactosidase alone, and one corresponding to the size of

beta-galactosidase with an additional 15 amino acids from tRNA ligase (data not shown). No protein was detected in the size range of beta-galactosidase plus the 122 amino acids of ORF1. However, these results did not preclude the possibility that very small amounts of ORF1 were translated alone *in vivo* and were perhaps responsible for some other required function unrelated to tRNA ligase.

To determine if translation of ORF1 was essential in yeast, a nonsense mutation was constructed in ORF1 by site-directed oligonucleotide mutagenesis in M13, using a synthetic oligonucleotide of the sequence: 5'-AAGTA**TAA**AGAAGAGCA-3'. This mutation changes a tyrosine (TAC) residue at amino acid 76 of the ORF1 open reading frame to an ochre stop codon (TAA) and also deletes an *Acc*I restriction enzyme cleavage site. It was necessary to make a relatively small disruption in this region, in this case only a one nucleotide change, in order to avoid disturbing either the ORF2 or tRNA ligase potential transcription signals. This mutated oligonucleotide was annealed, extended, ligated, and transformed to *E. coli* strain JM101 competent cells. Double-stranded M13 was rescued which contained the mutation as detected by *Acc*I restriction site analysis (data not shown). Single stranded M13 was also isolated and sequenced to insure the presence of the mutation (data not shown). An *Eco*RI to *Bcl*I DNA fragment containing the mutation was then substituted into a plasmid containing the entire tRNA ligase gene as well as all upstream DNA containing the ORF2 gene, and transferred to a yIP5 vector (Struhl *et al.*, 1979). Figure 4 is a diagram of the cloning procedure. The final vector, pYSW100, was cleaved once within the ORF2 gene at a unique *Xho*I restriction site, and used to transform competent EMPY20 yeast cells (Phizicky *et al.*, 1992).

Transformants were selected on media minus uracil and genomic DNA was prepared. DNA was digested with *Acc*I, separated on a 1% TBE agarose gel, transferred to GeneScreen™ and probed with a fragment corresponding to the *Eco*RI to *Hpa*I region encompassing the ORF1 gene. Figure 5, panel *A* is an autoradiograph of a representative yeast clone containing the mutation integrated in the chromosome at the ORF1 locus.

The strain containing the mutation (SWY484A) was then grown on 5-fluoroorotic acid to force loss by recombination of the *URA3*$^+$ gene, accompanying yIP5 DNA, and the extra copy of the ORF2-ORF1-ligase DNA sequence (Boeke *et al.*, 1984). Genomic DNA was prepared from diploids which were now *ura*$^-$ and reanalyzed for the presence of the lesion. Figure 5, panel *B* is an autoradiograph of a representative yeast clone (SWY497) containing one wild type and one mutated chromosome. This diploid was sporulated and tetrads were dissected. Figure 6 is a photograph of a dissection plate and indicates that 10 of 10 tetrads dissected from SWY497 segregated 4 live and none dead. These haploids all demonstrated the correct segregation of markers (data not shown) and genomic DNA was prepared from each set. Figure 7 is an autoradiograph of a Southern blot with representative tetrad sets (SWY506A-D, 507A-D, and 508A-D), indicating that in each tetrad, two of the four live spores still retained the mutation. The results of the ochre mutation indicate that ORF1 is not an essential protein in yeast. Taken together with the lack of polyadenylated RNA detected for ORF1 (Westaway *et al.*, 1988), these results suggest that ORF1 is probably not produced in yeast, either as an RNA or a protein, or if it is, it is not an essential gene.

ORF2. The initial analysis of ORF2 function was undertaken in collaboration with Eric Phizicky. These studies were propelled by an early result of unusual genetic complementation with a chromosomal interruption in the yeast tRNA ligase gene. These results at first indicated that portions of the ligase coding region might be necessary for the production of ORF2, and are described below.

As previously reported (Abelson *et al.*, 1986), a disruption of the yeast tRNA ligase gene was made by cloning the *HIS3*$^+$ gene into a unique *Bgl*II site in the middle of the ligase gene. This disrupted gene was transformed into a *his*$^-$ diploid, integrated into the chromosomal tRNA ligase locus, and sporulated. Dissection of the resulting tetrads yielded two live and two dead spores, with the live spores all *his*$^-$. This indicated that such

a disruption of the only chromosomal copy of tRNA ligase in a haploid yeast cell was lethal to the cell (Abelson *et al.*, 1986). The *RLG*$^+$/*rlg*::*HIS3*$^+$ diploid was then transformed with the 4.2 kilobase *Eco*RI fragment containing the entire yeast tRNA ligase gene (see Figure 1), either on a single copy *URA3*$^+$ plasmid (CEN) or a multicopy *URA3*$^+$ plasmid (2 micron). The transformants were selected on media lacking uracil and sporulated. The intact tRNA ligase gene on either plasmid failed to complement the disruption of the chromosomal copy of ligase (data not shown), even when the tRNA ligase gene was placed under the control of the *GAL10* promoter in order to induce overproduction of the protein (see below). However, assays for ligase activity *in vitro* using yeast extracts from the parental transformed diploids indicated that ligase activity was elevated 5-25 times over wild type activity, and protein gels hybridized to anti-tRNA ligase antibody demonstrated the presence of a correct size tRNA ligase polypeptide. Only when at least 2.3 kilobases of DNA upstream from the tRNA ligase gene was included on the plasmid was complementation obtained. Figure 8 is a schematic of these complementation results (Eric Phizicky, unpublished data). The 2.3 kilobases of DNA upstream from tRNA ligase necessary to complement the ligase disruption was approximately the same size as the 2.1 kb transcript of ORF2. This result led us to believe that tRNA ligase was indeed being produced from the smaller plasmids, but that the *HIS3*$^+$ disruption at the chromosomal locus of tRNA ligase was deleterious to the upstream gene, ORF2.

A different disruption of the yeast tRNA ligase gene was constructed using two unique *Kpn*I sites, deleting 1371 nucleotides (or 457 amino acids) from the center of the ligase coding region (Phizicky *et al.*, 1992). A yeast diploid with this disruption in one chromosomal homologue was transformed with a plasmid containing the ligase gene under the control of the *GAL10* promoter, in which the *GAL10* promoter is fused near the beginning of the tRNA ligase coding region, omitting DNA upstream of the ligase gene (Phizicky *et al.*, 1986). When this strain was sporulated, complementation by this plasmid was obtained (Phizicky *et al.*, 1992), leading to the hypothesis that the original *HIS3*$^+$

disruption was moving some region of the ligase gene too far downstream, possibly interfering with the transcription of ORF2. Alternatively, the *Kpn*I deletion disruption was moving this essential region closer, and therefore not interfering in the production of ORF2.

Due to these unusual complementation results, it was necessary to determine if ORF2 was an essential gene in yeast. To accomplish this, a small disruption in the ORF2 coding region was constructed by cleaving at a unique *Xho*I site mapped within the length of the ORF2 RNA transcript, filling in the restriction site overhang, and religating. Again, this type of small disruption was preferred, as a larger change may have affected the upstream tRNA ligase gene. This disruption added an additional four nucleotides within the putative ORF2 coding region, presumably throwing the coding region out of frame by one nucleotide, as well as destroying the unique *Xho*I site. Although the sequence of ORF2 was not known when this mutation was made, it was anticipated that it should have created a new stop codon downstream or created an altered amino acid sequence which presumably would no longer be capable of carrying out the ORF2 cellular function. This alteration is now known to have created a stop codon 8 amino acids downstream of the *Xho*I site, thus shortening the length of the ORF2 polypeptide from 623 amino acids to 297 (Komatsoulis *et al.*, 1987).

The mutant ORF2 sequence was cloned into the yIP5 vector (Struhl *et al.*, 1979), and the resulting plasmid (pYSW20) was linearized at a unique *Nhe*I site located within the tRNA ligase gene and transformed into EMPY20 diploid yeast cells. Transformants were selected on media minus uracil, and genomic DNA was prepared. DNA from eight transformants was digested with *Eco*RV and *Xho*I and Southern analysis was performed to verify integration at the ORF2 locus (*lanes 1-8*, Figure 9). These diploids were subjected to 5-fluoroorotic acid treatment, and genomic DNA was prepared from the resulting *ura*$^-$ diploids and redigested with *Eco*RV and *Xho*I. Southern analysis was performed to determine which diploids retained the *Xho*I mutation. Figure 9 is an autoradiograph of the

DNA from diploids before and after 5-fluoroorotic acid treatment. These diploid yeast were sporulated, and those retaining the mutation segregated 2 live:2 dead (data not shown). These tetrad sets correspond to the parent diploids from the lanes labeled *1E5*, *3B8* (SWY333, see below), *4E2*, and *6B9* in Figure 9. Those tetrad sets not retaining the mutation segregated 4 live:0 dead (data not shown), and these tetrads correspond to the parent diploids from the lanes labeled *2C1*, *5D4*, *7C6*, and *8B3* in Figure 9. These results confirm that ORF2 is an essential gene in yeast.

A yeast strain defective in ORF2 could now be tested *in vivo* using plasmid complementation to determine how much ORF2 and tRNA ligase sequence is necessary to restore life. In collaboration with Eric Phizicky, a set of plasmids was constructed and transformed to a yeast diploid with the ORF2 mutation in one chromosomal homologue (SWY333). These plasmids were created by exonuclease BAL31 treatment of an approximately 7-kilobase *Xba*I DNA fragment containing both the tRNA ligase and ORF2 genes. Yeast diploids harboring the chromosomal ORF2 mutation and deletions of ORF2 and/or ligase sequence on single copy *URA3*$^+$ centromere plasmids were sporulated, and live spores were analyzed for the correct segregation of markers in the strain itself, or on a plasmid, if present (data not shown). Strains lacking sufficient plasmid sequences to complement the ORF2 chromosomal lesion segregated 2 live:2 dead, while those with an activity complementing the ORF2 mutation segregated 4 live:0 dead, 3 live:1 dead, or 2 live:2 dead, depending on the segregation of the plasmid's *URA3*$^+$ marker. These results are diagrammed in Figure 10.

Nucleotide numbering in Figure 10 corresponds to the distance upstream or downstream of the first *Eco*RI site in ORF2. The first set of complementation results indicate that at least 1700 base pairs downstream from the *Eco*RI site in ORF2 are necessary to complement the ORF2 disruption. The beginning of ORF2 is 204 nucleotides upstream of the *Eco*RI site (Komatsoulis *et al.*, 1987). Therefore approximately 1900 nucleotides of ORF2 are necessary to complement the chromosomal disruption. This

represents all of the ORF2 open reading frame as well as approximately 40 nucleotides of downstream 3'-untranslated DNA (Komatsoulis *et al.*, 1987). A plasmid with only 1560 nucleotides downstream of the *Eco*RI site, or approximately 1700 nucleotides of ORF2, is insufficient to complement the ORF2 disruption. This plasmid is missing about 150 nucleotides, or 50 amino acids, of the carboxy-terminus of the coding region of ORF2.

The second set of complementation results are depicted at the bottom of Figure 10. These results indicate that 680 nucleotides upstream of the *Eco*RI site, or approximately 470 nucleotides upstream of the ORF2 start codon, are not sufficient to rescue the ORF2 disruption phenotype. It should be noted that this noncomplementing region contains all four of the transcription start sites mapped for ORF2 (Westaway *et al.*, 1988). Some part or all of the stretch of DNA between 680 and 1680 nucleotides upstream of the *Eco*RI site is necessary in "cis" on the plasmid to restore life, and this region is contained entirely within the coding region of tRNA ligase. To narrow down the required nucleotides, plasmid pYSW93 containing all of ORF2 and tRNA ligase but missing the center *Kpn*I fragment containing 457 amino acids of tRNA ligase was used to test complementation of the ORF2 disruption. This plasmid contains up to 937 nucleotides upstream of the *Eco*RI site (same numbering as in Figure 10), and does restore life (data not shown). This result indicated that a region between nucleotides 680 and 937, or what corresponds to amino acids 44 through 130 of the tRNA ligase coding region, is sufficient to restore complementation of the ORF2 disruption. At this time, it is unknown what, if any, transcriptional regulatory signals might be contained within this region necessary for ORF2 production. Additional evidence has shown that a plasmid construct containing the ORF2 coding region fused to the *GAL10* promoter 58 nucleotides upstream of the first ORF2 transcription start site (see below) is also sufficient to complement the ORF2 chromosomal disruption when grown on galactose media, but not on glucose (data not shown). It is assumed that this construct is overproducing the ORF2 transcript when grown on

galactose, and the *GAL10* promoter may be taking the place of a possible transcription regulatory signal in the 680-937 nucleotide region of tRNA ligase.

Because of the close proximity of the transcriptional start sites of ORF2 and tRNA ligase and the unusual genetic complementation results described above, it was of interest to determine if the cellular function of ORF2 was related in any way to tRNA processing. A haploid yeast strain containing the chromosomal ORF2 disruption and harboring a suppressible copy of the wild-type ORF2 gene was needed. This strain would be used to detect defective tRNAs made *in vivo* after suppression of transcription of the plasmid-borne gene. Strain SWY515 contains a single copy centromere plasmid with the wild-type ORF2 sequence under control of the *GAL10* promoter (pYSW101A). Growth curves on glucose or galactose media were first performed for SWY515, in comparison with EMPY439, a similar strain containing a chromosomal tRNA ligase disruption harboring the wild-type tRNA ligase gene under control of the *GAL10* promoter (Phizicky *et al.*, 1992). This growth curve is shown in Figure 11 and indicates that cells deficient in ORF2 production cease multiplying on glucose in fewer generations than the 6-8 generations needed for tRNA ligase deficient cells (Phizicky *et al.*, 1992).

To ascertain whether a subset of incompletely processed tRNAs (i.e., precursor tRNAs or tRNA half molecules) accumulated in an ORF2-deficient strain, SWY515 was grown in galactose, shifted to either galactose or glucose media, and labeled with [^{32}P]-orthophosphate. RNA was extracted and separated by two-dimensional gel electrophoresis and the resulting autoradiographs were compared to those resulting from the treatment of temperature sensitive *rna1* cells at the nonpermissive temperature, which accumulate precursor tRNAs containing introns when incubated at the nonpermissive temperature (Hopper *et al.*, 1978; Knapp *et al.*, 1978).

Figure 12, panels *a* and *b*, are autoradiographs of *rna1* cells at the permissive and nonpermissive temperatures. Several spots are seen to accumulate in the upper portion of the autoradiograph (see arrow) in the RNA extracted from cells treated at 37°C (*b*), but not

from cells grown at 23°C (*a*). These spots have been previously identified as precursor tRNAs containing introns (Knapp *et al.*, 1978). However, no such accumulation is seen for SWY515 cells on either galactose or glucose (Figure 12, panels *c* and *d*), indicating that ORF2 is probably not a component of the yeast tRNA endonuclease and may not play any role in tRNA processing.

The function of ORF2 *in vivo* is still unknown, however, ORF2 has been completely sequenced and contains an open reading frame 1869 nucleotides long, coding for a putative protein of 623 amino acids, or 71.3 kDa (Komatsoulis *et al.*, 1987). ORF2 probably does not contain an intron, as the sequence does not contain the canonical "TACTAAC" box presumed necessary for mRNA splicing in yeast (Komatsoulis *et al.*, 1987). The ORF2 amino acid sequence was not found to be significantly homologous to any protein in the NBRF database (National Biomedical Research Foundation protein database of the BIONET™ National Computer Resource for Molecular Biology) (Komatsoulis *et al.*, 1987). Sequences yet to be reported for essential yeast genes, or future sequences of proteins from yeast or other organisms could reveal a potential function for the essential ORF2.

ACKNOWLEDGMENTS

I am tremendously grateful to Eric Phizicky and Elizabeth Grayhack for instructing me in yeast genetics, protein biochemistry, analysis of RNA, and the labyrinthine methods of plasmid construction. I would also like to acknowledge Jennifer Normanly and Marjorie Strobel for their instruction in bacterial, plasmid, and M13 phage procedures, and Lianna Johnson for helpful tips on yeast tetrad preparation and dissection. I would like to thank Michael Clark for his assistance in cellular biology and protein biochemistry methods, and Margaret Saks for critical reading of this manuscript. This work was supported in part by American Cancer Society Grant ACS NP302D (to John Abelson) and by National Research

Service Award T32 GM07616 from the National Institute of General Medical Sciences (to S. W.)

REFERENCES

Abelson, J.N., Brody, E.N., Cheng, S.-C., Clark, M.W., Green, P.R., Dalbadie-McFarland, G., Lin, R.-J., Newman, A.J., Phizicky, E.M., and Vijayraghavan, U. (1986) *Chemica Scripta* **26B**, 127-137

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984) *Mol. Gen. Genet.* **197**, 345-346

Casadaban, M.J., Chou, J., and Cohen, S.N. (1980) *J. Bact.* **143**, 971-980

Clare, J., and Farabaugh, P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2829-2833

Clark, M.W., and Abelson, J. (1987) *J. Cell Biol.* **105**, 1515-1526

Davis, R.W., Thomas, M., Cameron, J., St. John, T.P., Scherer, S., and Padgett, R.A. (1980) *Methods Enzymol.* **65**, 404-411

Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267

Goodenow, R.S., McMillan, M., Örn, A., Nicolson, M., Davidson, N., Frelinger, J.A., and Hood, L. (1982) *Science* **215**, 677-679

Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

Hall, M.N., Hereford, L., and Herskowitz, I. (1984) *Cell* **36**, 1057-1065

Hilger, F., Prévot, M., and Mortimer, R. (1982) *Curr. Genet.* **6**, 93-98

Hopper, A.K., Banks, F., and Evangelidis, V. (1978) *Cell* **14**, 211-219

Huygen, R., Crabeel, M., and Glansdorff, N. (1987) *Eur. J. Biochem.* **166**, 371-377

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* **153**, 163-168

Johnston, M., and Davis, R.W. (1984) *Mol. Cell. Biol.* **4**, 1440-1448

Kalderon, D., Richardson, W.D., Markham, A.F., and Smith, A.E. (1984) *Nature* **311**, 33-38

Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhrman, S.A., and Abelson, J. (1978) *Cell* **14**, 221-236

Komatsoulis, G.A., Westaway, S.K., and Abelson, J.N. (1987) *Nucleic Acids Res.* **15**, 9079

Laemmli, U.K. (1970) *Nature* **227**, 680-685

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J. (1985) *Nature* **313**, 243-246

Messing, J. (1983) *Methods Enzymol.* **101**, 20-78

Messing, J. (1979) *Recomb. DNA Tech. Bull.* **2**, 43-48

Moore, K.W., Sher, B.T., Sun, Y.H., Eakle, K.A., and Hood, L. (1982) *Science* **215**, 679-682

Moreland, R.B., Langevin, G.L., Singer, R.H., Garcea, R.L., and Hereford, L.M. (1987) *Mol. Cell. Biol.* **7**, 4048-4057

Ogden, R.C., Beckman, J.S., Abelson, J., Kang, H.S., Söll, D., and Schmidt, O. (1979) *Cell* **17**, 399-406

Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986) *J. Biol. Chem.* **261**, 2978-2986

Phizicky, E.M., Consaul, S.A., Nehrke, K.W., and Abelson, J. (1992) *J. Biol. Chem.* **267**, 4577-4582

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467

Sherman, F., Fink, G. R., and Hicks, J. B. (1982) in *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Struhl, K., Stinchcomb, D. T., Scherer, S., and Davis, R. W. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1035-1039

Uehara, H., Ewenstein, B.M., Martinko, J.M., Nathenson, S.G., Coligan, J.E., and Kindt, T.J. (1980a) *Biochemistry* **19**, 306-315

Uehara, H., Ewenstein, B.M., Martinko, J.M., Nathenson, S.G., Kindt, T.J., and Coligan, J.E. (1980b) *Biochemistry* **19**, 6182-6188

Vieira, J., and Messing, J. (1982) *Gene* **19**, 259-268

Westaway, S.K., Phizicky, E.M., and Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171-3176

Wray, S., Boulikas, T., Wray, V. P., and Hancock, R. (1981) *Anal. Biochem.* **118**, 197-203

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103-119

FIGURE LEGENDS

FIG. 1. **Open Reading Frames Near the tRNA Ligase Gene.** Open reading frames coding for potential proteins of greater than 150 amino acids are shown as linear arrows (not to scale) above and below the linear representation of the yeast genomic 4.2-kilobase *Eco*RI fragment containing the tRNA ligase gene (from Westaway *et al.*, 1988). The designations (*ORF1, ORF2, ORF3,* and *ORF4(ARG3)*) are shown above each arrow (*ORF4* is the yeast *ARG3* gene, see *Rationale*) and their potential length in amino acids (*aa's*) is shown below. Relevant restriction enzyme recognition sites are depicted to provide points of reference.

FIG. 2. **Enlargement of the Overlap Between the Predicted ORF1 Sequence and the tRNA Ligase Gene.** The nucleotide sequence of the region of overlap between ORF1 and tRNA ligase is depicted with small brackets above and below indicating the codons and their corresponding amino acids for both open reading frames. *OPEN READING FRAME* indicates ORF1; *tRNA LIGASE* indicates the tRNA ligase open reading frame. The large black arrow indicates the initiation codon for tRNA ligase.

FIG. 3. **Hydrophobicity Profiles of ORF1 vs. Murine Transplantation Antigens.** A hydrophobicity profile (see text) is shown for the predicted ORF1 amino acid sequence compared to two transplantation antigens of the mouse histocompatibility complex, mouse H-2L^{d} (Goodenow *et al.*, 1982; Moore *et al.*, 1982) and mouse H-2K^{b} (Uehara *et al.*, 1980a and 1980b). Only the profile of the transmembrane domains of the mouse antigens are shown, and are raised two units above the origin line for clarity. Hydrophobic regions are plotted above the axis and hydrophilic regions below the axis. *5'-OPEN READING FRAME* designates the profile for ORF1; symbols are as follows: (•), ORF1; (o), mouse H-2K^{b}; (**x**), mouse H-2L^{d}.

FIG. 4. **Diagram of the Cloning Procedure for the ORF1 Ochre Mutation.** A diagram of the cloning procedure for pYSW100 is presented. Arcs filled in black correspond to fragments of or an intact tRNA ligase gene. Arcs with cross hatching indicate an intact ORF2 gene. Arcs with light stippling correspond to the putative ORF1 reading frame, and the "*X*" within this arc represents the ochre mutation described in the text. Pertinent genes in the yIP5 and pYSW100 vectors are indicated as follows: *AMP*, ampicillin resistance gene of *E. coli* (beta-lactamase); *TET*, tetracycline resistance gene; *URA3*, yeast URA3$^+$ marker; and *T*, the disrupted TET gene in pYSW100 conferring tetracycline sensitivity for selection of the insert in *E. coli*. Pertinent restriction sites in the plasmids are indicated by the following abbreviations: *E*, *Eco*RI; *AccI*, *Acc*I; *Bc*, *Bcl*I; *Xb*, *Xba*I; *Hp/Hc*, the *Hpa*I/*Hin*cII junction of M13 phage mpSW116 (see *Experimental Procedures*); *Na*, *Nae*I; *Nh*, *Nhe*I; *B*, *Bam*HI; and *Xh*, *Xho*I. Restriction digests are indicated by the letters inside the bridges between the arrows in the flowchart. The construction of plasmid pSW30d and M13 phage mpSW162 is described in *Experimental Procedures*. Double-stranded mpSW162 and pSW30d were digested as indicated and the *Eco*RI - *Bcl*I fragment of mpSW162 containing the ochre mutation ("*X*") and deleting the *Acc*I site was inserted in the vector portion of pSW30d to create pSW54. The construction of pEP108 is described in *Experimental Procedures*. Plasmids pSW54 and pEP108 were digested as indicated and the *Nae*I - *Nhe*I fragment of pSW54 was inserted in the vector portion of pEP108 to create pSW99. Plasmids pSW99 and yIP5 were digested with *Bam*HI and the fragment of pSW99 containing the ochre mutation was inserted in the tetracycline resistance gene of yIP5 to create pYSW100.

FIG. 5. **Southern Hybridization of Genomic DNA From Strains with an Ochre Mutation in ORF1.** Panel *A* shows genomic DNA isolated from a strain (*SWY484A*) transformed with an integrating plasmid containing the ochre mutation before treatment with 5-fluoroorotic acid. The DNA was digested with *Acc*I, separated on an

agarose gel, and hybridized to a probe spanning the region of the ochre mutation (see *Experimental Procedures*). Panel *B* shows genomic DNA from SWY484A after treatment with 5-fluoroorotic acid (*SWY497*). Approximate lengths of *Acc*I-digested DNA fragments are shown at left. Controls in both panels are as follows: *pEP108* (*lane 1, panels A and B*), plasmid DNA containing the wild type sequence indicates the migration of two expected *Acc*I fragments; *pYSW100* (*lanes 2 and 3, panels A and B*), plasmid DNA containing the ochre mutation indicates migration of an *Acc*I fragment which is the sum of the two fragments in *lane 1*, and indicates the deletion of the *Acc*I site at the ochre mutation; *EMPY20* (*lane 4, panels A and B*), genomic DNA from the parent strain without the ORF1 mutation; *SWY484A* (*lane 5, panel A*), genomic DNA containing the wild type ORF1 sequence on both chromosomal homologues as well as the extra copy of the ORF1 locus containing the ochre mutation; *SWY497* (*lane 5, panel B*), genomic DNA containing the wild type ORF1 sequence on one chromosomal homologue and retaining the ORF1 ochre mutation with the *Acc*I site deleted on the other homologue after treatment with 5-fluoroorotic acid.

FIG. 6. **Segregation of Spores From SWY497 Tetrads.** A photograph of a representative tetrad dissection plate is shown for diploid strain SWY497 containing one wild type and one mutated ORF1 chromosomal locus. The patch at the top is SWY497 taken from a sporulation plate and treated with glusulase to dissolve the encapsidated tetrads. Ten tetrads were dissected top to bottom and indicate a segregation for the mutation of four live spores.

FIG. 7. **Southern Hybridization of Haploid Strains From the Sporulation of SWY497.** Three representative tetrad sets from the plate in Figure 6 were analyzed for the presence of the ORF1 mutation (*SWY506A-D, SWY507A-D*, and *SWY508A-D*). Genomic DNA was isolated, digested with *Acc*I, separated on an agarose gel, and probed

with the appropriate DNA fragment (see *Experimental Procedures*). Approximate sizes of *Acc*I DNA fragments are shown to the right. Controls (*EMPY20* and *pYSW100*) are the same as for Figure 5. For each tetrad set, two haploids still retain the ochre mutation in ORF1.

FIG. 8. **Plasmid Complementation of a *HIS3* Chromosomal Disruption of Yeast tRNA Ligase.** A diagram of the genetic complementation results of a *RLG*$^{+}$/*rlg::HIS3*$^{+}$ chromosomally disrupted strain is shown (Abelson *et al.*, 1986). The *HIS3* gene insert at the *Bgl*II site of the yeast tRNA ligase chromosomal locus is shown at the top. The 4.2 kilobase *Eco*RI fragment is the line with hash marks at each end and the tRNA ligase open reading frame is indicated below with a solid black circle indicating the ligase start codon, followed by an arrow for the length of the polypeptide. Vectors are described as either yeast single copy plasmids (*CEN*), yeast multicopy plasmids (*2μ*), or a single copy plasmid with the tRNA ligase open reading frame fused to the yeast *GAL10* upstream activation sequence (*CEN + GAL UAS*). Splicing activity of diploids with the *HIS3* disruption containing either the 4.2 kilobase *Eco*RI fragment on a 2μ plasmid or a CEN plasmid with ligase fused to the GAL UAS is indicated as 5-fold (*5X*) or 25-fold (*25X*) over wild type ligase activity. For these two strains, the presence of the correct size ligase polypeptide is indicated, as measured by Western blot. Hash marks on lines in the lower section of the figure indicate the presence or absence of additional genomic DNA upstream of the yeast tRNA ligase gene, including all or part of the ORF2 open reading frame. Complementation by each plasmid as determined by sporulation of the respective diploid strain is indicated in the far right column. *POOR(GALACTOSE)* indicates that growth on galactose of spores from the tetrad is much slower than wild type (data not shown). When the strain is sporulated on glucose (*NO(GLUCOSE)*), which represses the GAL-controlled plasmid copy of tRNA ligase, the *CEN + GAL UAS* plasmid does not rescue the chromosomal disruption.

FIG. 9. **Southern Hybridization of Diploid Strains Containing a Disruption in ORF2.** Genomic DNA was isolated from EMPY20 transformed with an integrating plasmid containing a 4 base pair insertion at the *Xho*I site in ORF2 (see *Experimental Procedures*). DNA was digested with *Xho*I and *Eco*RV, separated on an agarose gel, and probed with the appropriate DNA fragment (see *Experimental Procedures*). Approximate sizes of hybridizing fragments are shown at the right. The ~2680 nucleotide band corresponds to an *Eco*RV - *Xho*I fragment of ORF2; the ~4080 nucleotide band corresponds to a larger *Eco*RV - *Eco*RV fragment of ORF2 resulting from the elimination of the *Xho*I recognition site. Controls (*pYSW100* and *pYSW18*) are plasmid DNA digested with *Eco*RV and *Xho*I. pYSW18 is a plasmid containing the wild type ORF2 sequence and pYSW20 is pYSW18 with the *Xho*I site in ORF2 cleaved, filled in, and religated, destroying the *Xho*I site (see *Experimental Procedures*). Lanes labeled *1-8* are DNA from eight isolates of EMPY20 transformed with the integrating plasmid containing the ORF2 disruption, before treatment with 5-fluoroorotic acid. Lanes labeled *2C1, 5D4, 7C6*, and *8B3* are DNA from four ura$^-$ isolates of the above strain after treatment with 5-fluoroorotic acid, and indicate that the ORF2 mutation has been "popped out," leaving behind a wild type ORF2 sequence on each chromosome. Lanes labeled *1E5, 3B8, 4E2*, and *6B9* are DNA from four ura$^-$ isolates of the above strain after treatment with 5-fluoroorotic acid and show that one chromosomal homologue still retains the ORF2 mutation.

FIG. 10. **Plasmid Complementation of the *Xho*I Disruption of ORF2.** A diagram of the genetic complementation results of an *ORF2*$^+$*/orf2*$^-$-∇*XhoI* strain (SWY333, see text and *Experimental Procedures*) is shown. The four base pair insert at the *Xho*I site of the ORF2 chromosomal locus is shown at the top. The ~7-kilobase genomic DNA fragment containing the ORF2 and tRNA ligase (*RLG*) genes is shown below the arrows indicating the reading frames for each gene, with relevant restriction

sites. All plasmids used for the complementation studies were single copy yeast vectors (*CEN*). Complementation by each plasmid as determined by sporulation of the respective transformed diploid strain is indicated in the column labeled *COMPLEMENTATION?*. *BAL31* brackets indicate in which orientation the exonuclease reaction was allowed to proceed. Numbers above each linear plasmid represent the approximate length in nucleotides left or right from the most 5' *Eco*RI recognition site (with respect to the direction of the gene) in ORF2.

FIG. 11. **Growth Curves for Galactose or Glucose Incubation of SWY515 and EMPY439.** Haploid strains containing chromosomal disruptions of either the ORF2 gene (*SWY515*) or the tRNA ligase gene (*EMPY439*) but harboring plasmids with their corresponding genes under the control of the *GAL10* promoter (see text and *Experimental Procedures*) were grown in galactose or glucose and A_{600} was monitored. Strains were diluted into fresh media in order to maintain log phase growth. Symbols for each plot are identified in the legend at the right of the graph. Legend designations are as follows: *GAL*, growth in galactose; *GLUC*, growth in glucose; *Diln*, first dilution in galactose for both strains; *Diln#1*, first dilution in glucose for both strains; *Diln#2*, second dilution in glucose for EMPY439 only (SWY515 ceased vegetative growth).

FIG. 12. **Two-Dimensional Gels of RNA Extracted From *rna1* and SWY515 Cells.** Autoradiographs of two-dimensional 4-5.8S small RNAs extracted from *rna1* cells grown at the permissive (*23°C*, panel *a*) or nonpermissive temperatures (*37°C*, panel *b*), or SWY515 cells grown in galactose (*GAL*, panel *c*) or glucose (*GLUC*, panel *d*) are shown (see *Experimental Procedures*). *Brackets* indicate migration of *pre-tRNAs* or mature *tRNAs* as identified by Knapp *et al.* (1978).

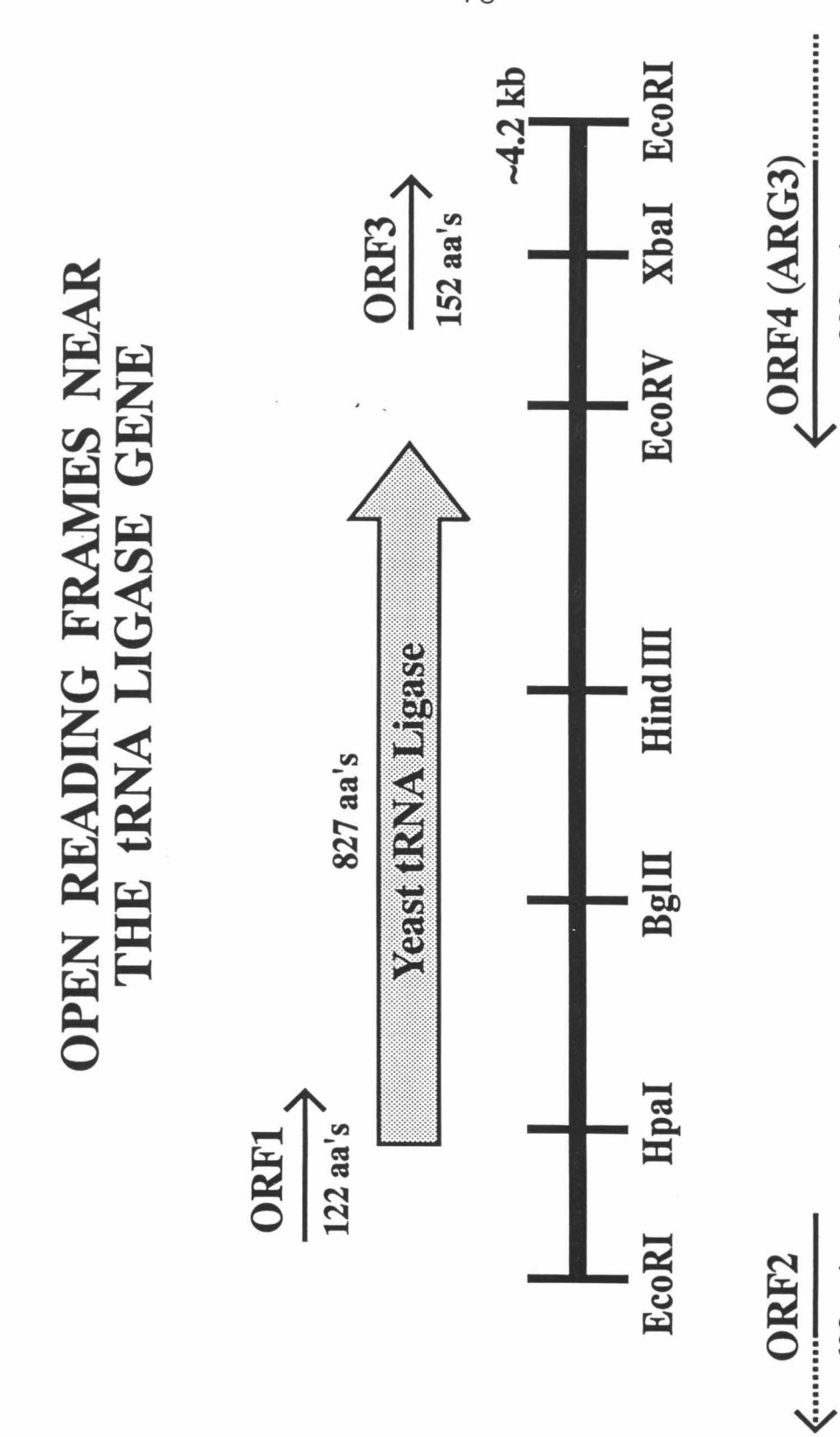

OPEN READING FRAMES NEAR
THE tRNA LIGASE GENE
ORF1
122 aa's
827 aa's
Yeast tRNA Ligase
ORF3
152 aa's
~4.2 kb
EcoRI
HpaI
BglII
HindIII
EcoRV
XbaI
EcoRI
ORF2
623 aa's
ORF4 (ARG3)
338 aa's

Figure 1

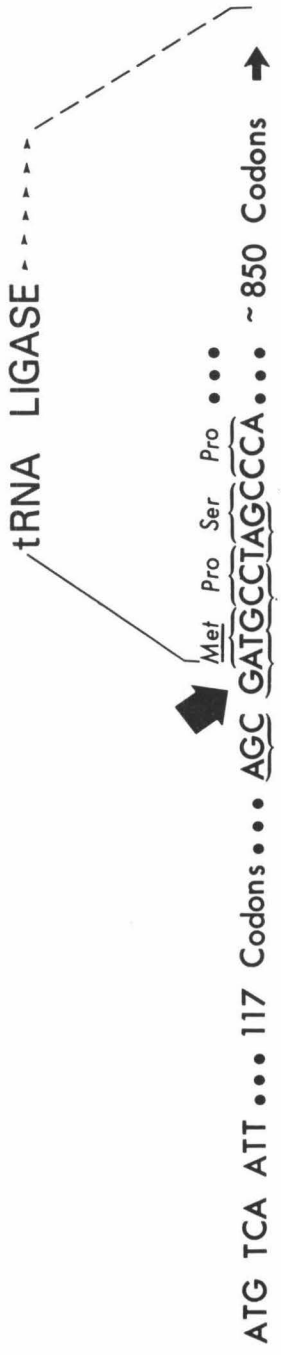

tRNA LIGASE
Met Pro Ser Pro •••
AGC GATGCCTAGCCCA••• ~850 Codons
Ser Asp Ala STOP
Promoter?
ATG TCA ATT ••• 117 Codons •••
Met Ser Ile
OPEN READING FRAME

Figure 2

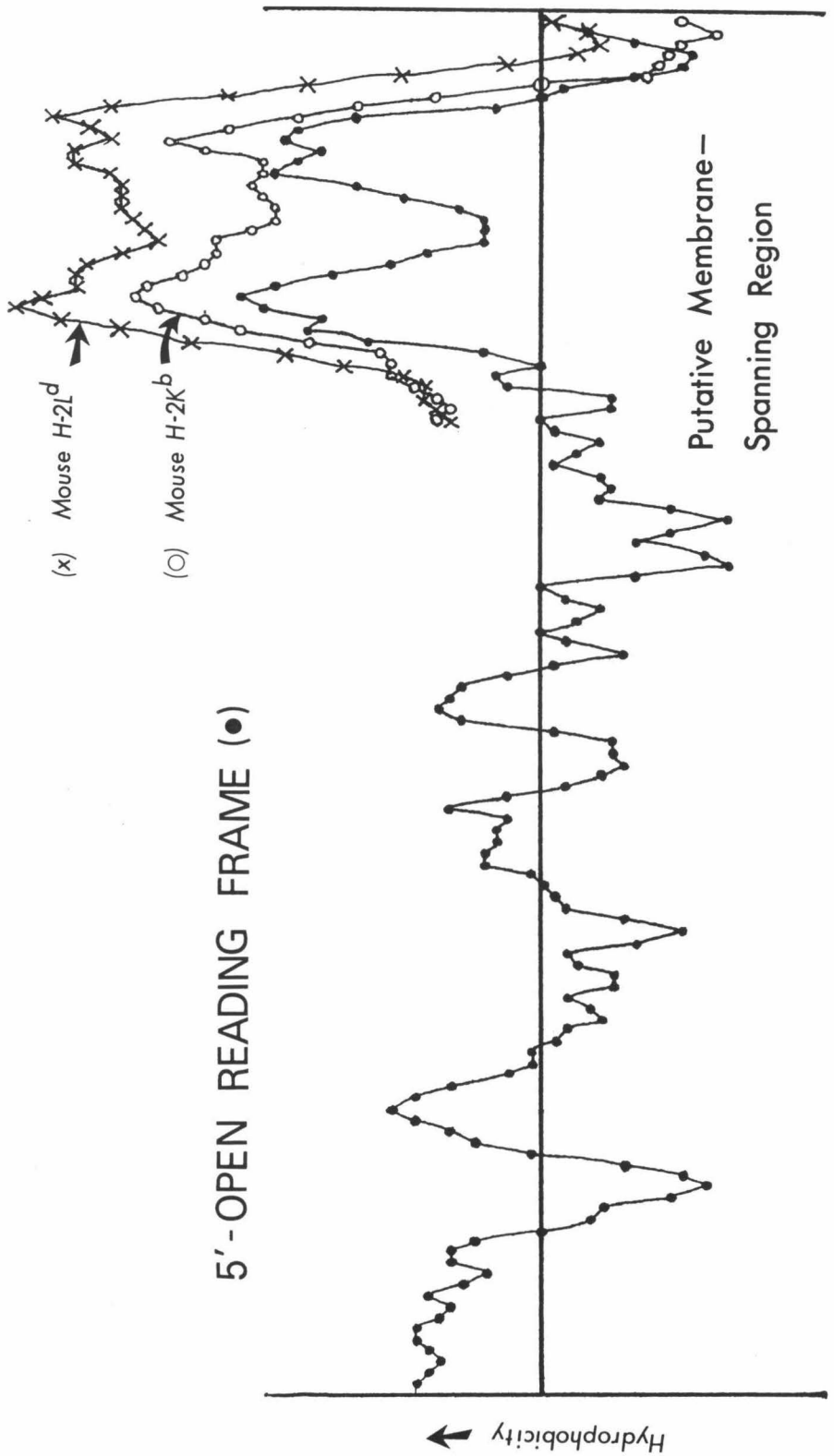

(x) Mouse H-2L^d
(○) Mouse H-2K^b
5'-OPEN READING FRAME (●)
Putative Membrane-Spanning Region
Hydrophobicity

Figure 3

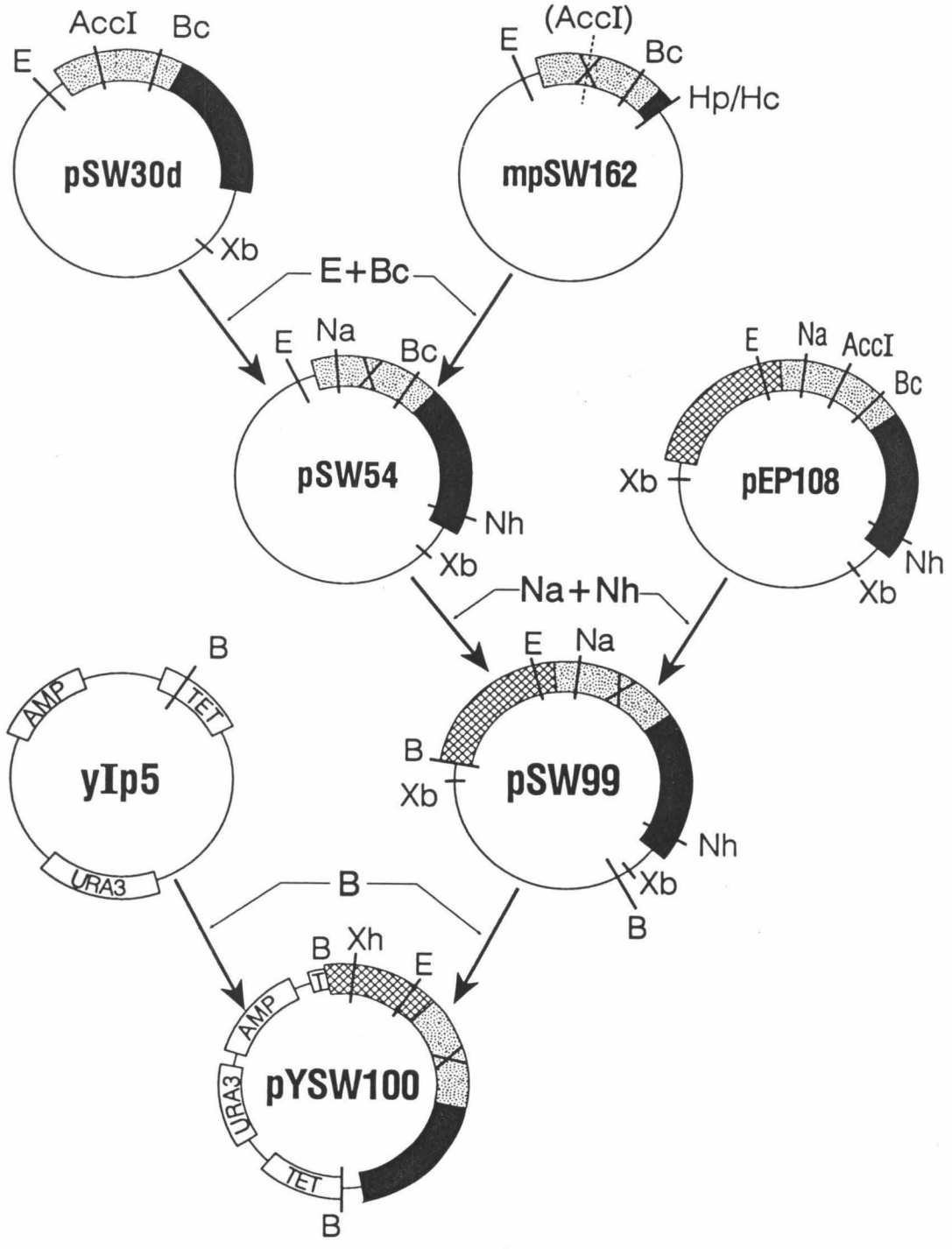

AccI
Bc
E
pSW30d
Xb
(AccI)
E
Bc
Hp/Hc
mpSW162
E+Bc
E
Na
Bc
pSW54
Nh
Xb
E
Na
AccI
Bc
Xb
pEP108
Nh
Xb
Na+Nh
B
AMP
TET
yIp5
URA3
E
Na
B
Xb
pSW99
Nh
Xb
B
B
B
Xh
E
AMP
TET
URA3
pYSW100
B

Figure 4

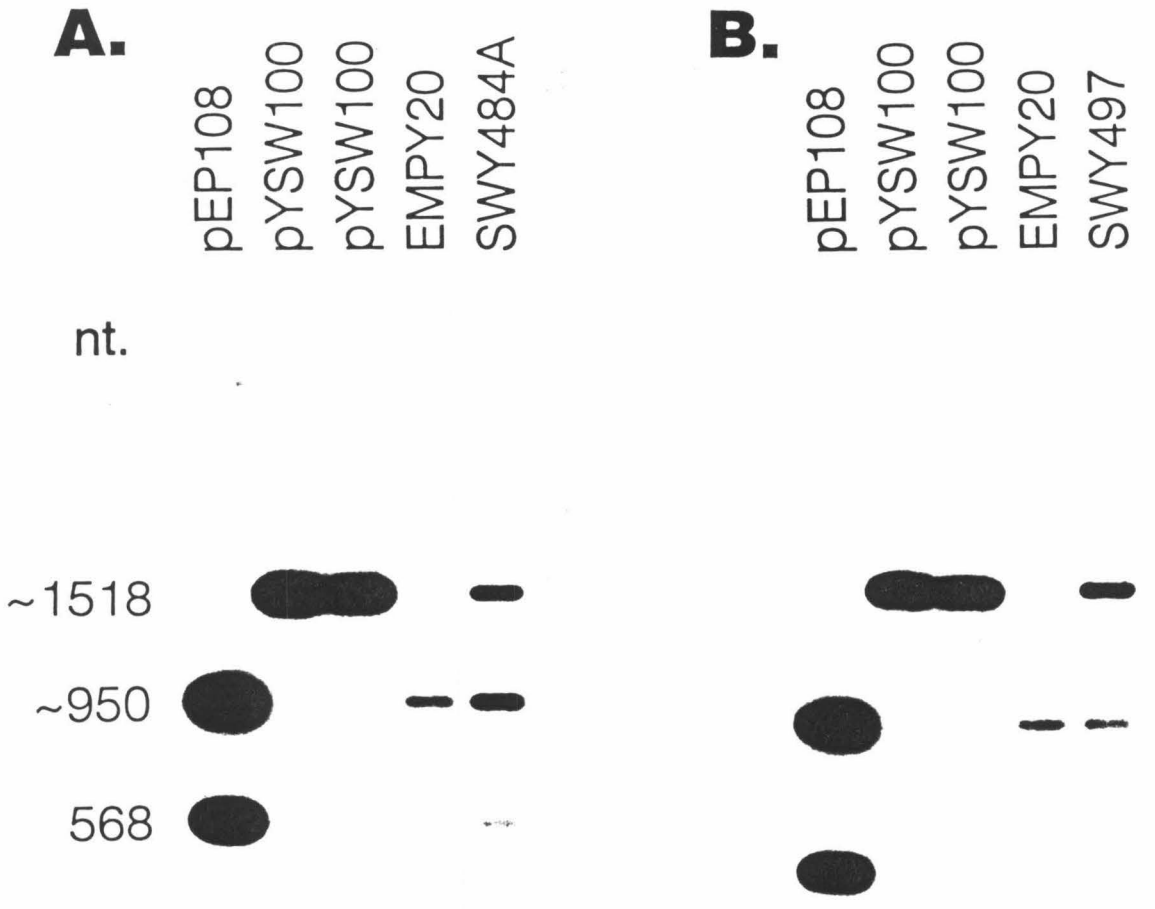

A.
pEP108
pYSW100
pYSW100
EMPY20
SWY484A
nt.
~1518
~950
568
B.
pEP108
pYSW100
pYSW100
EMPY20
SWY497

Figure 5

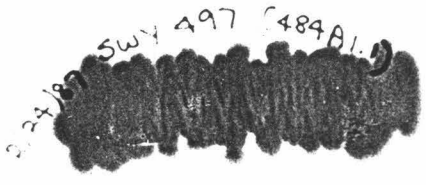


Figure 6

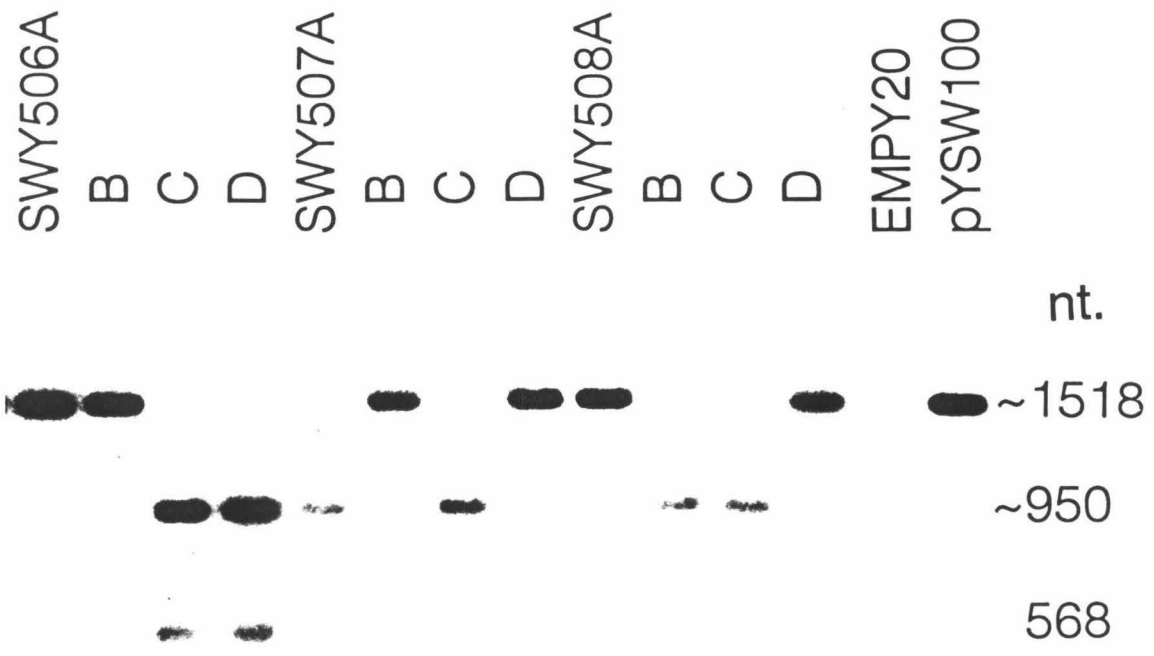

SWY506A
B
C
D
SWY507A
B
C
D
SWY508A
B
C
D
EMPY20
pYSW100
nt.
~1518
~950
568

Figure 7

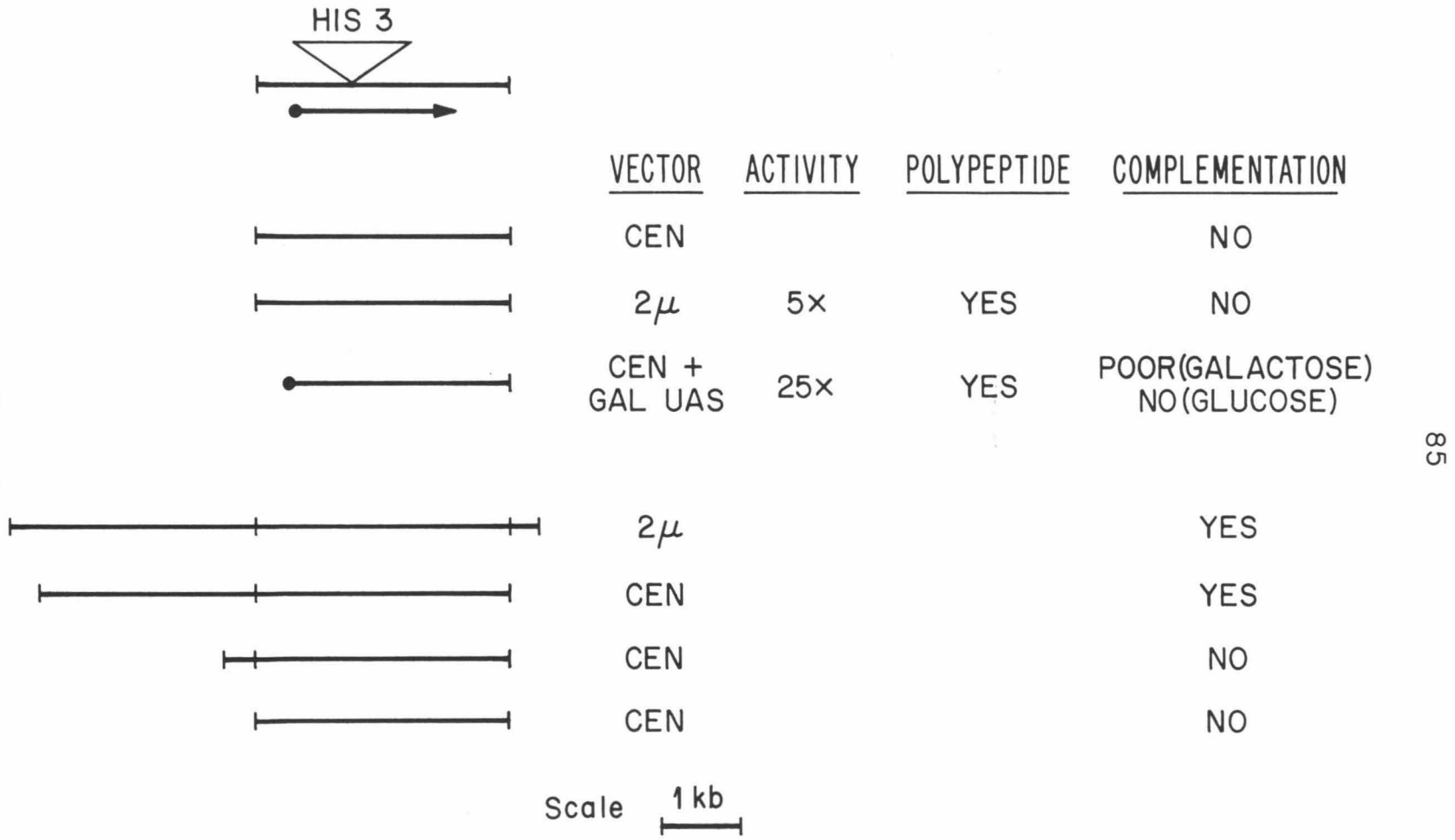

HIS 3
VECTOR
ACTIVITY
POLYPEPTIDE
COMPLEMENTATION
CEN
NO
2μ
5×
YES
NO
CEN +
GAL UAS
25×
YES
POOR(GALACTOSE)
NO(GLUCOSE)
2μ
YES
CEN
YES
CEN
NO
CEN
NO
Scale
1 kb

Figure 8

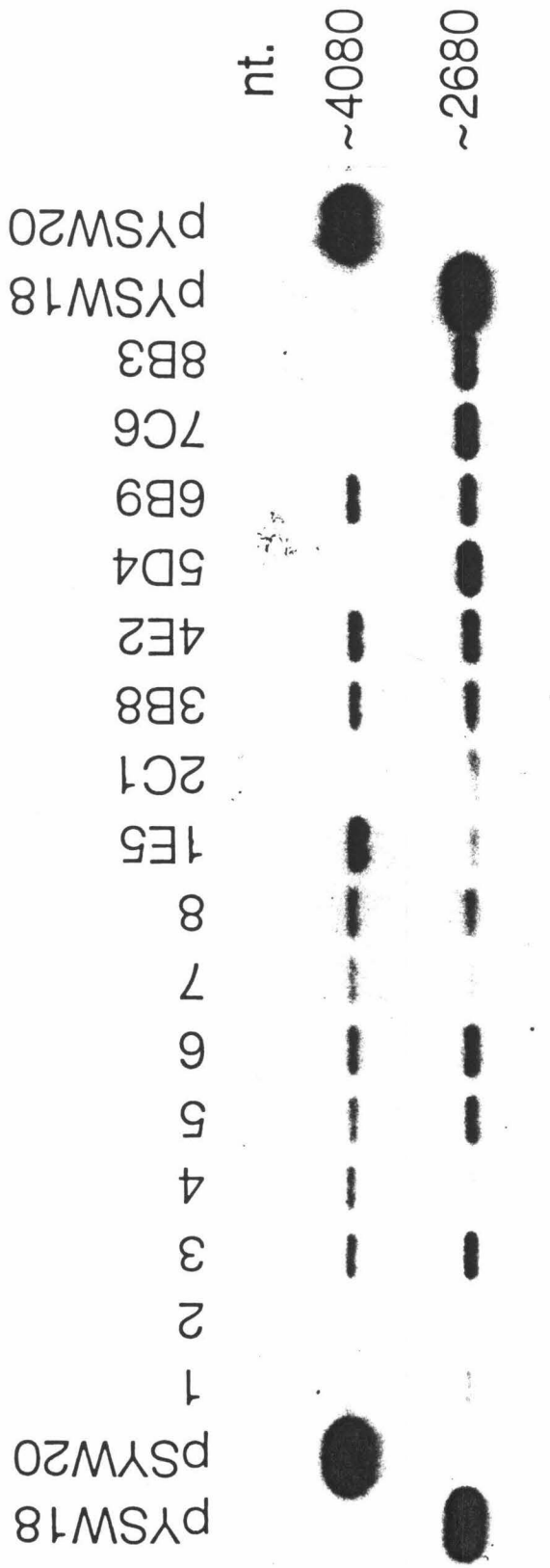

nt.
~4080
~2680
pYSW18
pSYW20
1
2
3
4
5
6
7
8
1E5
2C1
3B8
4E2
5D4
6B9
7C6
8B3
pYSW18
pYSW20

Figure 9

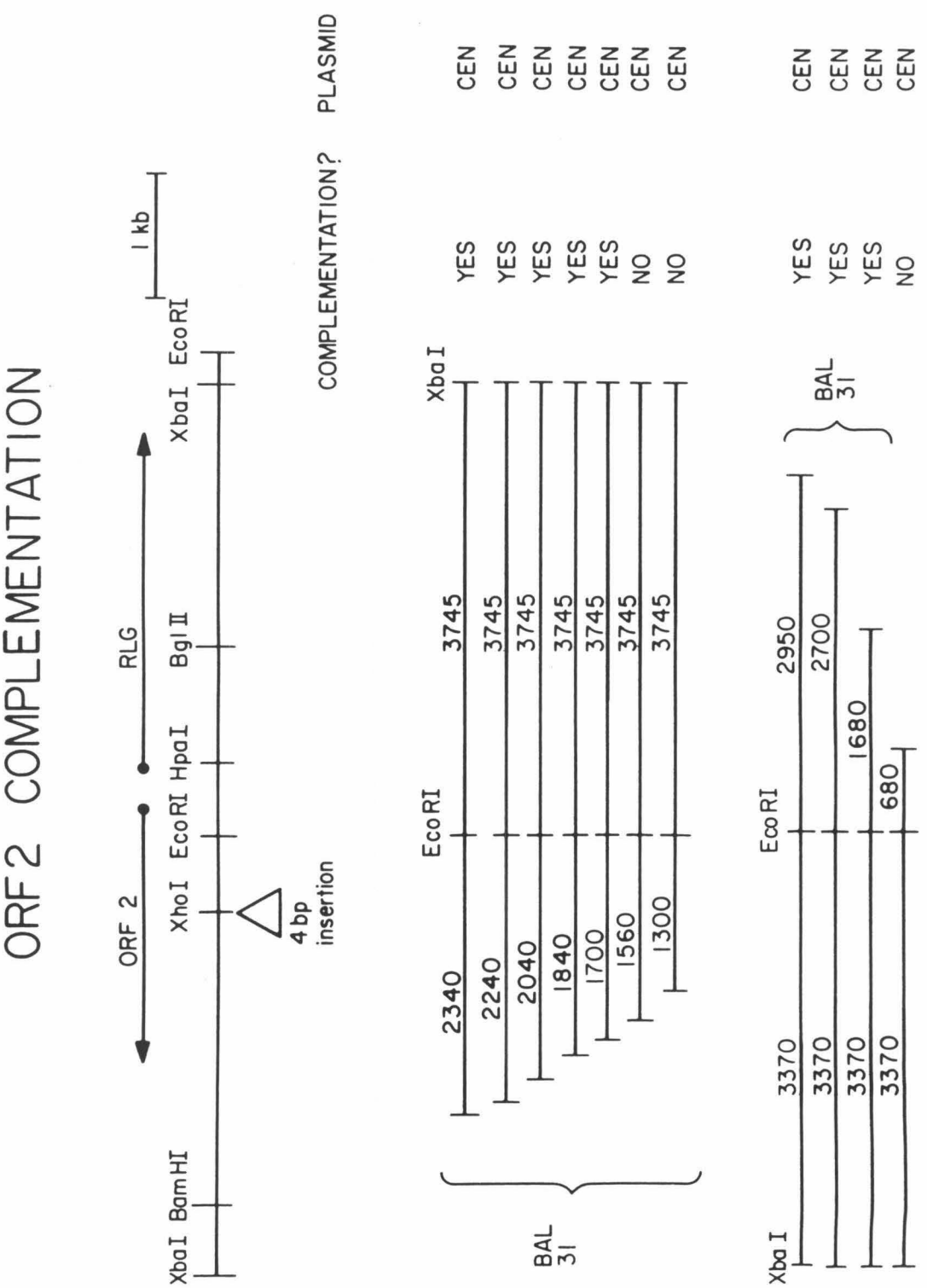

ORF2 COMPLEMENTATION
ORF 2
RLG
1 kb
XbaI BamHI
XhoI EcoRI HpaI
BglII
XbaI EcoRI
4 bp insertion
COMPLEMENTATION?
PLASMID
XbaI
EcoRI
BAL 31
2340
2240
2040
1840
1700
1560
1300
3745
YES
NO
CEN
EcoRI
XbaI
BAL 31
3370
2950
2700
1680
680
YES
NO
CEN

Figure 10

GROWTH CURVES: GALACTOSE TO GLUCOSE SHIFT

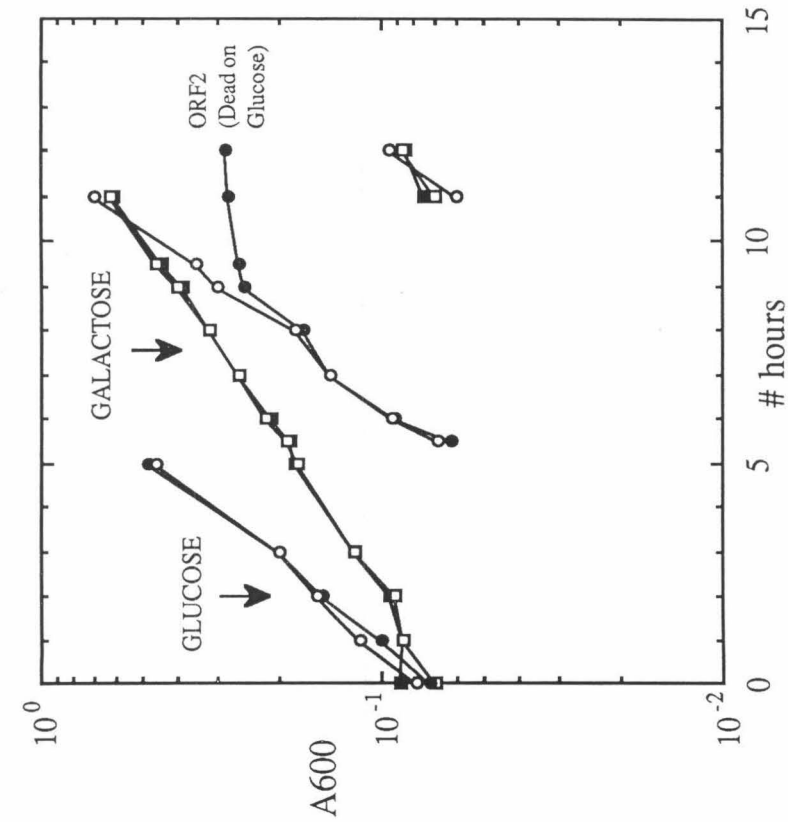


Figure 11

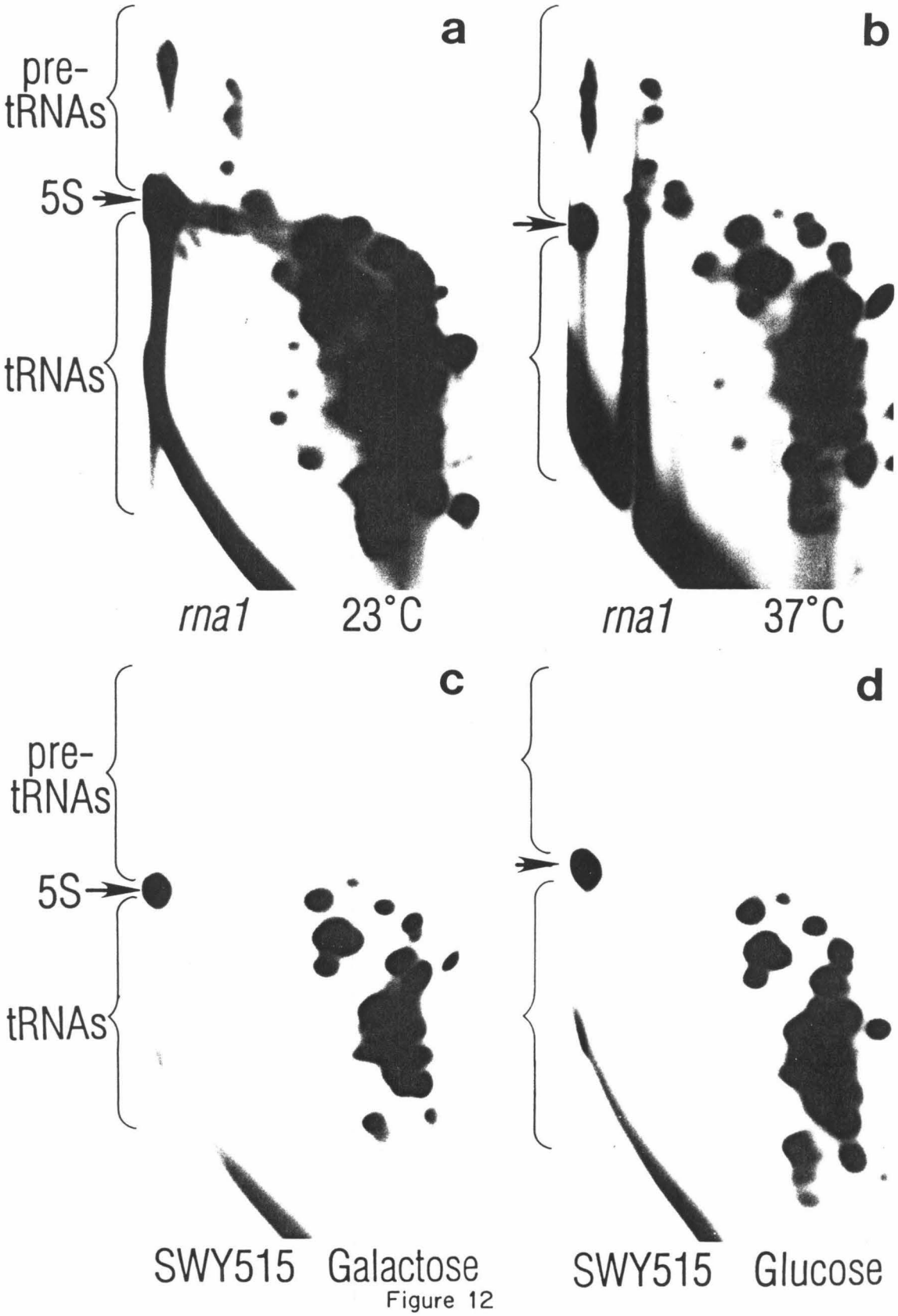

a
b
pre-
tRNAs
5S
tRNAs
rna1 23°C
rna1 37°C
c
d
pre-
tRNAs
5S
tRNAs
SWY515 Galactose
SWY515 Glucose

Figure 12

INTRODUCTION TO CHAPTERS III & IV

In the following research, attention has been focused specifically on the structure and function of tRNA ligase from *Saccharomyces cerevisiae*. Yeast tRNA ligase is responsible for the second step in tRNA splicing: ligation of the two tRNA halves resulting from endonucleolytic cleavage of those precursor tRNAs containing an intron (Peebles *et al.*, 1979; Greer *et al.*, 1983; Peebles *et al.*, 1983). In the first step of tRNA splicing, yeast endonuclease removes the intron in an ATP-independent reaction by cleaving at the 5' and 3' splice sites, leaving a 2',3'-cyclic phosphate at the 3' end of the 5' half tRNA and a 5'-hydroxyl on the 3' half tRNA (Knapp *et al.*, 1979; Peebles *et al.*, 1983). Yeast tRNA ligase possesses multiple activities which are necessary to correctly splice the two termini, resulting in a mature tRNA. An ATP-independent cyclic phosphodiesterase activity opens the 2',3'-cyclic phosphate of the 5' half tRNA to a 2'-phosphate. An ATP-dependent kinase activity phosphorylates the 5'-hydroxyl of the 3' half tRNA. An ATP-dependent adenylylate synthetase activity adenylylates the newly created 5'-phosphate of the 3' half tRNA through an adenylylated enzyme intermediate. Finally, a ligase activity joins the two activated tRNA halves, resulting in a 2'-phosphomonoester, 3',5'-phosphodiester bond at the splice junction.

Yeast tRNA ligase has been cloned and overexpressed in *E. coli* (Phizicky *et al.*, 1986), and the gene has been sequenced (Westaway *et al.*, 1988). Protease resistant fragments of tRNA ligase purified from *E. coli* still retain a subset of the multiple activities listed above (Xu *et al.*, 1990). In order to futher delineate the domains of ligase responsible for each of these activities, a gene deletion analysis was performed in collaboration with Barbara Apostol, Heather Belford, and Chris Greer of the University of California, Irvine. The results of this deletion study are summarized below and are presented in Apostol *et al.* (1991).

Since some of the deletions might result in completely inactive protein, to facilitate protein purification, these deletion mutations were joined at the amino-terminus to *E. coli* DHFR (dihydrofolate reductase). The plasmid used to construct all the deletions has an inducible bacterial promoter (ptac), fused to DHFR, then fused in frame to full-length yeast tRNA ligase using a linker which was designed to contain a Factor Xa cleavage site. This site allowed removal of the DHFR portion of the protein after purification, in case DHFR was shown to interfere with any of the ligase activities. The yeast ligase sequence is followed by translation termination codons in all three reading frames, and the *rrnB* bacterial transcription terminator. After transformation into *E. coli*, the fusion proteins were purified using an affinity column containing methotrexate, which is an analog of DHFR's natural substrate. The DHFR portion binds the column and the fusion protein can then be washed and eluted with folate.

Figure 1 is a schematic of the deletion mutants constructed (after Apostol *et al.*, 1991). D stands for *E. coli* dihydrofolate reductase, designated by the single black line. The light stippled section is the A region and contains the site of adenylylation. K stands for the proposed kinase region, in white, assigned on the basis of sequence similarities and enzymatic assays of protein fragments. C, the dark stippled section, stands for the cyclic phosphodiesterase region. Each possible combination of domains was constructed, i.e., full-length fusion protein DAKC, containing all three domains; single domain constructs, DA, DK, and DC; double domain constructs DAK, DKC, and DAC; and two other deletions made using available restriction sites within the ligase gene, DRΔNdeI and DRΔXmnI. The amino acids of yeast tRNA ligase that are included in each deletion construct are shown in Figure 1. The construction of these deletions is described in detail in Apostol *et al.* (1991). Important points to note are that a new *Hpa*I site has been created at amino acid 396 of the tRNA ligase coding region using oligonucleotide mutagenesis. This mutation changes amino acid #397 from an aspartate to a serine, and amino acid #401

from a lysine to a serine. These amino acid substitutions had no effect on the joining activity of the full-length fusion protein, DAKC.

These deletion constructs were purified and tested for all of the separate activities of tRNA ligase. For complete splicing activity, each fusion protein was tested for joining of radiolabeled $tRNA_3^{Leu}$SUP53 halves in the presence of ATP, and separated on denaturing polyacrylamide gels. For kinase activity, the ability of the fusion protein to radiolabel a cold oligonucleotide using [γ-^{32}P]ATP was tested, and the products were separated on denaturing polyacrylamide gels. For cyclic phosphodiesterase activity, a radiolabeled RNA of repeating 6-mers was transcribed *in vitro* and cleaved by Nuclease N1 to yield radiolabeled 2',3'-cyclic phosphate ends. These oligomers were used to test the ability of the fusion protein to open a 2',3'-cyclic phosphate to a 2'-phosphate, and the products were digested with RNase A, separated by thin layer chromatography, and autoradiographed. The assay for protein-AMP adduct formation, or adenylylate synthetase, was autoradiography of the fusion protein on SDS-polyacrylamide gels after incubation with [α-^{32}P]ATP. Finally, substrate binding was measured by incubation of the fusion protein with radiolabeled pre-$tRNA_3^{Leu}$ followed by gel shift in non-denaturing protein-RNA gels. Table 1 depicts the results of these activity assays for each of the fusion proteins.

Several interesting results emerge from the activity assays catalogued in Table 1. The cyclic phosphodiesterase assays indicate that only those deletion constructs containing the carboxyl terminus, or C region, of ligase contain cyclic phosphodiesterase activity. This confirms the result of Xu and coworkers that a C-terminal fragment of intact tRNA ligase still retains the cyclic phosphodiesterase activity (Xu *et al.*, 1990).

[1]Table 1. Activities of DHFR/Ligase Deletion Constructs.

Name of Deletion	Ligation	Kinase	Phospho-diesterase	Adenylylate Synthetase	tRNA Binding
DAKC	YES	YES	YES	YES	YES
DA	POOR[2]	NO	NO	YES[3]	NO
DK	NO	NO	NO	NO	NO
DC	NO	NO	YES	NO	NO
DAK	POOR[2]	NO	NO	YES	YES
DKC	NO	YES	YES	NO	YES
DAC	YES	NO	YES	YES[3]	YES
DRΔNdeI	NO	NO	YES	POOR	NO
DRΔXmnI	NO	NO	YES	NO	NO

[1]Results are from Apostol *et al.* (1991).
[2]Fusion proteins DA and DAK ligated tRNA halves only poorly, requiring a 100-1000 fold excess of protein to detect spliced product.
[3]Fusion proteins DA and DAC were adenylylated approximately twice as efficiently as DAKC.

The adenylylate synthetase activity is also as expected. Only those deletion constructs containing an approximately 40 kilodalton amino-terminal fragment of tRNA ligase, or the A region, can be adenylylated with ATP. DA and DAC adenylylate more efficiently perhaps because they are more stable than wild-type, or perhaps the K domain has an inhibitory effect on adenylylation. The deletion construct DRΔNdeI adenylylates only poorly, and DRΔXmnI does not have any adenylylate synthetase activity, indicating that almost all of the amino-terminal 396 amino acids are required for this activity.

Pre-tRNA binding is negative for the single domain constructs, but positive for any of the two domain constructs, and therefore requires any two of the three domains, indicating that the single pre-tRNA binding site of ligase (Apostol and Greer, 1991) may be composed of more than one element within the linear ligase sequence. Neither the NdeI or

XmnI deletion constructs were capable of pre-tRNA binding, and this could reflect a general instability of these proteins.

A surprising result was that kinase activity was seen only for DKC, and not for DK or DAK, which also contain the proposed kinase domain. Perhaps the kinase domain extends beyond the boundary at amino acid #595, and into the C domain, as neither DK nor DAK contain any C-terminal region.

The most surprising result obtained is that joining activity is seen for DAC, which does not contain the proposed kinase domain. The kinase assay at this time indicated that DAC did not possess a kinase activity. However, the mechanism of joining tRNA halves for the wild-type yeast enzyme requires the phosphorylation of the 3' half tRNA (Greer *et al.*, 1983). The joining mechanism of DAC might be different than that of tRNA ligase, or the structure of the final product may not be authentic mature tRNA.

For this reason, we undertook a detailed analysis of the DAC joining reaction to determine the exact structure of the spliced product, as well as any other characteristics of the splicing reaction which might differ from the full-length DHFR/ligase fusion protein, DAKC. These results are covered in Chapter 3, and Chapter 4 presents experiments which deal with the extrapolation of the results with the DAC mutant protein to the full-length DAKC.

REFERENCES

Apostol, B. L. and Greer, C. L. (1991) *Nucleic Acids Res.* **19**, 1853-1860

Apostol, B. L., Westaway, S. K., Abelson, J., and Greer, C. L. (1991) *J. Biol. Chem.* **266**, 7445-7455

Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979) *Cell* **18**, 37-45

Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536

Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J. (1979) *Cell* **18**, 27-35

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986) *J. Biol. Chem.* **261**, 2978-2986

Westaway, S. K., Phizicky, E. M., and Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171-3176

Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990) *Biochemistry* **29**, 6132-6138

FIGURE LEGEND

FIG. 1. **DHFR/Ligase Deletion Constructs.** A schematic is shown (as presented in Apostol *et al.*, 1991) depicting the DHFR fusion to a set of deletion constructs of the yeast tRNA ligase gene. *D* stands for the amino-terminal dihydrofolate reductase gene from *E. coli*, indicated by the single black line. *A* stands for the adenylylate synthetase region, indicated by the light stippled box. *K* stands for the proposed kinase domain, indicated by the white box. *C* stands for the cyclic phosphodiesterase domain, indicated by the dark stippled box. *Amino Acids* of RNA ligase which each deletion construct still retains are indicated. Numbering above the top line indicates the amino acid number at each junction. Letters corresponding to hash marks below the top line indicate relevant restriction enzyme sites used in the constructions. Abbreviations are as follows: *H*, *Hpa*I; *X*, *Xba*I; *N*, *Nde*I; and *K*, *Kpn*I.

DHFR/LIGASE DELETION CONSTRUCTS

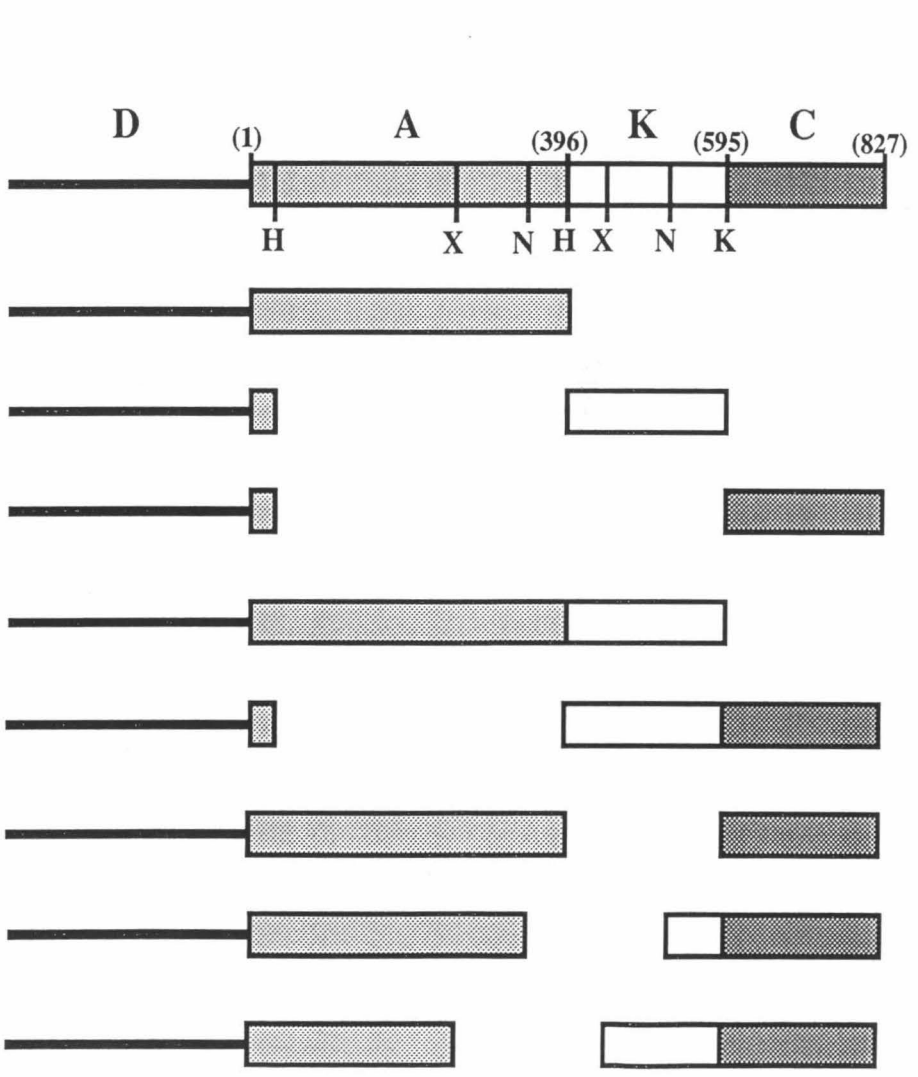


Figure 1

CHAPTER III

Novel Activity of a Yeast Ligase Deletion Polypeptide:
EVIDENCE FOR GTP-DEPENDENT tRNA SPLICING

Shawn K. Westaway[2], Heather G. Belford[1], Barbara Apostol[1], John Abelson[2], and Chris Greer[1]

[1]Department of Biological Chemistry, College of Medicine, University of California, Irvine; and

[2]Division of Biology, California Institute of Technology, Pasadena, California

[J. Biol. Chem. (1993), in press]

SUMMARY

Yeast tRNA ligase possesses multiple activities which are required for the joining of tRNA halves during the tRNA splicing process: cyclic phosphodiesterase, kinase, adenylylate synthetase, and ligase. A deletion polypeptide of a DHFR-ligase fusion protein, designated DAC, was previously shown to join tRNA halves although ATP-dependent kinase activity was not measurable in the assay used. We describe here a characterization of the mechanism of joining used by DAC and the structure of the tRNA product. DAC produces a joined tRNA and a splice junction with a structure identical to that produced by DAKC, the full-length DHFR-ligase fusion. Furthermore, DAC can use GTP as the sole cofactor in the joining reaction, in contrast to DAKC, which can only complete splicing in the presence of ATP. Both enzymes exhibit GTP-dependent kinase activity at 100-fold greater efficiency than with ATP. These results suggest that a potential function for the center domain of tRNA ligase (missing in DAC) is to provide structural integrity and aid in substrate interactions and specificity. They also support the hypothesis that ligase may prefer to use two different cofactors during tRNA splicing.

INTRODUCTION

In the yeast, *Saccharomyces cerevisiae*, there are about 400 tRNA genes, of which approximately 20% contain an intron which interrupts the functional portion of the tRNA (Ogden *et al.*, 1984, reviewed in Abelson, 1991). These introns are small (13-60 bases), and are all located at the same position in each gene, one base downstream from the anticodon. For those precursor tRNAs (pre-tRNAs[1]) containing an intron, processing of transcripts includes a three-step reaction called tRNA splicing. The tRNA splicing pathway in yeast has been extensively characterized (Peebles *et al.*, 1979; Peebles *et al.*, 1983; Greer *et al.*, 1983). The first step requires tRNA endonuclease, which removes the intron in an

ATP-independent reaction by cleaving at the 5' and 3' splice sites of the intron, resulting in a 2',3'-cyclic phosphate at the 3'-end of the 5'-half tRNA, and a 5'-hydroxyl on the 3'-half tRNA. The second step requires tRNA ligase, which joins the two halves to form a tRNA with a 2'-phosphate and a 3'-5' phosphodiester bond at the splice junction. The final step of splicing is the removal of the 2'-phosphate by a novel NAD-dependent 2'-phosphotransferase (McCraith and Phizicky, 1991).

In this research, we have focused specifically on the second enzyme in the splicing pathway, the tRNA ligase. The gene encoding tRNA ligase is essential in yeast (Abelson *et al.*, 1986). It has been cloned (Phizicky *et al.*, 1986) and sequenced (Westaway *et al.*, 1988), and encodes an 827 amino acid basic protein of molecular mass 95 kDa. The protein can be overexpressed and purified from *E. coli*. The ability to obtain large quantities of purified protein which is fully active in splicing provides a unique opportunity to study structure/function relationships in this complex enzyme.

Yeast tRNA ligase possesses multiple enzymatic activities which accomplish the joining step of splicing: 1) cyclic phosphodiesterase activity to open the 2',3'-cyclic phosphate of the 5'-half tRNA to a 2'-phosphate; 2) kinase activity to phosphorylate the 5'-hydroxyl of the 3'-half; 3) adenylylate synthetase activity to activate the 3'-half by transferring an AMP residue to the 5'-phosphate through an adenylylated enzyme intermediate; and 4) ligase activity to join the activated tRNA halves.

Evidence is compelling that the several activities of tRNA ligase are organized into distinct enzymatic domains. Partial proteolysis of the intact polypeptide and assay of the fragments for ligase-associated activities has confirmed that the region responsible for cyclic phosphodiesterase activity resides in the carboxy-terminus of the polypeptide, while the site of adenylylation lies in the amino-terminal half, specifically, at amino acid residue lysine-114 (Xu *et al.*, 1990). It is appropriate to note that bacteriophage T4 RNA ligase and T4 polynucleotide kinase together can substitute for yeast tRNA ligase in *in vitro* splicing reactions (Greer *et al.*, 1983), because T4 RNA ligase also adenylylates the 5'-

phosphate of its RNA substrate through an adenylylated enzyme intermediate (reviewed in Uhlenbeck and Gumport, 1982). There are distinct regions of limited sequence similarity between yeast tRNA ligase and both of the T4 enzymes, including the sites of adenylylation of the yeast and T4 ligases (Xu *et al.*, 1990; Apostol *et al.*, 1991), adding further corroboration for a domain-like organization of tRNA ligase.

Previous work has focused on the delineation and characterization of the domains of tRNA ligase (Apostol *et al.*, 1991). We reported the fusion of the tRNA ligase gene to the *E. coli* dihydrofolate reductase gene (DHFR[1]) to facilitate purification of the protein. The full-length DHFR-ligase fusion and large deletions of that fusion were characterized with respect to cyclic phosphodiesterase, kinase, adenylylate synthetase, and ligase activities, and binding of pre-tRNA substrate. The nomenclature of these fusion proteins reflects the domains and proposed activities which are still present in each deletion. Thus, the full-length DAKC fusion includes the DHFR domain "D" (present in all deletions and the full-length fusion), the "A" domain (adenylylate synthetase), the "K" domain (for proposed kinase), and the "C" domain (cyclic phosphodiesterase). Using these deletions, we showed that any two domains of tRNA ligase (i.e., A + K, K + C, or A + C) were sufficient to bind pre-tRNA. We have shown that the region responsible for adenylylation lies within amino acids 1-396 (the A region) and the region responsible for cyclic phosphodiesterase lies in the carboxy-terminus within amino acids 595-827 (the C region). Further, we showed that regions K and C were both necessary for measurable kinase activity. One interesting deletion construct, DAC, missing amino acids 397-594 in the proposed K region of the polypeptide, was still able to complete the splicing reaction at an efficiency of 60-100% compared to DAKC, the full-length ligase fusion. This result appeared to contradict the finding that both the K and C regions are required for kinase activity, essential in phosphorylating the 3'-half tRNA during splicing.

In this paper, we report the further characterization of DAC and its tRNA splicing product. As previously reported, DAC can be adenylylated, and is able to join tRNA

halves in the presence of ATP, bind precursor tRNA, and open a 2',3'-cyclic phosphate to a 2'-phosphate. However, ATP-dependent kinase activity was not detected by assaying incorporation of radiolabeled phosphate to an oligonucleotide substrate (Apostol *et al.*, 1991). Phosphorylation is an intrinsic step in the activation of the 3'-half tRNA in the normal splicing reaction. The result is that the phosphate in the newly-formed splice junction is derived from the γ-position of a nucleotide triphosphate cofactor (Greer *et al.*, 1983). Two possibilities exist to explain the ability of DAC to join tRNA halves. Either DAC uses a novel mechanism for ligation without kinase activity, or DAC uses the same mechanism for ligation as DAKC, but with different cofactor or substrate requirements. We favored the second hypothesis, because tRNA joined by DAC appeared to be equivalent to that spliced by DAKC. Radiolabeled tRNA spliced by either enzyme migrates in the same position on polyacrylamide gels (Apostol *et al.*, 1991). When those spliced tRNAs are analyzed by RNase T1 digestion followed by gel electrophoresis, a T1-resistant oligonucleotide of unique size is produced for both enzymes indicating the presence of a 2'-phosphate at the splice junction (data not shown). Since the RNase T1 patterns are the same for both enzymes, it is conceivable that DAC produces a mature tRNA with a structure analogous to that produced by the wild type enzyme, and therefore most likely uses the same mechanism in splicing.

In order to test this hypothesis, we further examined the nature of the splice junction created by DAC and the cofactor requirements of DAC and DAKC in joining and phosphorylation reactions. We describe below experiments which demonstrate that DAC joins tRNA halves to produce a splice junction identical to that produced by DAKC, and, unexpectedly, can use GTP as the only cofactor in the splicing reaction. This is in contrast to DAKC, which has no measurable ligase activity with GTP alone. We further demonstrate that both DAC and DAKC have GTP kinase activity with both non-tRNA and authentic tRNA substrates. These results have raised the possibility that the mechanism of

tRNA splicing by native yeast tRNA ligase *in vivo* may include the use of two different nucleotide cofactors.

EXPERIMENTAL PROCEDURES

Materials----Yeast tRNA splicing endonuclease, fraction 71 of a Sepharose CL-6B gel filtration step, was kindly provided by R. Rauhut (Rauhut *et al.*, 1990). DEAE-Sephadex (A-25), polyoxyethylene 20 cetyl ether (Brij 58), and DNase I were obtained from Sigma. Phage T7 RNA polymerase and Ultrapure ATP, CTP, GTP, and UTP were obtained from Pharmacia LKB Biotechnology, Inc. Phage T4 polynucleotide kinase (T4 PNK[1]) was obtained from United States Biochemical Corp. Ribonuclease T1, P1, and T2 were from Calbiochem. Snake venom phosphodiesterase (SVP[1]), restriction enzyme *Eco*RI, and calf intestine alkaline phosphatase were obtained from Boehringer Mannheim Biochemicals. Cellulose polyethyleneimine impregnated thin-layer chromatography (TLC[1]) plates were obtained from Brinkmann Instruments, Inc. Restriction enzyme *Bst*NI was obtained from New England Biolabs. [α-^{32}P]-labeled nucleotide triphosphates (NTPs[1]) (3000 Ci/mmol) and [γ-^{32}P]GTP (5000 Ci/mmol) were obtained from Amersham Corp. Methotrexate-agarose and immunopure sulfosuccinimidyl-2-(biotinamido)ethyl-1,3'-dithioproprionate (NHS-SS-Biotin[1]) were obtained from Pierce. Biotin-Sephacryl S1000 and succinylavidin were from EY Laboratories. N^6-(6-Aminohexyl)ATP was obtained from Bethesda Research Laboratories.

Strains and Plasmids----E. coli strain RJ1438 [*thi*1 *leuB*6 *thr*1 *lacY*1 *tonA*21 *supE*44 *hsdR galK* Δ*lon*100 (F' *lacI*^{SQ} *fzz*::Tn5)], used for expression of DHFR-ligase fusion proteins or derivatives, was kindly provided by R. Johnson. The construction of plasmids pDAKC and pDAC has been previously described (Apostol *et al.*, 1991). Plasmid ptac10, which expresses DHFR alone, was kindly provided by J. Normanly (Normanly *et al.*, 1986).

Plasmid pUC13Phe was kindly provided by V. Reyes (Reyes and Abelson, 1987). Plasmid pBS+ was obtained from Stratagene.

Cell Growth and Protein Purification----E. coli strain RJ1438, transformed with pDAKC, pDAC, or ptac10, was grown at 37°C in Luria-Bertani media containing 50 μg/ml each kanamycin and ampicillin. Cells were harvested at an A_{600} of 1.2-1.6, resulting in 1.75-2.4 g of wet cell weight per liter. Cells were stored at -70°C. The purification method for DHFR fusion proteins was basically as described by Baccanari *et al.* (1975), modified by Apostol *et al.* (1991). Crude extracts were obtained by a single lysozyme-nonionic detergent lysis. Cells were resuspended in 40 mM Tris-HCl (pH 8.0) at 2.14 ml/g of cells. Cells were brought to 1 mM phenylmethanesulfonyl flouride, 5.73 mM EDTA, 0.65 mg/ml lysozyme, and cell lysate was stirred for 30 min. Brij 58 was added to 1.85 mg/ml and lysate was stirred 30 min. Lysate was brought to 13.5 mM $MgCl_2$, 48 μg/ml DNase I, and stirred until viscosity was lost (approximately 30 min). Lysate was centrifuged 30 min at 24,000 X *g*, and the supernatant was adjusted to 40% saturation with ammonium sulfate and centrifuged. The supernatant was further adjusted to 75% saturation with ammonium sulfate and centrifuged. The second ammonium sulfate pellet was resuspended at 25-50 mg/ml protein in 50 mM potassium phosphate (pH 6.0), 200 mM KCl, 0.5 mM EDTA, 1 mM β-mercaptoethanol and dialyzed with 7-15 mls of settled methotrexate-agarose resin. Resin and bound protein were washed with 200 mM potassium phosphate (pH 6.0), 1 M KCl, 1 mM β-mercaptoethanol until the A_{280} of the eluate was less than 0.5. Protein was eluted in 100 mM potassium borate (pH 9.0), 1 M KCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 2 mM folic acid, then dialyzed in 100 mM potassium phosphate (pH 8.0), 250 mM KCl, 1 mM EDTA, 1 mM β-mercaptoethanol to remove folate, and concentrated by precipitation with ammonium sulfate at 75% saturation. Protein was dialyzed in 50 mM potassium phosphate (pH 8.0), 10% glycerol, 1 mM β-mercaptoethanol, 180 mM KCl, and passed over a 15-30 ml DEAE-Sephadex column to

remove nucleic acid. Protein was equilibrated in Buffer F (50 mM potassium phosphate (pH 8.0), 100 mM KCl, 35% glycerol, 5 mM β-mercaptoethanol), and stored at -20°C. This procedure typically yielded approximately 0.2 mg protein per liter of cell culture. Splicing assays after the methotrexate elution step indicated that proteins were as active as those that were also passed through the DEAE-Sephadex column, therefore removal of nucleic acids was unnecessary for our purposes and was omitted in some of the protein preparations. Protein determinations were by the method of Bradford (Bradford, 1976). Sodium dodecyl sulfate (SDS[1])-polyacrylamide gel electrophoresis was by the method of Laemmli (Laemmli, 1970).

Preparation of Substrates----The substrate used for splicing assays was a radiolabeled precursor $tRNA^{Phe}$ (pre-$tRNA^{Phe}$) prepared *in vitro* by runoff transcription with T7 RNA polymerase of plasmid pUC13Phe linearized with restriction enzyme *Bst*NI (Reyes and Abelson, 1987). Transcription reactions (100 µl) consisted of: 4 µg *Bst*NI-cut pUC13Phe; 500 µM each ATP, CTP, and GTP; 10 µM UTP; 100 µCi [α-^{32}P]UTP; 10 mM dithiothreitol (DTT[1]); 50 mM Tris-HCl (pH 8.0); 10 mM $MgCl_2$; 10 mM NaCl; and 250 units T7 RNA polymerase. Transcriptions were incubated at 37°C for 1 h, stopped by heating at 50°C for 20 min with addition of 10 µl 2 mg/ml proteinase K, 2% SDS, and 100 mM EDTA, and products were precipitated with addition of 100% ethanol. Pellets were resuspended in 3.5 M urea, 10% sucrose, and 0.05% each bromphenol blue and xylene cyanol, heated at 65°C for 5 min, and resolved on 12% acrylamide, 0.4% bis-acrylamide, 8M urea, 90 mM Tris-borate (pH 8.3), 2.5 mM EDTA gels. The predominant transcription product was excised from the gel, eluted at 37°C in 300 mM NaCl, 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1% phenol, filtered, precipitated with ethanol, and resuspended at 33,000 dpm/µl, equivalent to an approximate specific activity of 5,000 dpm/fmol.

The substrate for the analysis of splice junction structure was pre-$tRNA^{Phe}$ transcribed with each of four [α-^{32}P]-labeled nucleotide triphosphates. Transcription

reactions were performed as above except for the substitution of one of the following nucleotide mixes: 1) For ATP, CTP, or UTP-labeled transcripts, the following proportions appropriate to the labeled species; 100 μCi of [α-^{32}P]ATP, 20 μM ATP, 200 μM each CTP, GTP, and UTP; or 2) For GTP-labeled transcripts; 100 μCi of [α-^{32}P]GTP, 500 μM GTP, 200 μM ATP, CTP, and UTP (concentrations of GTP greater than 20 μM are required to effect complete transcription, since the sequence of tRNA^{Phe} begins with 5'-GGCG.) Transcriptions yielded pre-tRNA^{Phe} with a specific activity of 2-3,000 dpm/fmol, except for [α-^{32}P]GTP-labeled pre-tRNA^{Phe}, which could be produced only as a low specific activity transcript (115 dpm/fmol), due to the constraints imposed by GTP-dependent transcription initiation.

The substrate for kinase assays was a 9-nucleotide radiolabeled RNA, 5'-GGGCGAAUU-3', prepared *in vitro* by runoff transcription with T7 RNA polymerase of plasmid pBS+ linearized with *Eco*RI. Transcription reactions (25 μl) with either 25 μCi [α-^{32}P]CTP or 25 μCi [α-^{32}P]UTP were incubated and precipitated as above. Precipitate was resuspended and incubated at 50°C for 1 h with 2 units alkaline phosphatase to remove 5'-phosphates. Phosphatased RNA was treated with 20 mM [ethylenebis(oxyethylene-nitrilo)]tetraacetic acid (EGTA) at 65°C for 10 min to deactivate phosphatase. RNAs were resolved on 18% acrylamide, 0.9% bis-acrylamide, 8M urea gels as described above. The specific activity of this oligonucleotide was 110-220 dpm/fmol.

Splicing Reactions----Small scale splicing reactions were done essentially as described by Greer *et al.* (1983). Briefly, 6.8 fmol pre-tRNA^{Phe} were preincubated with 0.1 μg/ml yeast tRNA splicing endonuclease in 20 mM Hepes[1] (pH 7.8), 2.5 mM spermidine-HCl (pH 7.0), 5 mM $MgCl_2$, 0.1 mM DTT, 0.4% Triton X-100, 25 mM NaCl at 30°C for 18 min. Following the addition of 50 μg/ml DHFR-ligase fusion protein (DAKC or DAC) and varying concentrations of NTP, reactions were incubated an additional 15 min. Spliced products were resolved on 12% polyacrylamide gels as described above. Gels

were autoradiographed and products were quantitated by measuring Cerenkov radiation in gel slices.

Kinase Assay----Kinase reactions (5 µl) contained 10-40,000 cpm substrate (150-1200 fmol 9-nucleotide oligomer described above), 30 µg/ml DAKC or DAC or 30 units T4 PNK, 20 mM Hepes (pH 7.8), 2.5 mM spermidine-HCl (pH 7.0), 5 mM $MgCl_2$, 0.1 mM DTT, 0.4% Triton X-100, 25 mM NaCl, 0.05 mM EDTA, 2% glycerol, and varying concentrations of NTP. Reactions were incubated for 30 min at 30°C, except for reactions containing T4 PNK, which were incubated at 37°C. Reactions were resolved on 18% polyacrylamide gels as described above. Gels were autoradiographed and reaction products were quantitated by measuring Cerenkov radiation in gel slices.

Origin of the 2'-Phosphate at the Splice Junction----Preparative splicing assays (100 µl) of pre-$tRNA^{Phe}$ labeled with each of four ribonucleotides were carried out essentially as described for small scale splicing assays. Briefly, 10^6 cpm substrate (equal to 1 pmol, or 30 pmol [α-^{32}P]GTP-labeled tRNA) were incubated with 0.02 µg endonuclease for 30 min at 30°C. ATP was added to 4 mM, followed by 1.5 µg DAKC or DAC, and reactions were incubated an additional 30 min. Reactions were stopped as above and precipitated with ethanol. Spliced products were separated by gel electrophoresis, eluted, and precipitated as described above.

Ribonuclease Digestion and Thin-Layer Chromatography----For RNase T1 analyses, spliced tRNAs were incubated in 6 µl reaction volumes for 30 min at 37°C in 10 mM Tris-HCl (pH 7.4), 1 mM EDTA with 10 units RNase T1 and 20 µg RNA carrier. RNase T1-digested products were resolved on 23% acrylamide, 0.8% bis-acrylamide, 7.4 M urea gels and visualized by autoradiography.

For RNase T2 analyses, spliced tRNAs were incubated in 4 µl for 2 h at 37°C in 20 mM sodium acetate (pH 5.0) with 5 units RNase T2 and 37 µg RNA carrier. Reaction mixtures were spotted on polyethyleneimine-cellulose TLC plates and developed in 1 M LiCl. Plates were dried and autoradiographed. Unlabeled marker nucleotides were visualized with shortwave ultraviolet light.

For P1 nuclease analyses, spliced tRNAs or tRNA halves were incubated in 5 µl for 2 h at 37°C in 20 mM ammonium acetate (pH 5.0) with 5 units P1 nuclease and 37 µg RNA carrier, and chromatographed as above.

P1-nuclease resistant dinucleotides were further analyzed by SVP digestion. After washing plates with 70% ethanol to remove LiCl, radioactive spots were eluted with ice-cold 30% triethylamine carbonate (pH 9) (Volckaert *et al.*, 1976), dried, and resuspended in H_2O. SVP digests (10 µl) containing the eluted RNA, 2 µg SVP, 37 µg RNA carrier, 25 mM Tris-HCl (pH 9.5), and 5 mM $MgCl_2$ were incubated 2 hours at 37°C. Some reactions also contained 5-10 µg unlabeled nucleotide marker. Reaction products were analyzed by TLC as described above.

Incorporation of γ-phosphate of GTP into Splice Junction----Biotinylated [α-^{32}P]UTP-labeled pre-tRNA^{Phe} was prepared as follows: 1) For tRNA of high specific activity, T7 RNA polymerase transcriptions were carried out as described above except that ATP was reduced to 250 µM and N^6-(6-Aminohexyl)ATP was added to 250 µM. Biotinylated [α-^{32}P]UTP-labeled pre-tRNA^{Phe} of low-specific activity was prepared as above except [α-^{32}P]UTP was reduced to 10 µCi/100 µl reaction and UTP was added to 500 µM. Purified transcripts were biotinylated by reaction with NHS-SS-Biotin (Ruby *et al.*, 1990).

For the preparation and use of the biotin Sephacryl:succinylavidin resin, all steps were done at 4°C. Biotin Sephacryl resin was prebound to succinylavidin at a concentration of 1 mg succinylavidin per 200 µl packed resin for 1-12 hours (Ruby *et al.*,

1990). Resin was washed with 10 mM Tris-HCl (pH 7.5), 200 mM NaCl (Buffer A) to remove unbound avidin, and resuspended in an equal volume of Buffer A.

Splicing assays were performed as described above, except that reactions contained 600 fmol of low-specific activity biotinylated [α-^{32}P]UTP-labeled pre-tRNA^{Phe} and 100 μCi of desiccated [γ-^{32}P]GTP as the only cofactor. Control reactions also contained 6 fmol of high-specific activity biotinylated [α-^{32}P]UTP-labeled pre-tRNA^{Phe} as well as 2 mM unlabeled ATP instead of [γ-^{32}P]GTP. Splicing reactions were stopped as above, precipitated with ethanol, resuspended in 5 mM Tris-HCl (pH 7.5), 100 mM NaCl, and bound to 60 μl avidin resin slurry. After incubation with the splicing reaction for 15-60 min, 200 μl Buffer A was added, resin was vortexed gently, microfuged for 10 seconds, and supernatant was removed. This wash step was repeated three times with 5-15 min incubations. Material bound to the resin was eluted three times in 150 μl of 150 mM DTT, 40 mM Tris-acetate (pH 8.5), 150 mM NaCl, and 3 mM $MgCl_2$ (Ruby *et al.*, 1990), and precipitated with ethanol. Reaction products were resolved by gel electrophoresis as described above.

RESULTS

DHFR-ligase fusion proteins were expressed in *E. coli* and purified as described in *Experimental Procedures*. Protein gels confirmed the presence and purity of each protein (data not shown). DAKC migrated at its expected molecular mass of 110 kDa and DAC at its expected molecular mass of 88 kDa, and estimates of specific activity for ATP-dependent joining by DAKC and DAC were comparable to that reported elsewhere (Apostol *et al.*, 1991).

We previously reported that the presence of DHFR in the full-length DHFR-ligase fusion protein did not interfere with or enhance any activity attributed to the ligase portion of the protein (Apostol *et al.*, 1991). Bacterial DHFR alone, purified by the same

procedure as that used for the fusion proteins, had no activity in splicing assays under any of the conditions tested (see Belford *et al.*, accompanying manuscript). Therefore for the purposes of this study, all experiments were done with DAC or DAKC, both with intact DHFR domains.

Determination of the Structure of Spliced Product----Analyses of RNase T1 digestion products revealed that the sequences of spliced tRNA produced by DAC and DAKC were identical (data not shown). However, it was necessary to show that the specific structure of the splice junction and the origins of the splice junction phosphates in tRNAs joined by the enzymes were also identical.

For native yeast tRNA ligase, the 2'-phosphate at the splice junction of mature tRNA originates from the pre-tRNA (Greer *et al.*, 1983). We previously determined that the 2'-phosphate at the splice junction of tRNA joined by DAC originated from the pre-tRNA substrate, not from an external cofactor (Apostol *et al.*, 1991). It was necessary to show further that this 2'-phosphate originated specifically from the 5'-splice site phosphate of the pre-tRNA. Pre-tRNA^{Phe} substrates labeled with each of the four ribonucleotides (in the T7 RNA polymerase transcription reaction) were used in preparative splicing reactions with endonuclease and DAKC or DAC. Fig. 1 shows the results of the analysis of these spliced tRNAs with respect to the origin of the 2'-phosphate. Fig. 1*A* is a diagram depicting the expected products of the splicing reaction (showing only the nucleotides surrounding the splice junction) and subsequent nuclease digestions of spliced tRNA. Spliced tRNA^{Phe} produced by both DAKC and DAC in the presence of ATP migrates identically on polyacrylamide gels (Fig. 1*B*). When these spliced tRNAs are digested with P1 nuclease, the predicted products are 5'-monophosphates and a P1 nuclease-resistant splice junction dinucleotide of sequence 5'-$_pG^{2'p}{}_pA$-3' (see Fig. 1*A*). Radiolabeled 5'-monophosphates will result from the ribonucleotide that was used to radiolabel the pre-tRNA (i.e., either 5'-AMP, 5'-GMP, 5'-UMP, or 5'-CMP). The last base of the 5'-half

tRNA is a guanine residue, therefore, in reactions using pre-tRNA^{Phe} transcribed with [α-^{32}P]GTP, the 5'-phosphate of the P1 nuclease-resistant dinucleotide should be radiolabeled. The first base of the intron is an adenine residue, therefore, in reactions using pre-tRNA^{Phe} transcribed with [α-^{32}P]ATP, the 2'-phosphate of the P1 nuclease-resistant dinucleotide should be radiolabeled. The products of P1 nuclease digestion of spliced tRNA were as expected for the conventional joining reaction and were identical for both DAC and DAKC (Fig. 1*C*). When the P1 nuclease-resistant dinucleotides were eluted and digested with SVP, the predicted products were 5'-AMP and 2',5'-GDP (see Fig. 1*A*). The 5'-AMP should be unlabeled because this phosphate should originate from the γ-position of the unlabeled ATP in the splicing reaction. Again, only when [α-^{32}P]GTP-labeled pre-tRNA^{Phe} or [α-^{32}P]ATP-labeled pre-tRNA^{Phe} is used in the splicing reaction will 2',5'-GDP be radiolabeled. Fig. 1*D* shows that the products of SVP digestion were indeed as predicted and were identical for both enzymes (*lanes 5-8*). The first four lanes in Fig. 1*D* are controls without SVP added. A radiolabeled species was detected which migrated with authentic 2',5'-GDP in the lanes for [α-^{32}P]GTP-labeled tRNA (*lanes 1 and 2*), but not for [α-^{32}P]ATP-labeled tRNA (*lanes 3 and 4*). We believe this species is [α-^{32}P]GTP, originating from the 5'-pppG which is the first nucleotide of T7-transcribed pre-tRNA^{Phe}. GTP migrates closely above the P1 nuclease-resistant dinucleotide position and nearly identically with 2',5'-GDP in 1 M LiCl (data not shown). The spot was present in amounts approximately equimolar to the P1 nuclease-resistant dinucleotide, consistent with a single occurrence per molecule. It is probable that both species were coeluted due to overexposure of the TLC in Fig. 1*C*. Regardless, the radiolabel in the dinucleotide that was present reappeared in 2',5'-GDP when digested with SVP, indicating that the P1 nuclease-resistant dinucleotide in this case was indeed 5'[^{32}P]$_p$G2'$^p{}_p$A-3'.

Four similarly labeled pre-tRNA^{Phe} transcripts were used in a splicing reaction with each enzyme, and the products were digested with RNase T2 instead of P1 nuclease. The identity and radiolabeling pattern of these products were as predicted, and again identical

for both DAC and DAKC (data not shown). These results are consistent with the supposition that the 2'-phosphate at the splice junction of tRNA joined by either DAKC or DAC originates from the phosphate present at the 5'-splice site of the intron of the pre-tRNA. Additionally, we have determined that a 5'-half tRNA pretreated with T4 PNK (which exhibits both cyclic phosphodiesterase and phosphatase activities capable of removing the 2',3'-cyclic phosphate of the 5'-half) and reannealed to a normal 3'-half tRNA could not be joined by either DAC or DAKC (data not shown). This is in agreement with previous results using yeast tRNA splicing extracts (Knapp *et al.*, 1979). Together, these results indicate that DAC possesses 2',3'-cyclic phosphodiesterase activity which opens the cyclic phosphate to a 2'-phosphate (Apostol *et al.*, 1991), and that this phosphate constitutes the 2'-phosphate at the splice junction and must be present on the 5'-half tRNA for splicing to occur.

Determination of Cofactor Requirements During Splicing----We began a systematic study of splicing in the presence of a variety of nucleotide cofactors to determine whether DAC has cofactor or substrate requirements differing from DAKC. Splicing reactions were carried out in a separate titration for each of the four NTPs. The results of splicing reactions containing either DAKC or DAC, and increasing amounts of NTP are shown in Fig. 2. Fig. 2*A* shows a representative autoradiograph of splicing reactions with a constant concentration of DAKC and increasing concentrations of ATP, from 5 μM to 20 mM, as described in *Experimental Procedures*. The amount of spliced tRNA product clearly increased, up to an ATP concentration of 5 mM. There was a corresponding decrease in the amount of tRNA halves, indicating that this tRNA product resulted from the joining of tRNA halves.

There was an abrupt decrease in production of spliced tRNA at ATP concentrations above 5 mM. This decrease was similar for both DAKC and DAC, and was observed with each of the four NTPs (data not shown). This decrease was apparently not due to NTP

concentration-dependent ribonuclease activity since the recovery of pre-tRNA and tRNA halves was unaffected. It is possible that this reduction in splicing activity resulted either from inhibition by the increasing ionic strength at high concentrations of NTP in these reactions, or from the NTP-mediated chelation of Mg^{2+} (ligase requires a divalent cation in the splicing reaction (Greer *et al.*, 1983)).

Fig. 2*B*, *C*, and *D* are graphs of representative data sets obtained from polyacrylamide gels of splicing reactions carried out with increasing concentrations of one of the four NTPs, and either DAKC or DAC. Enzyme concentrations were kept constant, equal, and limiting for both DAKC and DAC, so as to emphasize the effect of NTP concentration on joining. K_{app}^{NTP} values were determined as the NTP concentration resulting in 50% maximal activity, and are average values generated from 2-4 separate experiments.

In the presence of ATP, DAC exhibited splicing activity which was comparable to DAKC (Fig. 2*B*), with average K_{app}^{ATP} values of 0.7 mM and 0.9 mM, respectively. Neither enzyme joined in the presence of CTP alone. Splicing reactions performed in the presence of GTP alone (Fig. 2*C*) or UTP alone (Fig. 2*D*) allowed us to conclude that DAKC cannot join tRNA halves with CTP, GTP, or UTP alone, even at concentrations up to 10 mM (data not shown). However, DAC can join tRNA halves with GTP or UTP alone, with an average K_{app}^{GTP} of 0.4 μM and an average K_{app}^{UTP} of 100 μM (Fig. 2*C* and *D*); values which are 10- to 1000-fold lower than the K_{app}^{ATP} for splicing with DAC. At first glance, this would appear to be a gain of function mutation for the deletion protein, DAC.

Determination of Cofactor Requirements in the Kinase Reaction----The discovery that DAC can splice tRNA using only GTP or UTP led us to evaluate cofactor requirements for the kinase reaction. Based on the greater efficiency of GTP- and UTP-dependent splicing (see above) and the previous lack of measurable ATP-dependent kinase activity (Apostol *et al.*,

1991), we hypothesized that DAC might require a cofactor other than ATP in the kinase reaction. It was also possible that DAC might require a higher concentration of ATP to detect kinase activity only in assays with synthetic substrates, and that reactions with authentic substrates might have different requirements.

To test these possibilities, a revised kinase assay was developed. The previous kinase measurements were obtained using an unlabeled size-fractionated poly(U_nG) substrate and a labeled cofactor, [γ-^{32}P]ATP, allowing only limited concentrations of cofactor to be added (Apostol *et al.*, 1991). A second kinase assay was developed using a radiolabeled 9-nucleotide RNA transcript with a 5'-hydroxyl, increasing the flexibility of the assay and allowing addition of greater concentrations of cofactor (see *Experimental Procedures*). Upon phosphorylation, the nonamer has greater mobility in an 18% polyacrylamide gel, presumably because of the added negative charges associated with the newly added phosphate group (Fig. 3*A*). This conclusion is supported by the comigration of the nonamer in the reaction with T4 PNK and ATP, as well as the comigration of the transcription product without phosphatase pretreatment (possessing the 5'-triphosphate from transcription initiation) (Fig. 3*A*, *lanes 9 and 10*). Fig. 3*B*, *C*, and *D* are graphs of representative data sets obtained from polyacrylamide gels of kinase reactions carried out with either DAKC or DAC. Kinase activity (expressed as the percent of oligomer phosphorylated) is plotted as a function of increasing concentration of NTP. DAC did not exhibit significant kinase activity with CTP or ATP using this artificial substrate (Fig. 3*B*). However, DAKC demonstrated both ATP- and CTP-dependent kinase activity, with K_{app}^{NTP} values which differed by approximately 3-fold.

Both DAKC and DAC were active with GTP alone (Fig. 3*C*), as well as with UTP alone (Fig. 3*D*). DAKC had an average K_{app}^{GTP} of 0.25 μM (see inset, Fig. 3*C*) and an average K_{app}^{UTP} of 25 μM for kinase activity, demonstrating a marked preference for GTP over UTP. DAC was also 30-fold more efficient at phosphorylation with GTP than with UTP, with average K_{app} values of 0.1 mM and >3 mM, respectively. This preference for

GTP during phosphorylation paralleled the efficiency of GTP over UTP in the data from DAC splicing reactions shown in Fig. 2. However, the lack of measurable ATP kinase activity for DAC was in direct contrast with those splicing results. If, as we propose, DAC must use the γ-phosphate of ATP to kinase the 3'-half of the tRNA during ATP-dependent splicing, then we can infer that DAC not only exhibits different cofactor utilization during joining and phosphorylation, but seems to display different reaction characteristics based on the nature of the substrate.

Incorporation of the γ-phosphate of GTP as the 3'-5' Splice Junction Phosphate----Fig. 1*D* shows that the 5'-AMP resulting from SVP cleavage of the splice junction dinucleotide from tRNA joined by DAC was not labeled for substrates labeled during transcription with any of the four NTPs. Thus this phosphate must be derived from a cofactor, as has been shown for native tRNA ligase (Greer *et al.*, 1983). We have shown above that both DAKC and DAC can use GTP in the kinase reaction. We have also demonstrated that DAC can splice tRNA using GTP alone. Therefore, if DAC employs the same mechanism as DAKC, it must then use the γ-phosphate of GTP to kinase the 3'-half tRNA. In order to unequivocally identify the source of this 3',5'-splice junction phosphate, [γ-^{32}P]GTP was used to label a low specific activity substrate in the splicing reaction.

Both DAKC and DAC were able to kinase a 9-nucleotide RNA which is not a natural substrate for tRNA splicing. It is more difficult to demonstrate phosphorylation of the native substrate by radiolabel incorporation, because DAKC, DAC, and tRNA endonuclease protein fractions all contain an appreciable quantity of cellular RNA. The nonspecific kinase activity alluded to above causes high levels of background labeling of RNA, completely masking the appearance of authentic tRNA splicing reaction products. Therefore we used biotin-avidin affinity chromatography (Ruby *et al.*, 1990) to separate pre-tRNA substrate and spliced products from other incidentally radiolabeled RNA contaminants. Briefly, the procedure was as follows: Pre-tRNA transcribed with N^6-(6-

Aminohexyl)ATP was biotinylated through a disulfide bond by reaction with NHS-SS-Biotin (see *Experimental Procedures*). This pre-tRNA was used in the splicing reaction and then bound to biotin Sephacryl:succinylavidin resin. Nonbiotinylated material was washed away and biotinylated tRNA splicing products were released by treatment with DTT (Ruby *et al.*, 1990).

An autoradiograph of a gel from such a splicing assay using biotinylated [α-^{32}P]UTP-labeled pre-tRNA is shown in Fig. 4*A*. In this figure, lanes labeled *^{32}P-tRNA* contained both low-specific activity and high-specific activity biotinylated [α-^{32}P]UTP-labeled pre-tRNA. Pre-tRNA treated only with tRNA endonuclease produced tRNA halves (Fig. 4*A*, *lane 1*). The control lanes show the comigration of tRNA spliced by DAKC or DAC in the presence of nonradioactive ATP (Fig. 4*A*, *lanes 2 and 3*). Low-specific activity pre-tRNA spliced by DAKC was not visible at this exposure (Fig. 4*A*, lane labeled *C*). Splicing reactions carried out with low-specific activity pre-tRNA, DAKC or DAC, and [γ-^{32}P]GTP as the only cofactor showed incorporation of radiolabel at a level far above the low-specific activity tRNA alone (Fig. 4*A*, compare lanes labeled *γ-^{32}P-GTP* to lane labeled *C*). Radiolabeled material comigrating with spliced tRNA was produced by DAC (Fig. 4*A*, *lane 5*). By contrast, DAKC produced a more diffuse band which did not precisely comigrate with authentic spliced tRNA (Fig. 4*A*, *lane 6*). Other biotinylated RNAs containing 5'-hydroxyls were kinased as well, including the 3'-half and the intron, which both possess a 5'-hydroxyl through cleavage by endonuclease (Knapp *et al.*, 1979).

Fig. 4*B* depicts the normal tRNA splice junction structure and subsequent nuclease digestion products. The predicted fate of a labeled junction phosphate is indicated. If the 3'-half tRNA is phosphorylated with the γ-phosphate of [γ-^{32}P]GTP, the radiolabeled phosphate should be incorporated into the 3'-5' phosphodiester at the splice junction. We detected the presence of the predicted splice junction dinucleotide following P1 nuclease digestion of tRNA spliced by DAC, but not by DAKC (Fig. 4*C*, lanes labeled *tRNA*). Instead, for DAKC we detected radiolabeled 5'-CMP, 5'-AMP, and 5'-GMP, indicating

that labeled species in the spliced tRNA area of the gel were probably due to 5'-phosphorylation of pre-tRNA breakdown products. When the P1 nuclease-resistant dinucleotide from Fig. 4*C*, (lane labeled *tRNA/DAC*), was further digested with SVP, 5'-[^{32}P]AMP was released as predicted for the structure shown in Fig. 4*B*. This result proves that DAC uses the γ-phosphate of GTP to kinase the 3'-half tRNA during GTP-dependent joining.

When the bands from the tRNA halves region of the gel were eluted and digested with P1 nuclease, the primary product was radiolabeled 5'-AMP, as predicted for a simple phosphorylation product (Fig. 4*C*, lanes labeled *halves*). The less intense radiolabeled 5'-UMP might result either from P1 nuclease digestion of low-specific activity [α-^{32}P]UTP-labeled tRNA halves, or through cleavage adjacent to the 3'-splice site of pre-tRNA by tRNA endonuclease. The 3'-splice site sequence of pre-tRNA^{Phe} is 5'-Up{down arrow}ApU-3', with the arrow indicating the normal cleavage site. If cleavage occurs either one base upstream or one base downstream of the normal site, the 5'-hydroxyl of a U residue will be available for phosphorylation, releasing radiolabeled 5'-UMP upon P1 nuclease digestion. Yeast tRNA endonuclease does not normally exhibit significant splice site cleavage infidelity with nonbiotinylated pre-tRNA^{Phe} (personal observation, and R. Rauhut, personal communication). We hypothesize that a miscleavage at the 3'-splice site might occur because the first base of the 3'-half is an adenine residue, and in part of this particular pre-tRNA population, this adenine will have been substituted with N^6-(6-Aminohexyl)AMP and should be biotinylated, which may stereochemically interfere with tRNA endonuclease cleavage. This hypothesis might also explain the reproducibly diffuse nature of the band corresponding to the 3'-half tRNA (see Fig. 4*A*); the group of bands may represent 3'-halves differing in size by one or two residues due to miscleavage at the 3'-splice site. In summary, the results of joining with [γ-^{32}P]GTP confirm the previous observations that both DAKC and DAC can use the γ-phosphate of [γ-^{32}P]GTP to

phosphorylate the 3'-half of the tRNA, and that while DAC can complete the joining reaction, DAKC cannot produce authentic spliced tRNA with GTP as its only cofactor.

DISCUSSION

We have shown that spliced $tRNA^{Phe}$ products joined by DAC and DAKC are structurally identical, therefore DAC most likely uses the same mechanism for joining as does DAKC or native tRNA ligase. Most other enzyme reaction mechanisms that could explain joining by DAC have been ruled out by our data. However, it is possible to propose mechanisms that are not incompatible with our data. For example, DAC could mediate the 3'-phosphorolysis of the 2',3'-cyclic phosphate on the 5'-half tRNA using the 5'-phosphate on the 3'-half, resulting in a spliced tRNA with a 2'-phosphate, 3',5'-splice junction phosphate. In this reaction, the origins of all splice junction phosphates would be as predicted for the original mechanism, and the product would be structurally indistinguishable from that produced by the original mechanism. However, no adenylylate or guanylylate synthetase activity is required in this alternate mechanism to activate the 5'-phosphate of the 3'-half tRNA. This could explain why DAC was able to join tRNA halves in the presence of GTP alone. However, this mechanism requires an intact 2',3'-cyclic phosphate at the 3'-end of the 5'-half tRNA. We have previously shown that DAC exhibits 2',3'-cyclic phosphodiesterase activity comparable to DAKC using a non-tRNA substrate (Apostol *et al.*, 1991). Therefore we believe that it is far more likely that DAC uses the same mechanism as DAKC, but uses GTP instead of ATP to activate the 5'-phosphate of the 3'-half during GTP-dependent splicing. If so, then DAC, which is missing the central portion of intact ligase, has probably undergone a rearrangement of active or substrate/cofactor sites such that GTP can now be used exclusively for both cofactor-requiring activities involved in splicing.

It is instructive to compare the relative activities of DAC and DAKC with respect to the constants for K_{app}, or 50% maximal activity for joining or kinase. ATP-dependent joining is approximately equal for both enzymes, with a K_{app}^{ATP} for DAKC of 0.9 mM and for DAC of 0.7 mM. However, DAC has much lower K_{app} values for GTP- and UTP-dependent joining, of 0.4 μM and 100 μM, respectively. Greer *et al.* (1983) reported that tRNA ligase purified from yeast was able to effect splicing using 1 mM UTP, and, to a lesser extent, CTP or GTP. However, they raised the possibility of either interconversion of nucleotides or up to 5% contamination of their cofactors, at which level splicing would still be able to occur. A partially purified fraction of tRNA ligase purified from yeast does not splice in the presence of any NTP other than ATP when assayed with Ultrapure NTPs (Belford *et al.*, accompanying manuscript), and as we have shown above, DAKC cannot join using GTP, UTP, or CTP alone.

A comparison of the constants for the kinase reaction reveals that they are very different. The K_{app}^{ATP} for kinasing by DAKC is 0.2 mM, which is 5-fold better than the K_{app}^{ATP} for joining by DAKC. DAKC may have a higher substrate affinity for the non-tRNA substrate used in the kinase assay than for the pre-$tRNA^{Phe}$ used in the joining assay. DAC has a K_{app}^{ATP} for kinase of greater than 4 mM, which is more than 5-fold greater than the K_{app}^{ATP} for joining by this same enzyme, and this may reflect distinct features of the reactions measured with synthetic oligonucleotide and authentic pre-tRNA substrates. DAKC exhibits K_{app} values for GTP-, UTP-, and CTP-dependent kinase activity of 0.25 μM, 25 μM, and 61 μM, respectively. By contrast, DAC has K_{app} values of 100 μM and 3 mM for GTP- and UTP-dependent kinase activity, respectively, with no activity detected for CTP. These values are up to 400-fold greater than the same values using DAKC, and again, may reflect unique requirements for use of a synthetic kinase substrate by DAC. It is perhaps not surprising that DAKC can utilize any of the four NTPs for phosphorylation, even though it can only complete the splicing reaction with ATP. T4 PNK has also been shown to use UTP, GTP, or CTP equally as well as ATP in the kinase

reaction (reviewed by Richardson, 1981), and we have found sequence similarities in two regions of the enzymes (Xu *et al.*, 1990; Apostol *et al.*, 1991). The most surprising result is that GTP is 100-fold more efficient than ATP when used in the kinase reaction with either DAC or DAKC.

We previously showed that DAC had half the binding affinity for pre-tRNA substrates as did DAKC (Apostol *et al.*, 1991). Since we do have GTP-dependent kinase activity with DAC, the region missing from the center of DAC (the K domain) is probably not responsible for GTP-dependent kinase activity, but instead may serve to hold the intact protein together when presented with its authentic substrate. We also demonstrated that any two of the three domains were sufficient for pre-tRNA binding, implicating a composite of multiple areas of the tRNA ligase linear sequence in forming the high affinity substrate binding site. Removal of the central "K" region could have a greater effect on substrate binding than on kinase or joining activities, such that the measurements of relative efficiencies of the individual activities during splicing vary greatly from substrate to substrate. Indeed, many protein preparations of DAC were found to be inactive for splicing (3 of 11 independent protein preparations were active for splicing, data not shown) and this could reflect a general instability of the protein.

It is provocative that the K_{app} for GTP-dependent joining (0.4 μM) by DAC closely approximates the K_{app} for GTP-dependent kinase activity (0.25 μM) for full-length DAKC. This results suggests that the step involving phosphorylation of the 3'-half tRNA is probably not the rate limiting step in the normal tRNA splicing reaction, and in addition may intimate that the characteristics of GTP-binding are probably similar both in the deletion and the full-length polypeptide.

In summary, the results presented here permit us to propose an amended mechanism for yeast tRNA splicing, including GTP as one of the cofactors. Fig. 5 is a diagram of this revised mechanism for splicing by endonuclease, yeast tRNA ligase, and NAD-dependent 2'-phosphotransferase. This model merges our *in vitro* observation that

GTP is used more efficiently than ATP by ligase in the kinase step with the data which indicate that ligase still requires ATP to complete the joining of tRNA halves, thus reserving ATP for the adenylylate synthetase function. This requirement for ATP in the adenylylate synthetase step is similar to the results seen with T4 RNA ligase, which, of the available nucleoside triphosphates, can utilize only ATP or dATP in either the circularization of RNA or in the ATP-PPi exchange reaction which indirectly measures enzyme adenylylation (reviewed by Uhlenbeck and Gumport, 1982). This model is very testable and experiments which address this are described in the accompanying manuscript.

ACKNOWLEDGMENTS

We wish to thank Reinhard Rauhut for generously providing purified yeast tRNA splicing endonuclease and for critical review of the manuscript. We are grateful to Craig Peebles, Eric Phizicky, Rosalind Haselbeck, Mark Berlin, and Daniel Herschlag for helpful discussions and comments.

FOOTNOTES

[1]The abbreviations used are: Pre-tRNA, precursor tRNA; DHFR, dihydrofolate reductase; T4 PNK, T4 polynucleotide kinase; SVP, snake venom phosphodiesterase; TLC, thin-layer chromatography; NTP, nucleotide triphosphate; NHS-SS-Biotin, sulfosuccinimidyl-2-(biotinamido)ethyl-1,3'-dithioproprionate; SDS, sodium dodecyl sulfate; DTT, dithiothreitol; Hepes, 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid.

REFERENCES

Abelson, J. (1991) *Harvey Lect.* **85**, 1-42

Abelson, J.N., Brody, E.N., Cheng, S.-C., Clark, M.W., Green, P.R., Dalbadie-McFarland, G., Lin, R.-J., Newman, A.J., Phizicky, E.M., and Vijayraghavan, U. (1986) *Chemica Scripta* **26B**, 127-137

Apostol, B.L., Westaway, S.K., Abelson, J., and Greer, C.L. (1991) *J. Biol. Chem.* **266**, 7445-7455

Baccanari, D., Phillips, A., Smith, S., Sinski, D., and Burchall, J. (1975) *Biochemistry* **14**, 5267-5273

Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248-254

Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979) *Cell* **18**, 37-45

Laemmli, U.K. (1970) *Nature* **227**, 680-685

McCraith, S.M., and Phizicky, E.M. (1991) *J. Biol. Chem.* **266**, 11986-11992

Normanly, J., Ogden, R.C., Horvath, S.J., and Abelson, J. (1986) *Nature* **321**, 213-219

Ogden, R. C., Lee, M.-C., and Knapp, G. (1984) *Nucleic Acids Res.* **12**, 9367-9382

Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536

Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J. (1979) *Cell* **18**, 27-35

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1983) *J. Biol. Chem.* **261**, 2978-2986

Rauhut, R., Green, P.R., and Abelson, J. (1990) *J. Biol. Chem.* **265**, 18180-18184

Reyes, V.M. and Abelson, J. (1987) *Anal. Biochem.* **166**, 90-106

Richardson, C.C. (1981) in *The Enzymes* (Boyer, P.D., ed) 3rd Ed, Vol. 14, 299-314, Academic Press, New York

Ruby, S.W., Goelz, S.E., Hostomsky, Z., and Abelson, J.N. (1990) *Methods Enzymol.* **181**, 97-121

Uhlenbeck, O. C., and Gumport, R. I. (1982) in *The Enzymes* (Boyer, P. D., ed.) Vol. 15, pp. 31-58, Academic Press, Orlando, FL

Volckaert, G., Min Jou, W., and Fiers, W. (1976) *Anal. Biochem.* **72**, 433-446

Westaway, S.K., Phizicky, E.M., and Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171-3176

Xu, Q., Teplow, D., Lee, T.D., and Abelson, J. (1990) *Biochemistry* **29**, 6132-6138

FIGURE LEGENDS

FIG. 1. **Origin of the 2'-phosphate of tRNA Ligated by DAKC and DAC.** Precursor $tRNA^{Phe}$ (pre-$tRNA^{Phe}$) radiolabeled with each of four ribonucleotides was treated with yeast tRNA splicing endonuclease, joined by DAKC or DAC in the presence of unlabeled ATP, the ligated product resolved on a polyacrylamide gel, eluted, digested with P1 nuclease, and analyzed by thin layer chromatography (TLC). The radiolabeled P1 nuclease-resistant dinucleotide, if any, was digested with snake venom phosphodiesterase (SVP) and analyzed by TLC as described in *Experimental Procedures*. *A* is a diagram of expected splicing and nuclease digestion products. Spliced product showing only the nucleotides surrounding the ligation junction is shown at the top. The products of P1 nuclease digestion of spliced product, which are radiolabeled mononucleotides and a P1-nuclease-resistant junction dinucleotide resulting from the presence of the 2'-phosphate, are shown in the middle. The products of SVP digestion of eluted P1-nuclease-resistant dinucleotide are 5'-adenosine monophosphate and 2',5'-guanosine diphosphate, and are shown at the bottom. * indicates those phosphates which would be radioactive in an [α-^{32}P]GTP-labeled substrate. Δ indicates those phosphates which would be radioactive in an [α-^{32}P]ATP-labeled substrate. † indicates those phosphates which would be radioactive in any [α-^{32}P]NTP-labeled substrate. *B* shows a representative autoradiograph of tRNA splicing by DAKC and DAC. *Pre-tRNA* is pre-$tRNA^{Phe}$ radiolabeled with [α-^{32}P]ATP, *tRNA* is spliced product containing a 2'-phosphate at the splice junction, and *halves* indicate unspliced tRNA halves. *C* is an autoradiograph of a TLC plate showing the P1 nuclease digestion products of tRNA spliced by DAKC or DAC. Location of unlabeled nucleotide markers, P1 nuclease-resistant junction dinucleotide, and origin (*O*) are shown at right. *Labeled nt* indicates which [α-^{32}P]NTP was used to radiolabel the pre-tRNA. Those lanes containing no protein or labeled nucleotide are controls containing only carrier RNA and P1 nuclease. Dotted circles indicate the migration of the products of P1 nuclease

digestion of carrier RNA. *D* is an autoradiograph of a TLC plate showing SVP digestion products of eluted dinucleotide spots from *C*. Appropriate markers are at left. *SVP*, (+ or -), indicates presence or absence of SVP. Solid circled spots indicate migration of authentic 2',5'-guanosine diphosphate marker visualized under ultraviolet light.

FIG. 2. **Cofactor Requirements in Splicing Reactions with DAKC and DAC.** Splicing activity was assayed as described in *Experimental Procedures*. Splicing reactions were resolved by electrophoresis in 12% polyacrylamide, 8M urea gels and quantitated by measuring Cerenkov radiation in gel slices. Percent joined was calculated a ((cpm of spliced product)/(cpm of spliced product + cpm of halves)) X 100. *A*, a representative autoradiograph of a splicing gel shows an ATP titration with 50 μg/ml (0.4 μM) DAKC. The millimolar concentration of ATP in each lane is shown at the top. The location of tRNA species in the reaction is shown at left: *pre-tRNA*, uncut precursor tRNA; *tRNA*, spliced product; *halves*, unjoined tRNA halves. *B*, the yield of spliced product in splicing reactions (expressed as *% Joined*) is plotted as a function of increasing ATP or CTP, up to 5 mM. Symbols are: DAKC with ATP (filled circles), DAKC with CTP (open circles), DAC with ATP (filled squares), DAC with CTP (open squares). *C*, *% Joined* is plotted as a function of increasing *GTP*, up to 25 μM. Symbols are: DAKC (filled circles), DAC (filled squares). *D*, *% Joined* is plotted as a function of increasing *UTP*, up to 1 mM. Symbols are the same as in *C*.

FIG. 3. **Polynucleotide Kinase Activity of DAKC and DAC.** Polynucleotide kinase activity was assayed as described in *Experimental Procedures*. Reaction products were resolved by electrophoresis in 18% polyacrylamide, 8M urea gels and quantitated by measuring Cerenkov radiation in gel slices. Percent phosphorylated was calculated as ((cpm of gel-shifted kinased band)/(total cpm per lane)) X 100. *A*, an autoradiograph of the kinase gel is shown, depicting the gel-shift of RNA in an ATP titration with DAKC.

The micromolar concentration of ATP in each lane is shown at the top. *U* indicates unphosphorylated oligonucleotide and *P* indicates phosphorylated oligonucleotide. *CIP*, (+ or -), indicates whether oligonucleotide substrates were pretreated with alkaline phosphatase to remove 5'-phosphates before the kinase assay. *DAKC* indicates the presence of the DHFR-full-length ligase fusion protein. *T4 PNK* is a control lane containing T4 polynucleotide kinase and ATP. *B*, the yield in phosphorylation reactions (expressed as *% Phosphorylated*) is plotted as a function of increasing ATP or CTP, up to 5-10 mM. Symbols are as follows: DAKC with ATP (filled circles), DAKC with CTP (open circles), DAC with ATP (filled squares), DAC with CTP (open squares). *C*, *% Phosphorylated* is plotted as a function of increasing *GTP*, up to 5 mM. Symbols are: DAKC (filled circles), DAC (filled squares). *D*, *% Phosphorylated* is plotted as a function of increasing *UTP*, up to 5 mM. Symbols are the same as in *C*.

FIG. 4. **Incorporation of the γ-phosphate of GTP as the Splice Junction Phosphate.** The analyses of splicing products from reactions containing either low-specific activity or high-specific activity [α-^{32}P]UTP-labeled biotinylated tRNA substrate, either nonradioactive ATP or [γ-^{32}P]GTP, and either DAKC or DAC, are described in *Experimental Procedures*. *A* is an autoradiograph of the splicing reaction. Lanes labeled with *^{32}P-tRNA* at top contain both low-specific activity and high-specific activity pre-tRNA^{Phe}, nonradioactive ATP, and enzymes. Lane labeled *C* contains only low-specific activity pre-tRNA^{Phe}, nonradioactive ATP, and enzymes. Lanes labeled *γ-^{32}P-GTP* contain only low-specific activity pre-tRNA^{Phe} incubated in the presence of [γ-^{32}P]GTP and enzymes. *B* is a diagram depicting the expected splicing and nuclease digestion products. The top line represents spliced product showing only the nucleotides surrounding the splice junction. In the middle are the products of P1 nuclease digestion. At the bottom are the products of snake venom phosphodiesterase (SVP) digestion. **p* indicates radiolabeled phosphate incorporated from the γ-phosphate of [γ-^{32}P]GTP. *C* is

an autoradiograph of a thin-layer chromatography (TLC) plate showing P1 nuclease digestion products of eluted tRNA bands from panel *A*. *O* indicates the origin. Positions of carrier RNA digestion products and P1 nuclease-resistant junction dinucleotide are shown at left. Positions of radiolabeled nucleotides are shown at right. *D* is an autoradiograph of the TLC plate showing the SVP digestion product of the eluted junction dinucleotide from panel *C*.

FIG. 5. **Mechanism for tRNA Splicing in Yeast.** A diagram is presented describing the mechanism for tRNA splicing in yeast, based on the model described in the text for the utilization *in vitro* of two different nucleotide triphosphates during tRNA splicing.

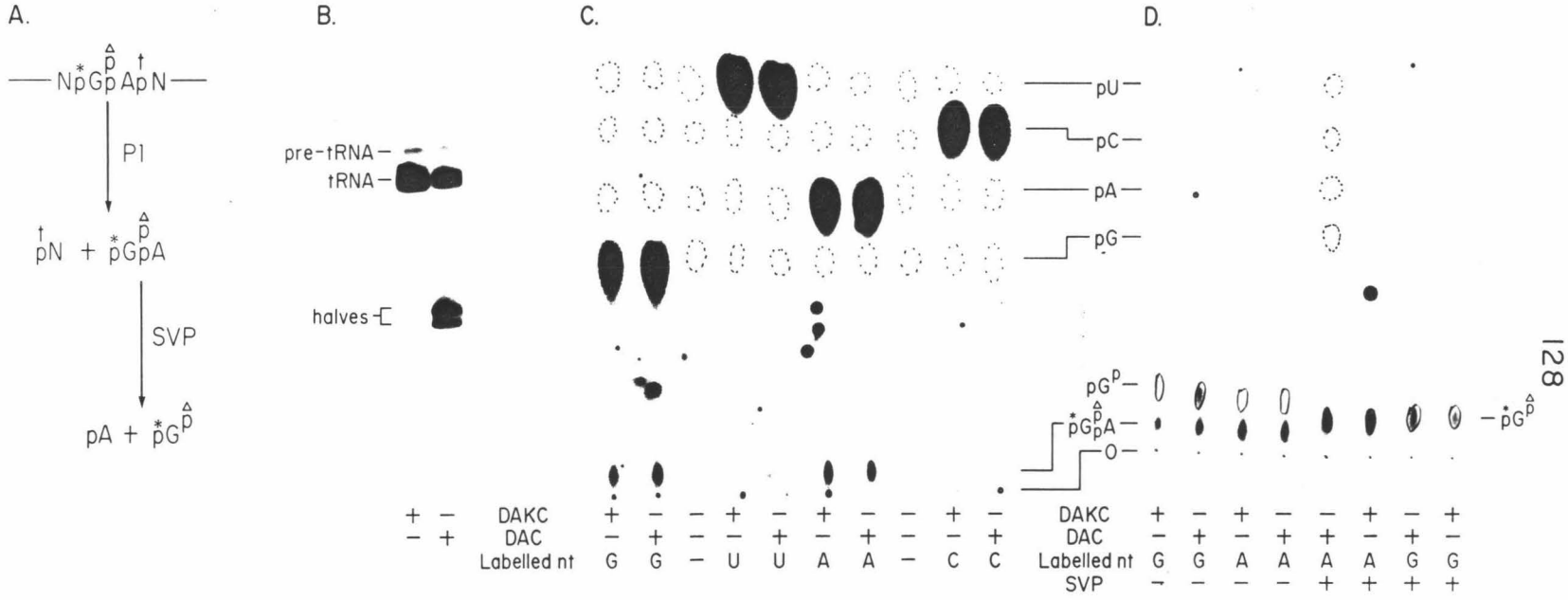

A.
B.
C.
D.
P1
SVP
pre-tRNA
tRNA
halves
pU
pC
pA
pG
DAKC
DAC
Labelled nt
SVP

Figure 1

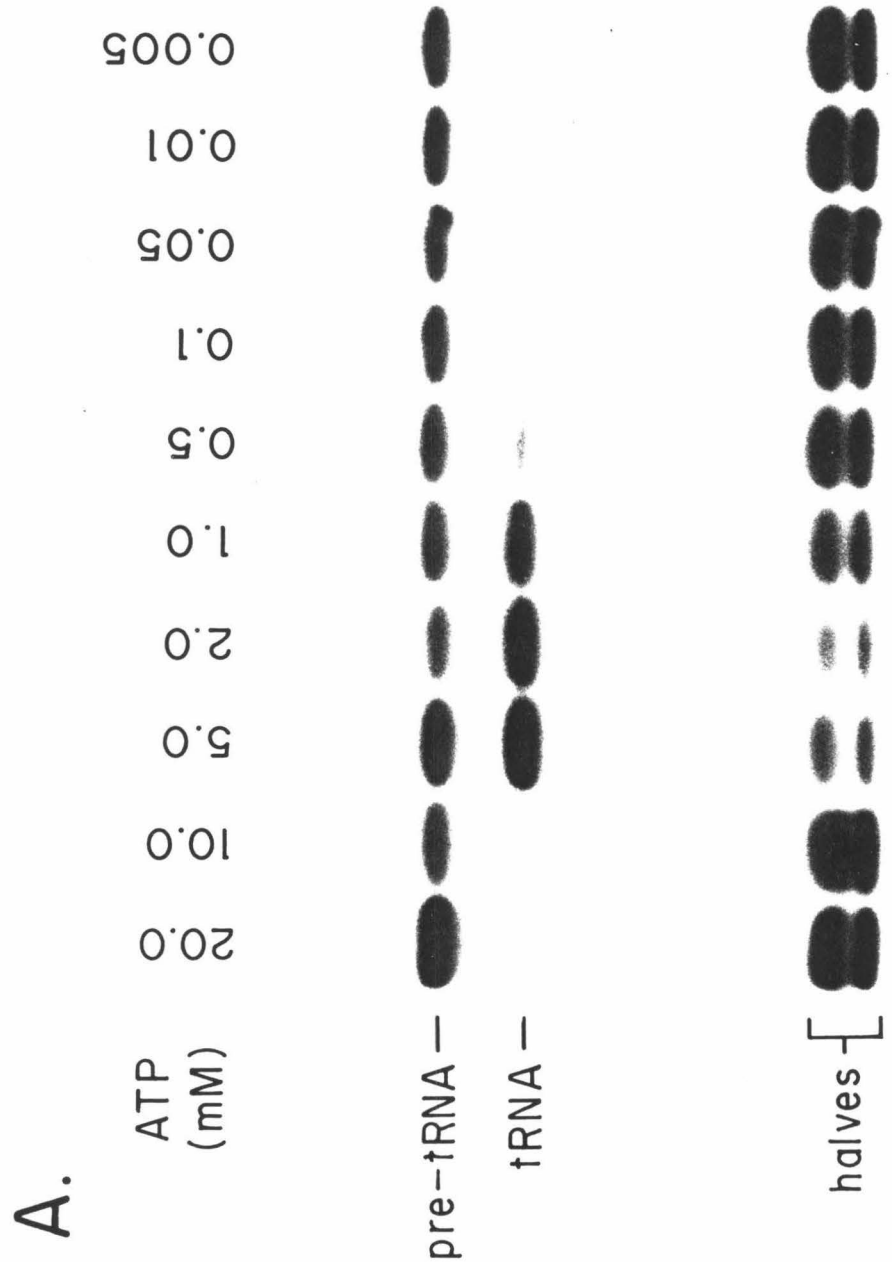


Figure 2A

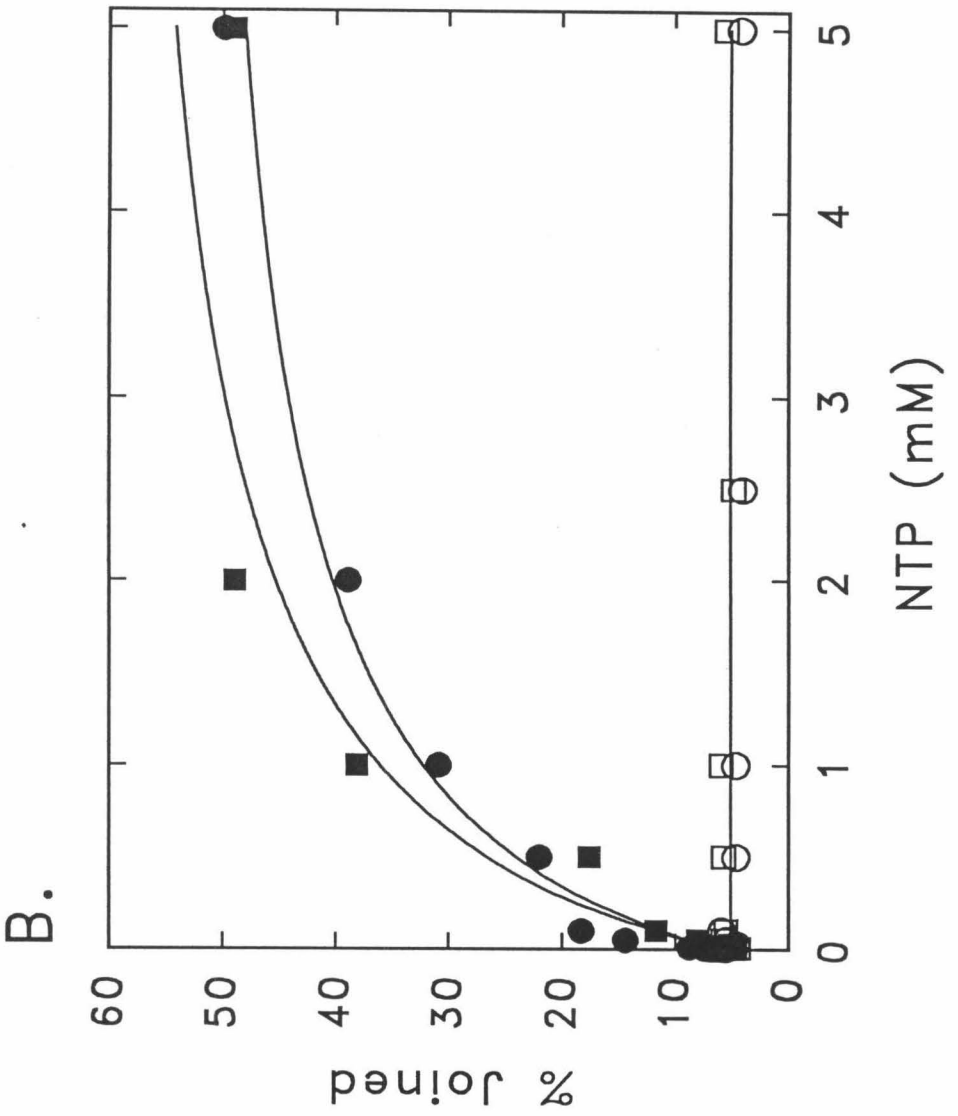


Figure 2B

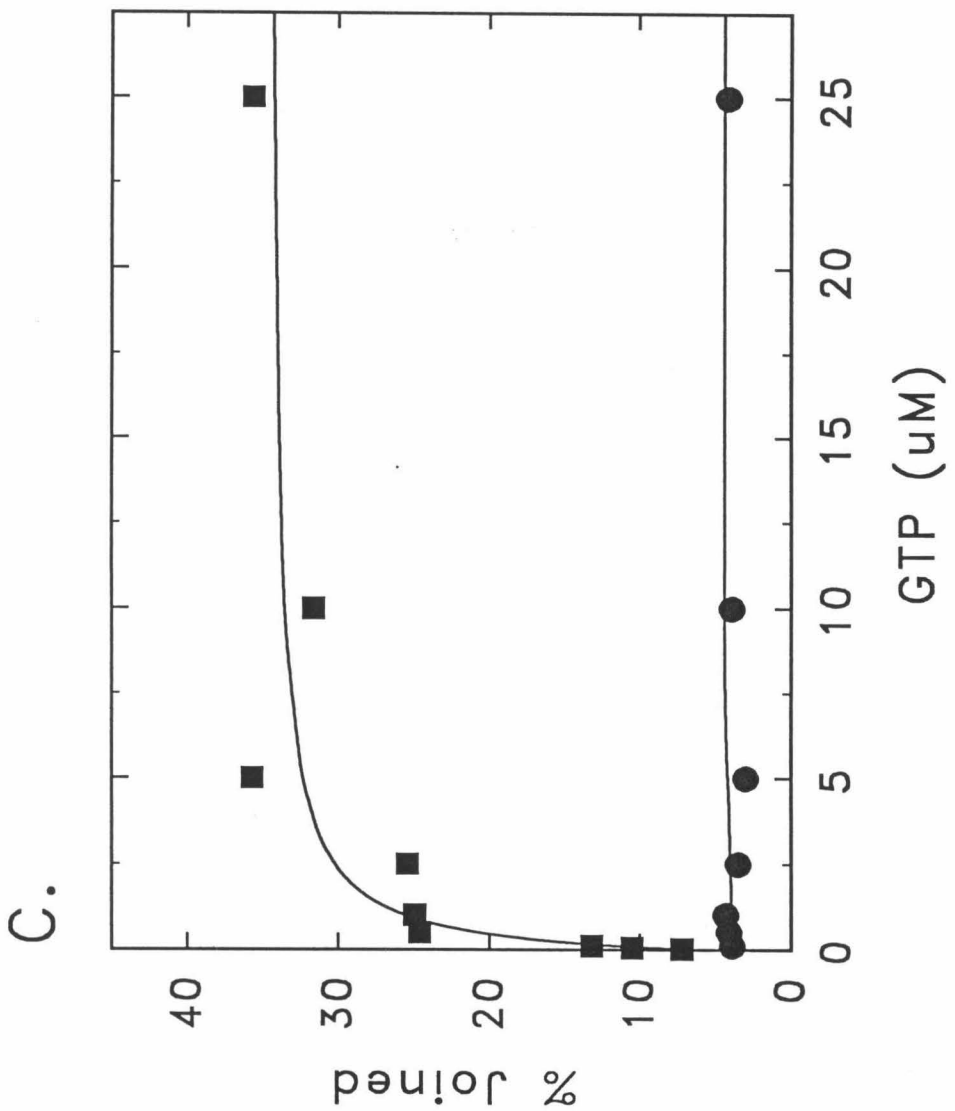


Figure 2C

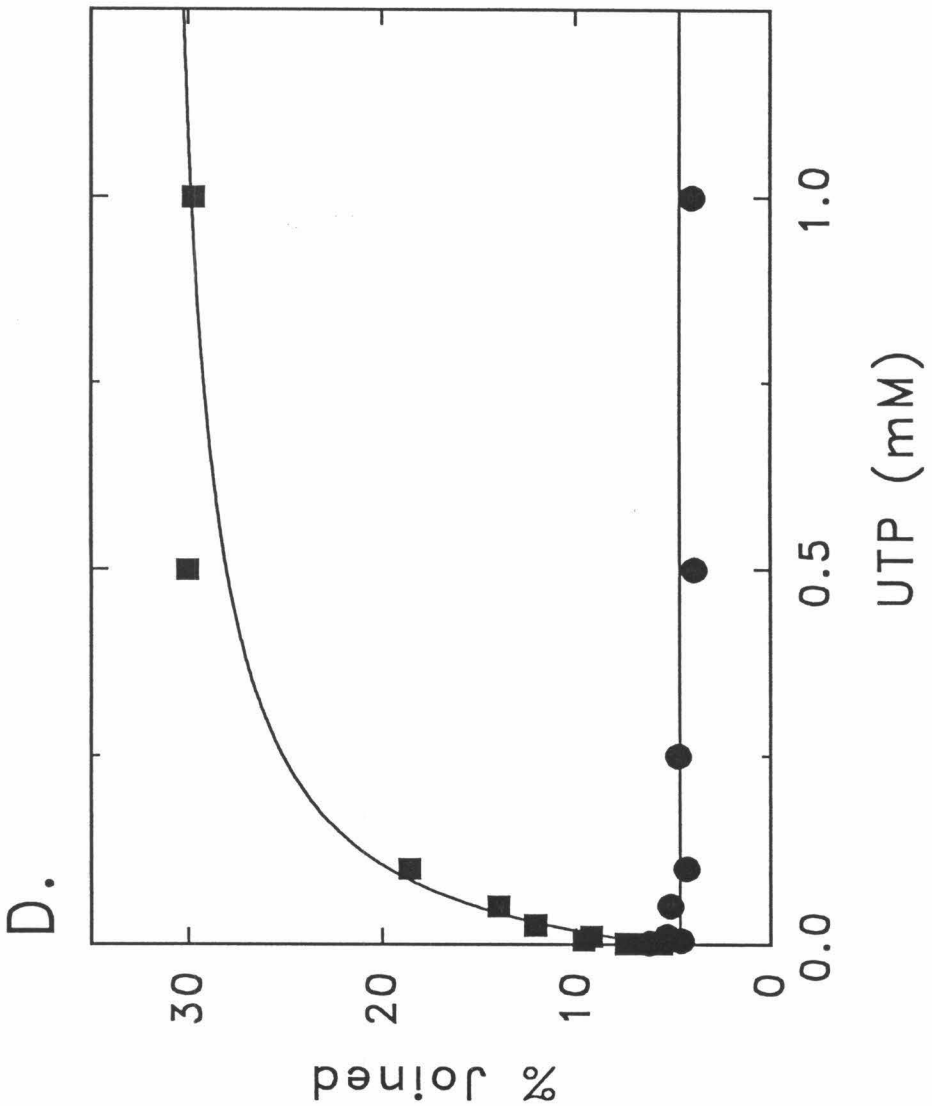


Figure 2D

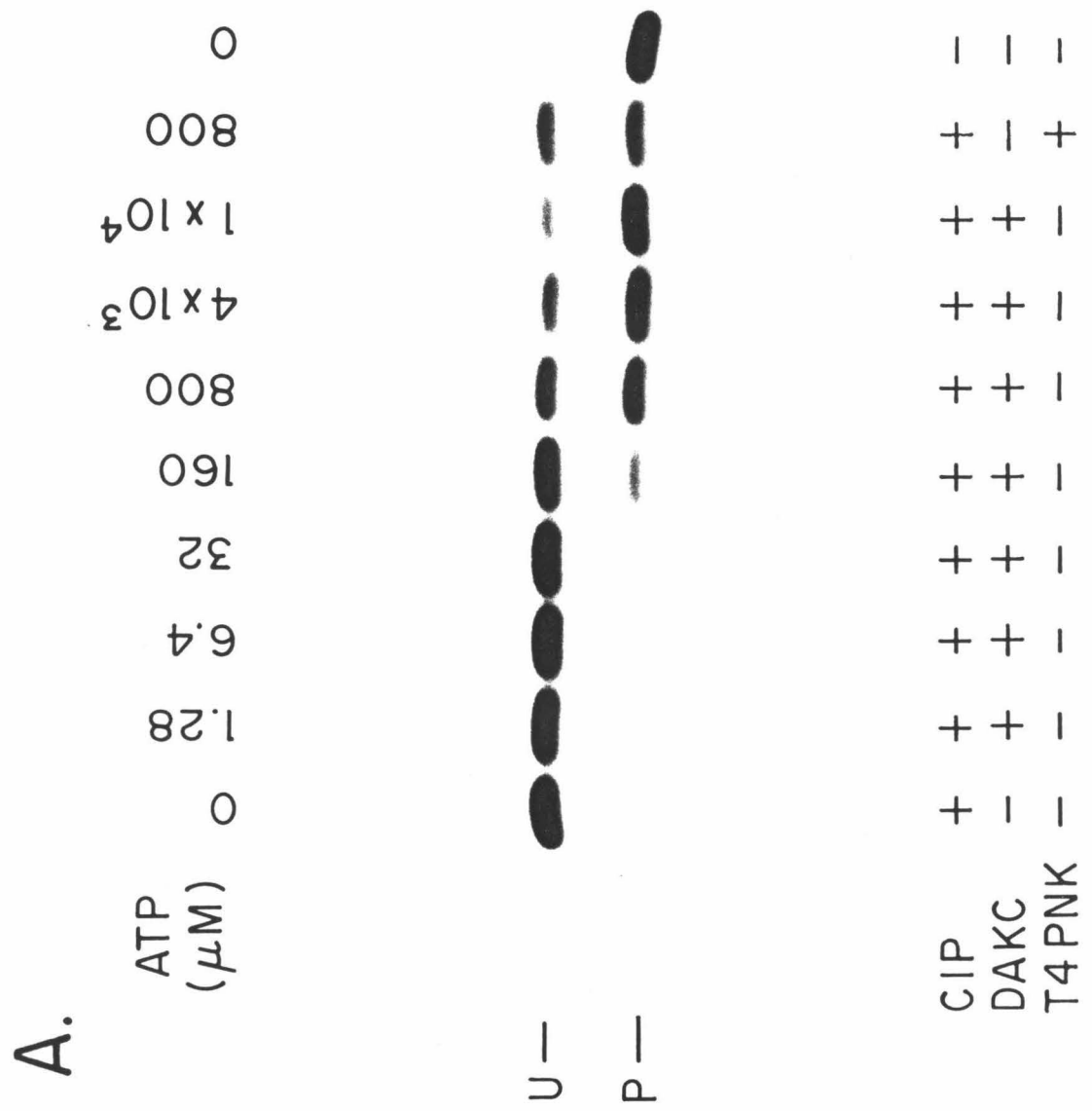


Figure 3A

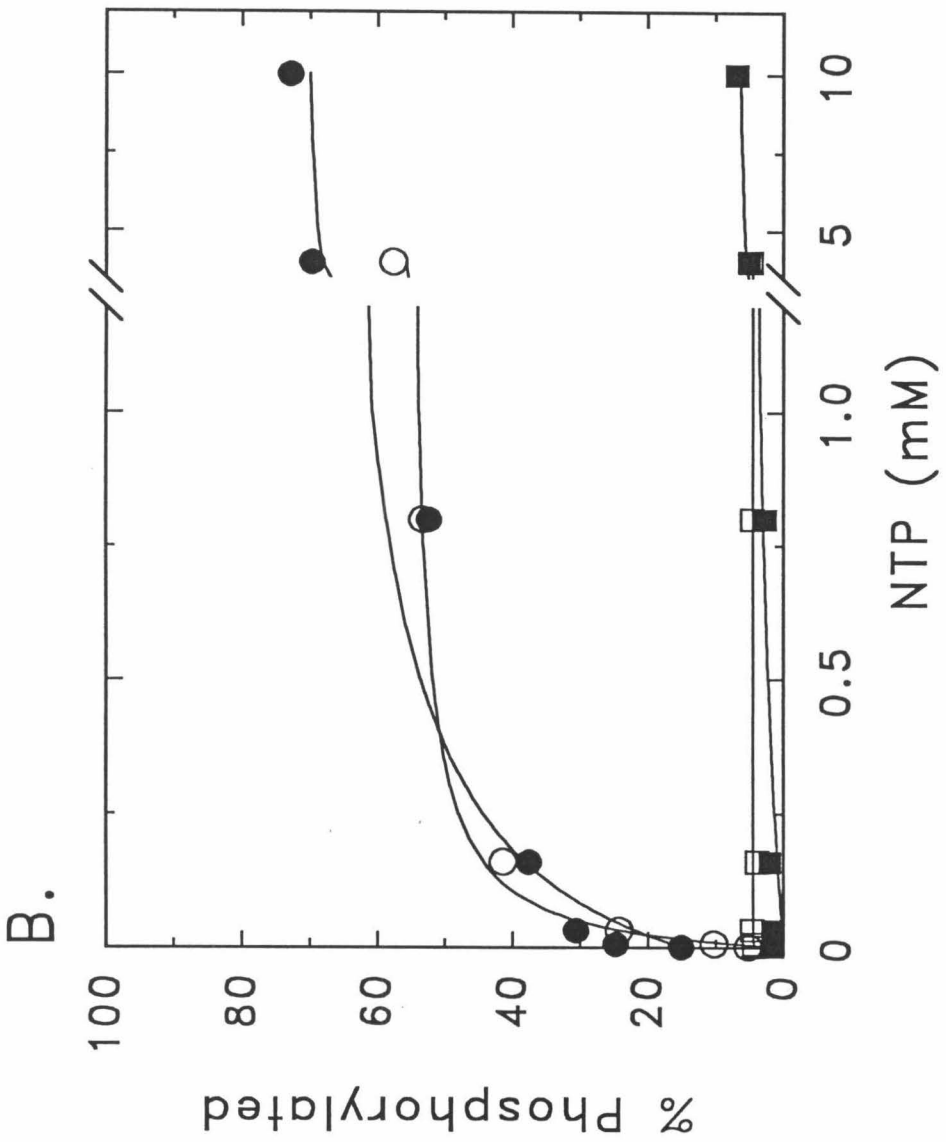


Figure 3B

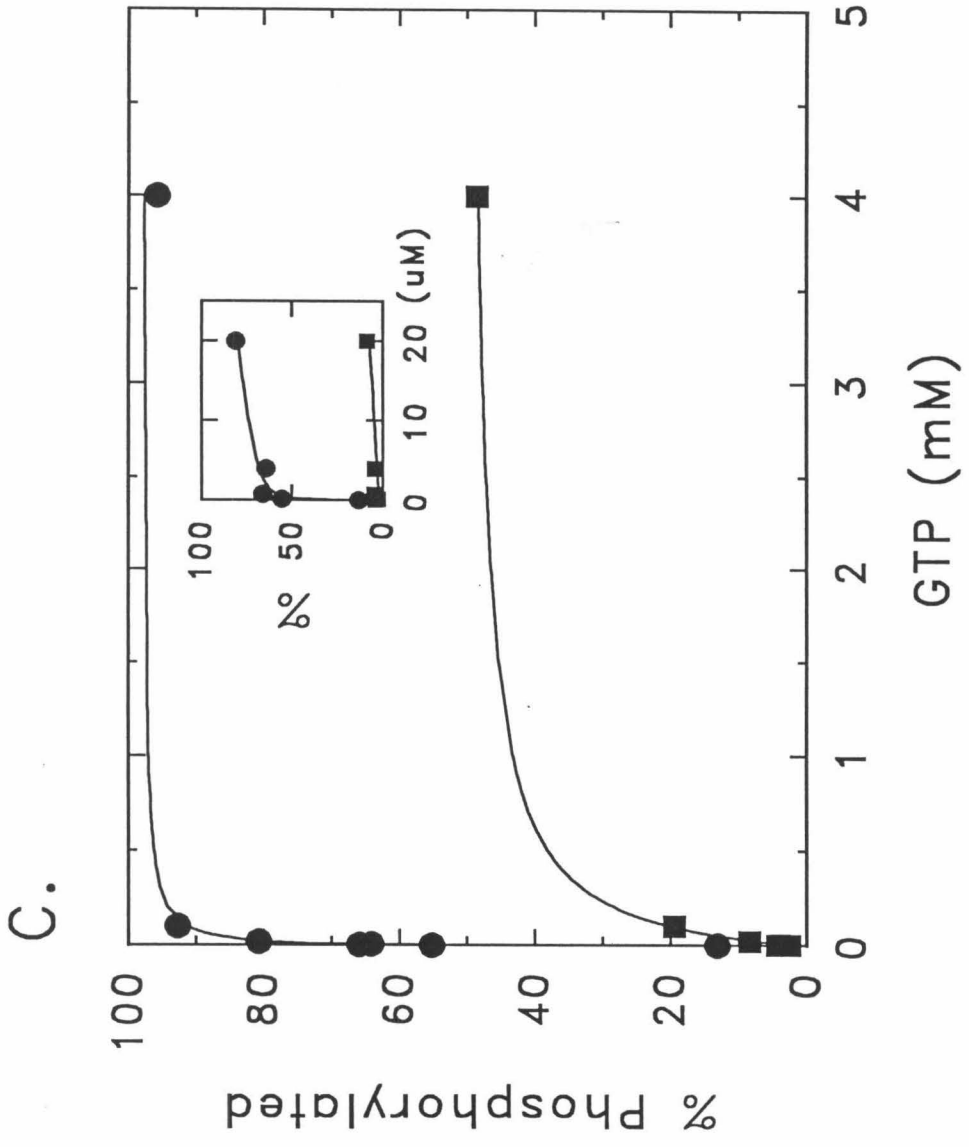

C.
% Phosphorylated
100
80
60
40
20
0
0
1
2
3
4
5
GTP (mM)
%
100
50
0
0
10
20 (uM)

Figure 3C

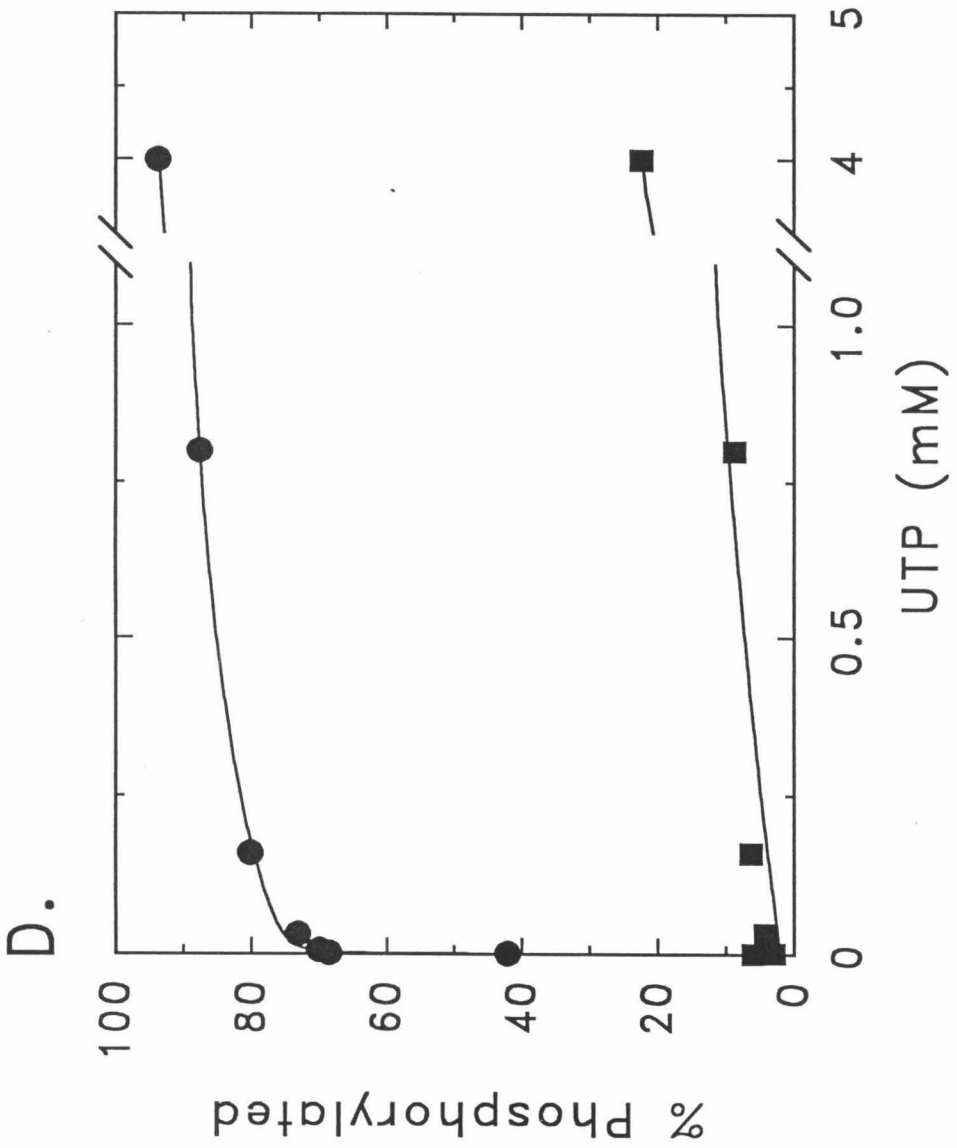


Figure 3D

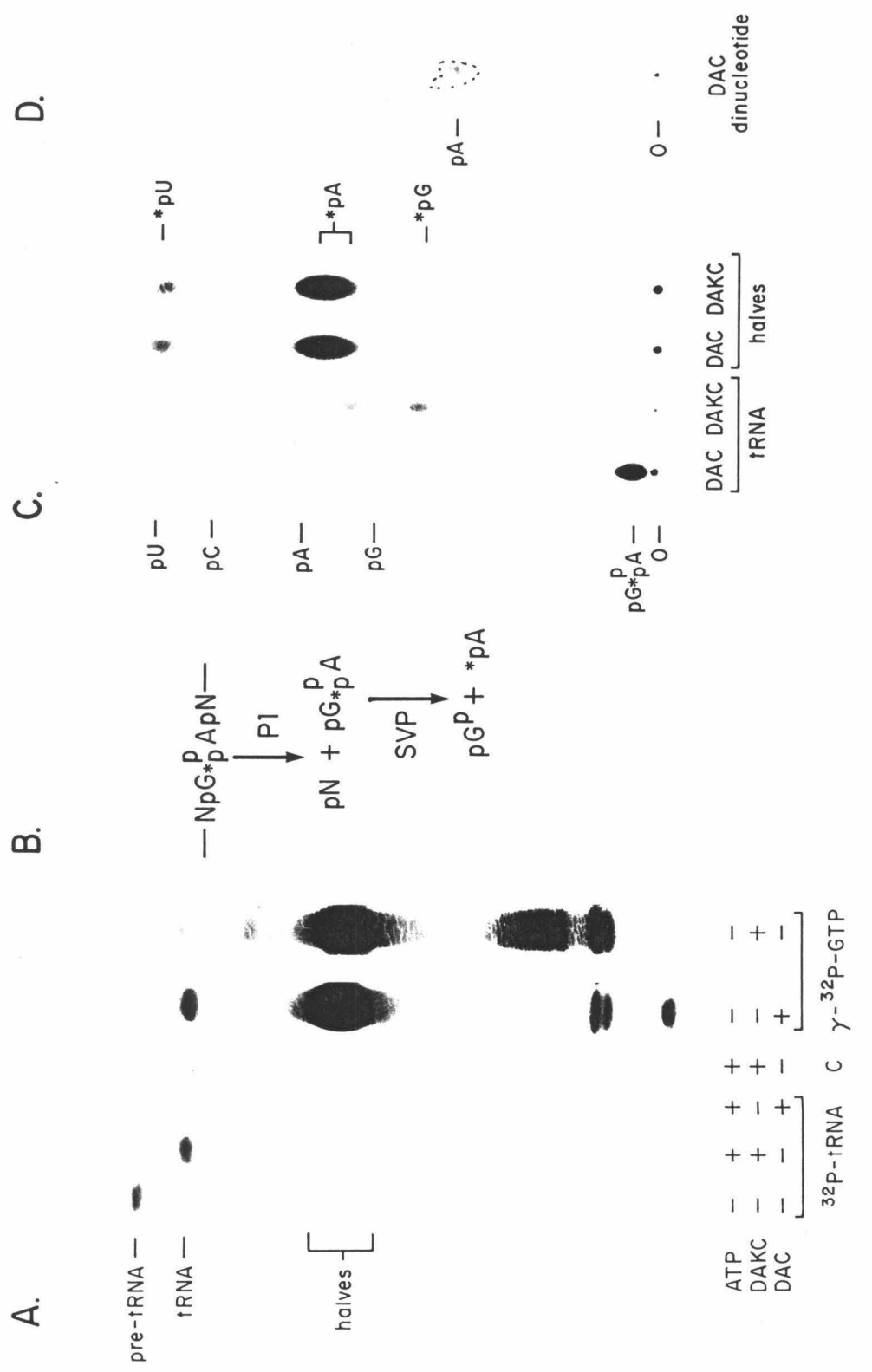

A.
B.
C.
D.
pre-tRNA
tRNA
halves
ATP
DAKC
DAC
32P-tRNA
C
γ-32P-GTP
P1
SVP
pU
pC
pA
pG
*pU
*pA
*pG
pA
O
DAC DAKC DAC DAKC
tRNA
halves
DAC
dinucleotide

Figure 4

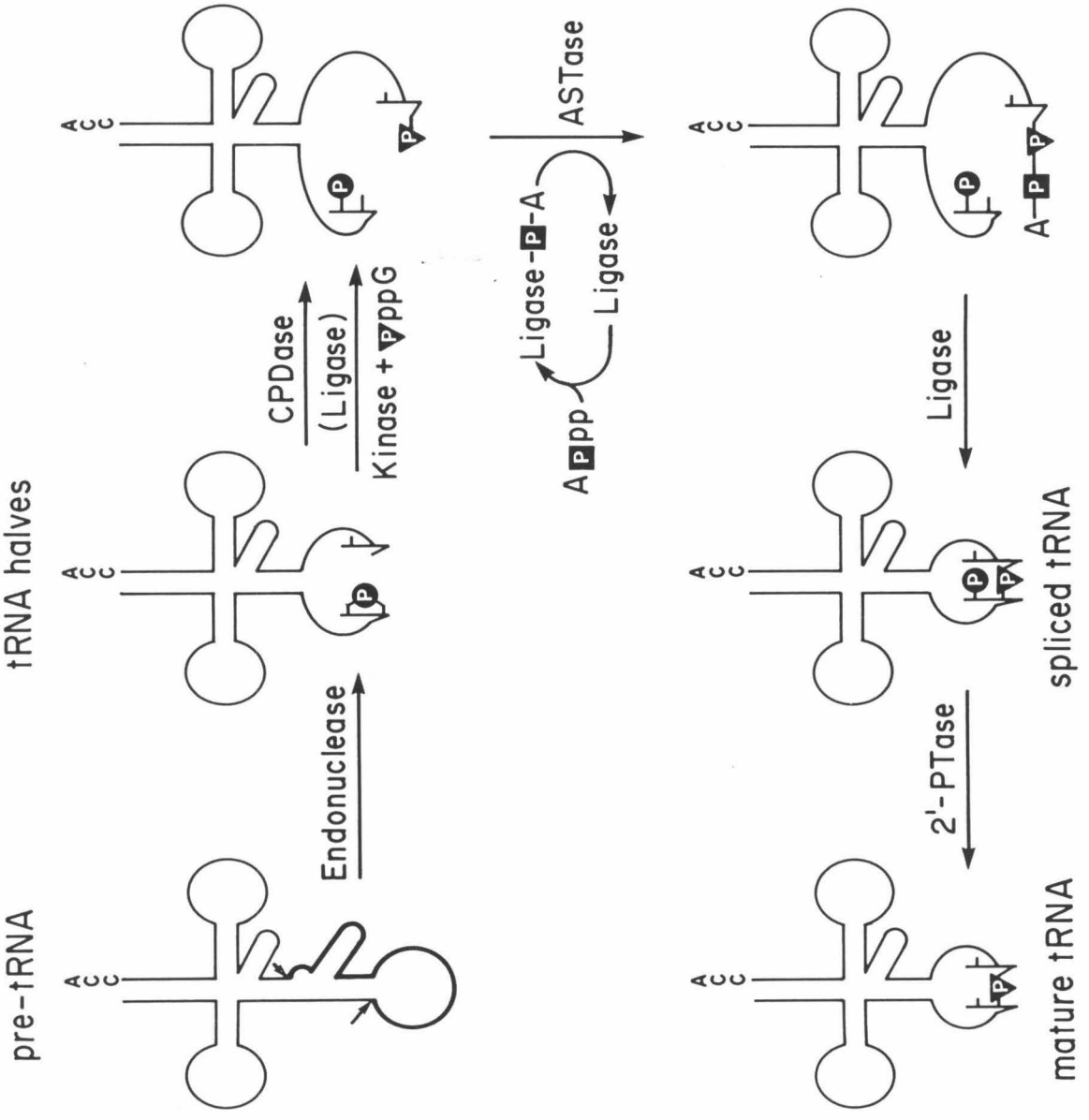

pre-tRNA
tRNA halves
Endonuclease
CPDase
(Ligase)
Kinase + pppG
Ligase-p-A
A p pp
Ligase
ASTase
2'-PTase
Ligase
mature tRNA
spliced tRNA

Figure 5

CHAPTER IV

Multiple Nucleotide Cofactor Use by Yeast Ligase in tRNA Splicing:
EVIDENCE FOR INDEPENDENT ATP AND GTP BINDING SITES

Heather G. Belford[@], Shawn K. Westaway[#], John Abelson[#],
and Chris L. Greer[@]

[@]Department of Biological Chemistry, College of Medicine,
University of California, Irvine; and
[#]Division of Biology, California Institute of Technology, Pasadena, CA 91125

[J. Biol. Chem. (1993), in press]

SUMMARY

We have examined multiple cofactor usage by yeast tRNA ligase in splicing *in vitro*. The ligase mechanism of action requires expenditure of two molar equivalents of nucleotide cofactor per mole of tRNA product. Recent evidence [Westaway, S. K., Belford, H. G., Apostol, B. A., Abelson, J. and Greer, C. L. (1992) *J. Biol. Chem.*, accompanying manuscript] demonstrated that the ligase-associated kinase activity is more efficient with GTP as cofactor than with ATP. Employing a ligase fusion construct with dihydrofolate reductase [Apostol, B. L., Westaway, S. K., Abelson, J. and Greer, C. L. (1991) *J. Biol. Chem.* **266**, 7445-7455] for purposes of enzyme purification, we performed joining assays demonstrating that ATP and GTP are the most effective combination of cofactors. ATP was essential to the joining reaction, while UTP, CTP, or ATP replaced GTP inefficiently. Specific and functionally independent binding sites were confirmed for ATP and GTP by direct binding measurement. A third site was implicated in UTP- and CTP-ligase interactions. Comparison of binding constants with K_{app} values determined for nucleotide-dependent joining suggested both that NTP binding may be limiting in tRNA joining and that tRNA ligation occurs most efficiently using GTP for the kinase reaction and ATP as the adenylylate synthetase cofactor.

INTRODUCTION

A significant number of nuclear tRNA genes from a wide range of eukaryotes contain a single intervening sequence (IVS) located uniformly one nucleotide downstream of the anticodon triplet position (Perlman *et al.*, 1990). In the yeast, *Saccharomyces cerevisiae*, an intervening sequence is present in about 20% of the tRNA genes comprising 10 isoaccepting families (Abelson, 1991). Penetrance is complete in that either all or none of the genes in a family contain an IVS. A required step in processing precursor tRNA

transcripts (pre-tRNAs) to functional maturity is splicing; the precise removal of the intervening sequence and subsequent ligation of the tRNA halves. Mutations which specifically prevent splicing of pre-tRNA to mature tRNA sequence are lethal (Phizicky *et al.*, 1992). Thus tRNA splicing is essential and the mechanism of this reaction has been the subject of our research.. The focus of the present study is the ligation of tRNA half molecules subsequent to the removal of the IVS.

IVS excision in yeast is accomplished by tRNA endonuclease which is a multisubunit, site-specific enzyme associated with the nuclear membrane (Peebles *et al.*, 1983; Rauhut *et al.*, 1990). The endonuclease generates a cyclic 2',3'-phosphodiester bond at the 3' end of the 5' half tRNA and a 5'-OH at the 5' terminus of the 3' half tRNA (Peebles *et al.*, 1983). These ends are joined by tRNA ligase which must perform three steps prior to the final ligation of halves (Greer *et al.*, 1983). The cyclic phosphate bond is opened to a 2'-phosphomonoester. The 3' half is phosphorylated at the 5'-OH position and the phosphorylated 3' half is activated by the transfer of AMP to the 5'-phosphate in a phosphoanhydride linkage. Subsequently the tRNA halves are joined and AMP released. The result is a splice junction structure with a 2'-phosphomonoester derived from the 5' splice site and a 3',5'-phosphodiester bond the phosphate of which derives from the cofactor, not from the tRNA. Inherent in this ligase reaction mechanism is a requirement for at least two molar equivalents of cofactor per mole of pre-tRNA spliced. The final step in the splicing reaction is the removal of the 2'-phosphate by an NAD^+-dependent phosphotransferase activity recently characterized by McCraith and Phizicky (1991).

The yeast tRNA ligase is a single polypeptide of 827 amino acids with an apparent molecular mass of 95 kDa containing all the enzymatic activities necessary for ligation (Greer *et al.*, 1983; Phizicky *et al.*, 1986; Westaway *et al.*, 1988) as well as binding sites for tRNA substrates (Apostol and Greer, 1991). Lysine-114 has been characterized as the enzyme adenylylation site (Xu *et al.*, 1990). The ligase gene is single copy and has been isolated (Phizicky *et al.*, 1986) and sequenced (Westaway *et al.*, 1988). Each ligase

molecule contains a single high affinity tRNA substrate binding site (Apostol and Greer, 1991) which is complex in nature requiring the minimum combination of two or more elements dispersed throughout the polypeptide linear sequence (Apostol *et al.*, 1991).

Previous work (Apostol *et al.*, 1991) has shown that the enzymatic domains are distributed within the protein sequence in a somewhat linear fashion and in accord with predictions made on the basis of sequence comparisons primarily with phage T4 RNA ligase and T4 polynucleotide kinase. Xu *et al.* (1990) generated proteolytic fragments of tRNA ligase which were shown to retain a subset of the normal enzymatic activities. Extending this work, Apostol *et al.* (1991) designed ligase gene deletions which were expressed in *E. coli* as fusions with bacterial dihydrofolate reductase (DHFR). The DHFR segment (D domain) was used to purify the ligase constructs by affinity chromatography with methotrexate agarose resin and did not affect the ligase activities measured. By comparing enzymatic activities of the full length ligase polypeptide (DAKC) to those of the deletion constructs, it was demonstrated that adenylylate synthetase activity is located within the amino-terminal half of the polypeptide (domain A), while cyclic phosphodiesterase activity is located within the carboxy-terminal quarter of the protein (domain C). A minimal kinase position was not so clearly delimited. The removal of any one of the terminal or middle regions of the ligase polypeptide significantly diminished or abolished kinase activity when assayed with ATP (Apostol *et al.*, 1991). Subsequent data (Westaway *et al.*, accompanying manuscript) suggested that ligase prefers to use GTP as cofactor in the kinase reaction over ATP by a margin of approximately 1000-fold and that the deletion of amino acids 396-595 (K domain; Apostol *et al.*, 1991) does not significantly reduce GTP-dependent kinase activity. Indeed, a primary function of the K domain may be determination of protein structure and substrate specificity. A further important observation from this latter work was that DAKC demonstrates CTP- and UTP-dependent kinase activity with K_{app} values intermediate between those for GTP and ATP. The novel inference is that ligase might bind and use multiple nucleotide triphosphate cofactors.

In this paper we have explored the potential and practicality of multiple cofactor usage. We employed the DAKC fusion gene construct (Apostol *et al.*, 1991) expressed in *E. coli* as a source of full length ligase polypeptide because of the advantages of reliable and efficient purification of quantities of active enzyme. Joining assays were used to demonstrate that while ATP is essential and sufficient for joining, ATP and GTP in micromolar concentrations are the most effective combination of cofactors. CTP or UTP may also replace GTP inefficiently. Functionally independent binding sites were confirmed for ATP and GTP and the evidence suggests that nucleoside triphosphate (NTP) binding may be a rate limiting step in tRNA joining. UTP and CTP binding characteristics suggest a third NTP binding site for ligase. We postulate that tRNA ligation occurs most efficiently using GTP for the kinase reaction and ATP as the adenylylate synthetase cofactor.

EXPERIMENTAL PROCEDURES

Materials----Restriction enzyme *Bst*N1 was obtained from New England Biolabs. Immunopure sulfosuccinimidyl-2-(biotinamido)ethyl-1,3'-dithiopropionate (NHS-SS-Biotin) were from Pierce. Bovine serum albumin (BSA) was obtained from BRL or ICN (Fraction V, BSA). T7 RNA polymerase and Ultrapure nucleotide triphosphates: ATP, CTP, GTP and UTP, were obtained from Pharmacia. Labeled nucleotides: [α-^{32}P]ATP (3000 Ci/mmol), [α-^{32}P]CTP (3000 Ci/mmol), [α-^{32}P]GTP (3000 Ci/mmol), [α-^{32}P]UTP (3000 Ci/mmol) and [γ-^{32}P]GTP (5000 Ci/mmol), were obtained from Amersham. N^6-(6-Aminohexyl)ATP was from BRL. Polygram CEL 300 PEI plates for thin layer chromatography were obtained from Brinkmann Instruments, Inc. Highly purified yeast tRNA endonuclease (4.8 μg protein/ml) was the kind gift of R. Rauhut and was the equivalent of the Sepharose CL-6B purified fraction (Rauhut *et al.*, 1990). E·Y succinyl avidin and E·Y biotin Sephacryl S1000^{TM} resin were from E·Y Labs, Inc. (San Mateo, CA). Ultrafree-MC 30,000 NMWL and 10,000 NMWL polysulfone Type PTTK

filter units were obtained from Millipore. Nuclease P1 was from Calbiochem and snake venom phosphodiesterase (SVP) was from Boehringer Mannheim.

Strains and plasmids----Plasmid pDAKC contains the full length yeast tRNA ligase gene fused to the carboxyl terminus of the *E. coli* dihydrofolate reductase (DHFR) gene all under control of the *E. coli* tac promoter as described by Apostol *et al.* (1991). Plasmid ptac10 has the DHFR gene alone under control of the tac promoter (Normanly *et al.*, 1986). *E. coli* strain RJ1438 [*thi*1 *leuB*6 *thr*1 *lacY*1 *tonA*21 *supE*44 *hsdR galK* Δ*lon*100 (F' *lacI*^{SQ} *fzz*::Tn5)], used for expression of DHFR and the DHFR-ligase fusion protein, was kindly provided by R. Johnson. Plasmid pUC13Phe containing the T7 promoter/pre-tRNA construct is described in Reyes and Abelson (1987).

Cell growth and protein purification----RJ1438 cells transformed with pDAKC or ptac10 were grown, harvested, extracted, and the ligase or DHFR polypeptides purified as described by Westaway *et al.* (accompanying manuscript). Yields were typically 0.2 mg per liter culture. Specific activities of the DAKC ligase varied from 700-9000 U/mg protein (1 Unit = 1 pmol tRNA joined per minute; Greer *et al.*, 1983).

Authentic yeast ligase used in this study was a partially purified soluble fraction from *S. cerevisiae* containing 0.18 mg/ml total protein with joining activity of 1100 U/mg protein.

Pre-tRNA substrate preparation----*In vitro* transcription with T7 RNA polymerase and pUC13Phe linearized with *Bst*N1 was used to generate radiolabeled pre-tRNA^{Phe} transcripts of authentic sequence and mature ends (Reyes and Abelson, 1987). To yield precursor of high specific activity (4.8 X 10^3 dpm/fmol), reactions containing 50 mM Tris-HCl (pH 8), 10 mM $MgCl_2$, 10 mM NaCl, 10 mM DTT, 0.5 mM each CTP, ATP and GTP, 0.01 mM UTP, 100 μCi [α-^{32}P]UTP, 4 μg restricted plasmid template and 250 U

T7 RNA polymerase in 100 µl were incubated 60 min at 37°C. Reactions were stopped with the addition of 0.2 mg/ml proteinase K, 0.2% SDS, 10 mM EDTA (stop mix) for 15 min at 50°C, adjusted to 0.4 M ammonium acetate, pH 7, and precipitated with ethanol. RNA products were purified by acrylamide gel electrophoresis (Apostol *et al.*, 1991).

Preparation of biotinylated transcripts was done in two steps to yield substituted RNAs from which the biotin moieties could be specifically cleaved. In step 1, T7 RNA polymerase reactions were carried out as above except that ATP concentration was reduced to 250 µM and 250 µM N^6-(6-Aminohexyl)ATP was added. In step 2, gel-purified, N^6-(6-Aminohexyl)ATP substituted pre-tRNA transcripts were biotinylated by reaction with NHS-SS-biotin (Ruby *et al.*, 1990). Note that at this level of substitution and biotinylation, splicing reactions are unaffected and 55-65% of transcripts can be bound to avidin-agarose resin (data not shown). To produce biotinylated precursor of low specific activity (9.6 dpm/fmol), radiolabeled and unlabeled UTP concentrations in the transcription reaction were adjusted to 10 µCi/100 µl and 0.5 mM, respectively. Biotinylated transcripts were purified by three repetitions of ethanol precipitation in the presence of 0.4 M ammonium acetate, pH 7.

Splicing reactions----Analytical splicing reactions contained 6.2 fmol pre-$tRNA^{Phe}$. Endonuclease fractions were dialyzed to remove $(NH_4)_2SO_4$. In order to minimize sample to sample variation, as much as 124 fmol pre-tRNA were first incubated with endonuclease (equivalent to approximately 3 X 10^{-5} U/fmol pre-tRNA (Rauhut *et al.*, 1990) or 0.15 ng/fmol pre-tRNA in 24 mM Hepes (pH 7.5), 6 mM $MgCl_2$, 3 mM spermidine, 0.12 mM DTT, 0.48% triton X-100, 30 mM NaCl, for 18-20 min at 30°C. Endonuclease-treated pre-tRNA (6.2 fmol in 8 µl) was added to 1 µl nucleoside triphosphate (NTP: ATP, CTP, GTP or UTP) and 1 µl ligase protein extract. Ligase concentrations were chosen operationally as the amount of ligase necessary to join 50% of available substrate (approximately 3 fmol) in a 15 min, 30°C incubation with 1 mM ATP. After 15 min

incubation at 30°C, reactions were stopped by the addition of stop mix and incubation at 50°C as described above. An equal volume of sample buffer (20% sucrose, 7 M urea, 0.1% each xylene cyanol and bromphenol blue) was added. Products were separated on 8 M urea, 12% polyacrylamide (acrylamide/bisacrylamide 30:1) gels, visualized by autoradiography and quantitated by Cerenkov cpm in gel slices.

Biotinylated pre-tRNA was used as substrate in experiments to label the splice junction phosphodiester bond. This use was necessary so that the specific splicing products of the reaction might by isolated from a background of RNAs in the DHFR-ligase enzyme fraction which were also labeled in the presence of [γ-^{32}P]GTP. As above, endonuclease was dialyzed to remove $(NH_4)_2SO_4$ and pre-tRNAs were precut (3 µg endonuclease per ml) in the absence of nucleoside triphosphates or ligase. Control splicing reactions contained 620 fmol biotinylated pre-tRNA of low specific activity and 6.2 fmol high specific activity biotinylated pre-tRNA. In reactions assaying incorporation of labeled phosphate into the splice junction, pre-tRNA of high specific activity was omitted and [γ-^{32}P]GTP was added at 100 µCi/10 µl total reaction volume by drying the labeled GTP in the tube to which the joining reaction was added. Assays were joined and stopped as described above, adjusted to 0.4 M ammonium acetate (pH 7) and precipitated with ethanol. Biotinylated products were then isolated by the bridge-binding method of Ruby *et al.* (1990) using E·Y biotin-Sephacryl combined with E·Y succinyl avidin. A brief description follows. All operations were carried out at 4°C. Biotin Sephacryl was reacted with succinyl avidin 1 h. Unbound avidin was removed by washing extensively with binding buffer (0.2 M NaCl, 10 mM Tris-HCl, pH 7.5). Splicing reaction precipitates were resuspended in binding buffer, added to the avidin-biotin-Sephacryl resin (1 µl resin/10 fmol pre-tRNA), and incubated 1 h with occasional vortexing. The resin was then washed extensively with binding buffer to remove unbound material. RNA was eluted from the washed resin with 3 X 5 volumes of 150 mM DTT, 150 mM NaCl, 40 mM Tris-acetate (pH 8.5), 3 mM $MgCl_2$. Eluted RNAs were concentrated by ethanol precipitation in

the presence of glycogen carrier (20 µg/ml). RNA precipitates were dissolved in sample buffer, and subjected to gel electrophoresis and quantification as described above.

Nuclease digestion----Nuclease P1 and snake venom phosphodiesterase (SVP) reactions were performed essentially as described by Apostol *et al.* (1991). In addition to labeled splicing reaction products, P1 digests contained 20 mM ammonium acetate (pH 5.0), 30 µg carrier RNA and 1 U P1 nuclease in 5 µl. Samples were incubated 1 h at 37°C, then spotted on polyethyleneimine (PEI) cellulose thin layer plates and developed with 1 M LiCl. Digestion products were visualized by UV-shadowing and autoradiography prior to recovery from the PEI-cellulose by elution with triethylamine, pH 8.0 (Volckaert and Fiers, 1976). SVP reactions (10 µl) containing radiolabeled P1 products, 37 µg carrier RNA and 2 µg snake venom phosphodiesterase in 25 mM Tris-HCl (pH 9), 10 mM $MgCl_2$, were incubated overnight at 37°C. Samples were chromatographed on PEI-cellulose in 1M LiCl and visualized as above.

Nucleotide triphosphate binding experiments----All procedures were carried out on ice at 4°C. BSA was used as a control for nonspecific binding of NTP to protein. A sample lacking protein was included to measure nonspecific nucleotide retention by filter units. I buffer (50 mM potassium phosphate (pH 8.0), 10 mM KCl, 10 mM $MgCl_2$, 0.5 mM β-mercaptoethanol, 3.5% glycerol) was used to dilute radiolabeled and unlabeled NTPs. F Buffer was used to dilute DHFR and DAKC protein fractions. [α-^{32}P]GTP binding assays contained 0.58 µM protein and 5 µCi [α-^{32}P]GTP (33 nM GTP). [α-^{32}P]ATP, -CTP, and -UTP binding assays contained 2.94-8.36 µM protein and 10 µCi [α-^{32}P]NTP (66 nM NTP). Protein and radioactive nucleotide concentrations were chosen to yield reproducible binding measurements at least ten-fold over background BSA/NTP binding. Unlabeled NTP was added in increasing concentration up to 10 mM. Binding assays were executed as follows. Sequentially, 10 µl DAKC, DHFR, or BSA were mixed with 30 µl I

Buffer. [α-^{32}P]NTP alone or with unlabeled NTP was mixed with I buffer in a volume of 10 µl. The protein-I buffer dilution was added to the NTP tube, and rapidly mixed by pipetting up and down. Immediately, the mixture was transferred to prechilled Ultrafree-MC filter units and centrifuged in an Eppendorf microfuge at 15,600 x *g*. 30,000 NMWL filters were centrifuged 15 min for NTP-ligase binding experiments; 10,000 NMWL filters used for NTP-DHFR binding assays required 40 min centrifugation. Filter units and filtrates were assayed for Cerenkov cpm. Binding is expressed as the dpm bound minus background with background defined as dpm retained by the filter in the presence of BSA.

RESULTS

The ligation of tRNA halves generated by tRNA endonuclease cleavage includes two steps requiring the expenditure of high energy equivalents. The 3' half must be sequentially phosphorylated and then activated by the addition of AMP through an adenylylated enzyme intermediate. Previous work demonstrated that the yeast tRNA ligase fusion polypeptide, DAKC, is able to kinase 3' tRNA halves using GTP as cofactor (Westaway *et al.*, accompanying manuscript). GTP was, in fact, the preferred phosphate donor over ATP, CTP, and UTP by a ratio of 100-fold or greater in experiments measuring DAKC kinase activity on a synthetic oligonucleotide substrate. However, the DAKC polypeptide was unable to complete the tRNA joining reaction in the presence of GTP alone. This finding suggests that GTP cannot be used for activation of the phosphorylated 3'-half as this is the only other ligase activity requisite to joining requiring a high energy cofactor.

Adenylylation of tRNA ligase has been demonstrated (Greer *et al.*, 1983; Xu *et al.*, 1990) and AMP transfer to the kinased 3'-half tRNA molecule has been postulated to activate the 3'-half prior to joining in reactions dependent solely on ATP (Greer *et al.*, 1983; Phizicky *et al.*, 1986). To determine whether ATP could be used to complete the joining of halves kinased by GTP, splicing reactions were carried out in the presence of [γ-^{32}P]GTP

in concert with unlabeled ATP. Radiolabeled reaction products were then analyzed to determine the position of ^{32}P-incorporation. The results are shown in Fig. 1 and suggest that ATP is, indeed, necessary and sufficient to complete the joining of halves phosphorylated with the γ-phosphate of GTP. Fig. 1*A* shows an autoradiogram of splicing products prepared and separated by polyacrylamide gel electrophoresis as described in *Experimental Procedures*. The first three lanes, labeled ^{32}P-tRNA, show control reactions in which precursor of high specific activity was included and splicing carried out in the presence or absence of DAKC and nonradioactive GTP and ATP as indicated. Note that as previously observed (Westaway *et al.*, accompanying manuscript), no spliced tRNA is formed with DAKC and GTP alone. Lanes 5 and 6 contain only precursor of low specific activity spliced in the presence of [γ-^{32}P]GTP with or without unlabeled ATP. In the absence of ATP, only RNA species in the halves area of the gel become radiolabeled while, in the presence of ATP, ^{32}P-labeled species are also observed in the position of joined product. That the source of labeled phosphate is the [γ-^{32}P]GTP is demonstrated by comparison of lanes 5 and 6 with lane 4 containing pre-tRNA of low specific activity spliced with unlabeled GTP and ATP, without labeled NTP.

If ATP were used to complete joining of 3'-halves phosphorylated by DAKC using ^{32}P-GTP, the splice junction of that tRNA would contain a ^{32}P-labeled 3',5' splice junction phosphate and an unlabeled 2'-phosphate (originating from the 5' splice site). As described in Fig. 1*B*, this splice junction structure is resistant to P1 nuclease and can be isolated as a dinucleotide. The dinucleotide can then be analyzed for labeled phosphate position by SVP digestion. The [γ-^{32}P]GTP-labeled halves and tRNA bands shown in Fig. 1*A* were thus eluted and characterized by sequential digestion with P1 nuclease and SVP as described in *Experimental Procedures*. The results are given in Fig. 1*C* and *D*. P1 nuclease digestion of tRNA halves resulted in labeled 5'-AMP as expected if the γ-^{32}P of GTP were used to phosphorylate the 5'-OH of the 3'-half tRNA. P1 nuclease digestion of RNA eluted from the spliced tRNA area of the polyacrylamide gel (1*A*) yielded a labeled

species migrating close to the origin on TLC plates as expected of a P1-resistant dinucleotide. SVP digestion of this dinucleotide yielded labeled 5'-AMP thereby confirming that not only was the γ-phosphate of GTP used by DAKC to kinase the 5'-OH position of the 3'-half tRNA; but that in the presence of ATP, the ligase could complete joining of such kinased halves in the normal fashion. These results suggest that ATP was used to adenylylate the ligase which then activated the kinased 3'-half and resulted in joined product with a splice junction 2'-phosphomonoester, 3'-5'-(^{32}P)-phosphodiester bond. The important implication here is that multiple cofactor usage is permitted by the ligase. The pursuant question is whether multiple cofactor use is preferred.

To examine the efficacy of multiple cofactor usage, joining reactions were carried out with pair-wise combinations of nucleotides. In the presence of a limiting concentration of either ATP or GTP, an increasing concentration of ATP, CTP, GTP or UTP was added to the joining assays. CTP and UTP were included in this titration series for two reasons. First, UTP and CTP were each previously shown to support kinase activity with the DAKC ligase. Secondly, UTP, like GTP, was demonstrated to allow the complete joining reaction with the deletion polypeptide DAC (although UTP use was not as efficient as GTP by approximately 250-fold; Westaway *et al.*, accompanying manuscript). Therefore it was of interest to investigate the potential for stimulation of single nucleotide joining by CTP or UTP as well as by ATP and GTP.

The results of joining assays performed with two nucleotides are given in Figure 2. 2*A* shows an autoradiograph of pre-tRNA halves joined by DAKC in the presence of limiting amounts of ATP (10 μM) and an increasing concentration of GTP as indicated. As previously determined (Westaway *et al.*, accompanying manuscript), DAKC ligase joined halves with ATP as the sole cofactor; half maximal joining activity occurred at 0.9 mM ATP (K_{app}^{ATP}). Thus with limiting ATP (10 μM), a limited amount of joining (21%; Fig. 2*A*) was observed. The joining yield at this limiting ATP concentration was augmented with the addition of GTP over the range of 0.01 μM to 500 μM. At higher concentrations

of GTP, the joining yield fell away (data not shown). This decline, which was observed with high concentrations of each of the four NTPs, may simply reflect an inhibition of the joining reaction due to increasing ionic strength or chelation of Mg^{2+} by nucleoside triphosphates and will not be discussed further. Experiments similar to that shown in Fig. 2*A* were performed for each nucleotide pair. Figure 2*B* is a graph of data for joining in the presence of limiting ATP concentrations with increasing concentrations of GTP added. Fig. 2*C* shows data for joining in the presence of limiting GTP concentration in conjunction with increasing ATP concentrations, while Fig. 2*D* summarizes joining data obtained in the presence of a limiting level of either ATP or GTP together with increasing concentrations of CTP or UTP. No joining was supported by GTP, UTP, or CTP alone (data not shown; also Westaway *et al.*, accompanying manuscript), or GTP in concert with UTP or CTP (Fig. 2*C*). GTP-dependent joining was enhanced only by the addition of ATP (Fig. 2*C*). Conversely, increasing concentrations of GTP, UTP and CTP stimulated ATP-dependent joining (Fig. 2*B* and *D*).

In Fig. 2*B* and *C*, joining curves are shown for two concentrations of DAKC ligase, 0.02 nM and 0.08 nM. The effect of the quadrupled enzyme concentration was to raise the maximum percent of substrate joined while the NTP concentration at which half maximal stimulation of joining occurs (K_{app}) remained unchanged. This is the result expected if reaction rates, rather than extents, are being measured as expected in this assay. K_{app} values were determined as the average K_{app} derived from 2-4 experiments and are summarized in Table 1. The measured K_{app}^{GTP} (derived from GTP stimulation of ATP joining) was 0.8 μM GTP. K_{app}^{ATP} (derived from ATP stimulation of GTP joining) was 2 μM ATP. CTP and UTP stimulation of ATP joining required significantly higher concentrations of second nucleotide to enhance the reaction. K_{app} measured for CTP was 0.2 mM , while K_{app} for UTP was 0.1 mM.

Joining data with authentic yeast ligase is included for comparison with DAKC ligase activity in Fig. 2*B* and *C*. Reaction extents vary, probably reflecting variations in the

concentration of active enzyme as observed with DAKC. However, the K_{app} values for ATP and GTP with authentic yeast ligase, 4 μM ATP and 0.4 μM GTP, closely match the values obtained with the DAKC ligase (2 μM and 0.8 μM, respectively). Since joining is the composite of its individual enzymatic activities, these similar K_{app} values suggest that activity measurements with DAKC provide a valid picture of the authentic yeast ligase mechanism and further support the previous observation (Apostol *et al.*, 1991) that the DHFR segment does not substantially interfere with the measured ligase properties of the fusion polypeptide.

The interaction of multiple cofactors with ligase might reflect a differential ability either to bind each cofactor or, once bound, to use a given cofactor in the kinase reaction or in the activation of the 3'-half tRNA, or a combination thereof. Furthermore, nucleotide binding might involve interaction at more than one binding site, each with unique NTP-specific affinity. Therefore, nucleotide binding to DAKC ligase was examined in order to determine the multiplicity and specificity of binding sites on the enzyme. Increasing amounts of unlabeled NTP were mixed with constant amounts of DAKC ligase and [α-^{32}P]NTP such that the molar concentration of labeled NTP was significantly below the concentration of DAKC (33-67 nM NTP versus 0.53-2.65 μM DAKC ligase). NTP binding was measured as the retention of radioactivity by 30,000 molecular weight cutoff filters in the presence of DAKC ligase as described in *Experimental Procedures*. BSA was used as a control for nonspecific NTP/protein binding and showed no significant increase in filter retention of ^{32}P-NTP over the no protein control. In order to assure that nucleotide binding measurements were dependent on the ligase moiety of the fusion protein, DHFR was also examined for nucleotide binding and showed no increase over the no protein or BSA controls. At the DAKC protein concentrations practically available for the binding assays, no significant binding of CTP or UTP was detected. Binding of labeled ATP and GTP was readily observed and could be reduced by the addition of unlabeled nucleotide to the reaction.

Binding reduction curves are shown Fig. 3; 3*A* presents the results with [α-^{32}P]ATP, 3*B* with [α-^{32}P]GTP. The unlabeled NTP concentration at which labeled NTP binding is reduced 50% is a measure of the affinity of the unlabeled NTP for the binding pocket of the labeled NTP or, conversely, a measure of the dissociation of the labeled NTP. K_D is the dissociation constant used to describe the self reaction, e.g., ^{32}P-ATP binding reduced by the addition of unlabeled ATP. $K_D{}^C$ is the constant used for the heterologous reaction, e.g., ^{32}P-ATP with unlabeled GTP. In the circumstance of these experiments wherein the concentration of labeled nucleotide is very much less than its K_D for DAKC ligase, $K_D{}^C$ is equivalent to an IC_{50} value (inhibitor concentration reducing the amount of bound label by 50%). Thus the 50% reduction in labeled GTP binding by unlabeled GTP at a concentration of 0.1 μM indicates a K_D of 0.1 μM, while the 50% reduction by unlabeled ATP at a concentration of 1 mM suggests a $K_D{}^C$ (IC_{50}) of 1 mM ATP for the GTP binding site. Conversely, the 50% reduction in labeled ATP binding requires 4 μM ATP or 5 mM GTP; a K_D of 4 μM and a $K_D{}^C$ of 5 mM, respectively, for the ATP site. Binding K_D and $K_D{}^C$ values were determined as the average from 2-6 experiments and are summarized in Table 1.

It might be argued that NTP binding by ligase would differ in the presence of tRNA substrate. The conformation of the ligase polypeptide would necessarily be restricted by binding to tRNA since it has been suggested that, although a ligase molecule contains only one high affinity tRNA binding site (Apostol and Greer, 1991), widely separated areas of the ligase linear sequence participate in that substrate binding (Apostol *et al.*, 1991). Thus when ligase is structurally confined in its substrate bound form, access to nucleotide binding sites might be specifically enhanced or restricted. Binding experiments were therefore carried out in the presence of tRNA. Binding assays were performed as described in *Experimental Procedures* with the addition of yeast $tRNA^{Phe}$ (Sigma) to the DAKC-binding buffer mix in molar concentration equal to DAKC. (At this concentration the tRNA is approximately 500- to 2500-fold above its measured affinity for yeast ligase;

Apostol and Greer, 1991). No significant changes in K_D or $K_D{}^C$ values were observed for GTP or ATP binding in the presence of $tRNA^{Phe}$ (data not shown) suggesting that NTP binding sites are either insensitive to conformational change induced by substrate binding or that no conformational change is induced. This finding is consistent with expectations based on the similarity of K_D values, determined in the absence of tRNA, with K_{app} values measured for ATP and GTP in the 2-nucleotide joining assays.

DISCUSSION

The data presented here suggest that yeast tRNA ligase joins tRNA halves most efficiently in the presence of multiple nucleotide cofactors. The results are summarized in Table 1 together with K_{app} values derived from single nucleotide joining assays and kinase assays presented previously (Westaway *et al.*, accompanying manuscript). The results of the two nucleotide joining assays suggest that while ATP is the only cofactor absolutely necessary for joining with tRNA ligase, multiple cofactor usage results in more efficient joining with respect to cofactor concentrations required. The best combination is ATP plus GTP. Joining with ATP as the sole cofactor requires ATP concentrations approximately 500-fold higher than when joining occurs in the presence of micromolar concentrations of GTP. That is, the K_{app} for ATP as the only joining cofactor is 0.9 mM, while the K_{app} values for ATP and GTP in mixed cofactor reactions are 2 μM and 0.8 μM, respectively. CTP and UTP also stimulate ATP-dependent joining though less dramatically, with K_{app} values 100- to 200-fold greater than for GTP. Earlier work suggesting that ligase could join at least inefficiently using GTP, CTP or UTP alone, recognized interpretation of that data was limited by use of a partially purified ligase extract as well as by use of nucleotides of qualified purity (Greer *et al.*, 1983). Given the micromolar amounts of GTP and ATP contaminants which would be required to stimulate joining, the results of this previous work are not inconsistent with our present observations.

In order to participate in the varied ligase activities, the cofactor must bind to the enzyme or substrate. Comparison of K_D with $K_D{}^C$ values indicates that GTP binds strongly and specifically to a GTP binding site. Likewise ATP binds tightly and specifically to an ATP binding site. Moreover UTP and CTP are equally poor at reducing specific nucleotide binding, requiring millimolar concentrations. This would suggest that ATP and GTP binding sites are not only preferential with respect to nucleotide use, but are operationally exclusive given the thousand-fold (or greater) difference in K_D and $K_D{}^C$ values. It is difficult to interpret what effect these NTP specificities may have *in vivo*. Only limited information on NTP concentration in either the whole cell or the nucleus is available. Compartmentation within the nucleus could also affect nucleotide pools available to the ligase. It is, however, interesting to note that relative estimates of nucleotide levels place the cellular GTP concentration approximately four to eight-fold below the ATP concentration (Burridge *et al.*, 1977). This order of magnitude difference parallels the order of magnitude difference observed in K_D values for ATP and GTP binding to ligase.

K_D values for ATP and GTP binding match the K_{app} values determined for ATP and GTP stimulation of joining in the presence of both nucleotides. Also as previously noted, the K_D for GTP binding matches well with the K_{app} for GTP in the kinase assay (Table 1). These similar values suggest that nucleotide binding to DAKC might be a limiting step for joining under these reaction conditions. The comparison of K_D and $K_D{}^C$ values for ATP or for GTP also suggests indirectly that GTP and ATP binding are independent events. That is, the presence of the second nucleotide does not influence the binding characteristics of the first as measured by the K_{app} for that nucleotide in joining stimulation assay. If binding of one nucleotide facilitated (or inhibited) binding of the second nucleotide, K_{app} values derived from 2-nucleotide joining assays should be lower (higher) than the corresponding K_D value. Since the K_D and K_{app} values are approximately equal within the error of the assays, independent binding events are implied.

The similarity of the ATP K_{app} determined for single nucleotide joining, 0.9 mM, and the $K_D{}^C$ of ATP for the GTP binding site, 1 mM, suggests that when ATP is used for the kinase activity, it may be bound at the GTP site rather than at the ATP binding site. This is in contrast to the mechanism of action suggested for GTP use by the DAKC deletion derivative DAC (Westaway *et al.*, accompanying manuscript). DAC can complete joining using only GTP as cofactor with an observed K_{app} for GTP of 0.4 μM. The similarity of this K_{app} with both the DAKC K_D for GTP binding (0.15 μM) and K_{app} for GTP-dependent kinase activity (0.25 μM) is consistent with the idea that GTP binds to the GTP-specific site of DAC and then is used by both the kinase and adenylylate synthetase activities to complete the joining of tRNA halves. The comparison of DAKC and DAC results also suggests that the K domain (amino acids 396-595) not only influences tRNA substrate recognition (Westaway *et al.*, accompanying manuscript), but also may mediate the transfer of bound cofactor to the active sites for the adenylylate synthetase and/or kinase activities.

The fact that the K_{app} values for GTP, CTP, and UTP stimulation of ATP-dependent joining closely parallel the K_{app} measurements for kinase activity, suggests that the joining stimulation reflects the ligase enzyme ability to use each of the four nucleotides in the kinase step; GTP>>UTP>CTP>ATP. GTP, CTP, and UTP do not allow joining in the absence of ATP despite the demonstrated ability of these nucleotides to bind the ATP site at high concentrations. These results are consistent with the hypothesis that only ATP can successfully complete the adenylylate synthetase reaction (Westaway *et al.*, accompanying manuscript) and further suggest that cofactor discrimination by the adenylylate synthetase activity occurs at a step subsequent to NTP binding at the ATP site.

Comparison of the UTP and CTP binding and displacement data with UTP and CTP kinase and two-nucleotide joining data suggests that DAKC ligase may have a third nucleotide binding site which can interact with the kinase active site. The K_{app} values for the stimulation of ATP-dependent joining reasonably match the kinase K_{app} values which

might suggest that the UTP and CTP cofactors enhance ATP-joining by serving as kinase activity cofactors. However, the results of binding displacement assays suggest that UTP and CTP do not bind to the GTP (or ATP) binding site except with extremely low affinity. Thus, in order to reconcile binding displacement data with joining stimulation and kinase data, it is reasonable to invoke a minimum of three distinct NTP binding sites for the DAKC ligase molecule. More than three sites are possible, but further experiments must be done to assay UTP and CTP binding to DAKC directly. Based on existing data for nucleotide usage it is predicted that 20-60 μM protein would be required to measure UTP or CTP binding to DAKC ligase reliably, while the highest molarity obtained for use in the current work was 8.3 μM.

In summary, Figure 4 presents a model for the interaction of cofactors with the nucleotide binding and enzyme active sites in yeast tRNA ligase. ATP binds and is used in the adenylylate synthetase reaction. GTP or ATP may bind at the GTP binding site, or UTP or CTP binds at a third NTP site, and subsequently are channeled to the kinase active site. GTP use is preferred in the kinase reaction The use of multiple cofactors by ligase may have important implications for integration of events in tRNA biosynthesis. *In vivo*, this requirement may effectively couple pre-tRNA splicing to the nucleotide requirements for transcription and/or translation increasing sensitivity and response to changes in intracellular environment. To assess the degree of interdependence among processes will require the identification of the ligase site of action within the nucleus as well as the ability to measure local levels of nucleotide concentrations.

ACKNOWLEDGMENTS

We thank Eric M. Phizicky and Craig L. Peebles for many helpful discussions, and Reinhard Rauhut for the generous gift of highly purified tRNA endonuclease.

REFERENCES

Abelson, J. (1991) *Harvey Lect.* 1989-90 **85**, 1-42

Apostol, B.L. and Greer, C.L. (1991) *Nucleic Acids Res.* **19**, 1853-1860

Apostol, B.L., Westaway, S.C., Abelson, J., and Greer, C.L. (1991) *J. Biol. Chem.* **266**, 7445-74

Burridge, P.W., Woods, R.A., and Henderson, J.F. (1977) *Can. J. Biochem.* **55**, 935-941

Greer, C.L., Peebles, C.L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 537-546

McCraith, S.M. and Phizicky, E.M. (1991) *J. Biol. Chem.* **266**, 11986-11992

Normanly, J., Ogden, R.C., Horvath, S.J. and Abelson, J. (1986) *Nature* **321**, 213-219

Peebles, C.L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32**, 525-536

Perlman, P.S., Peebles, C.L., and Daniels, C. (1990) in *Intervening Sequences in Evolution and Development* (Stone, E.M. and Swartz, R.J., eds) pp. 112-161, Oxford University Press, Oxford

Phizicky, E.M., Consaul, S.A., Nehrke, K.W., and Abelson, J. (1992) *J. Biol. Chem.* **267**, 4577-4582

Phizicky, E.M., Schwartz, R.C., and Abelson, J. (1986) *J. Biol. Chem.* **261**, 2978-2986

Rauhut, R., Green, P.R., and Abelson, J. (1990) *J. Biol. Chem.* **265**, 18180-18184

Reyes, V.M., and Abelson, J. (1987) *Anal. Biochem.* **166**, 90-106

Ruby, S.W., Goelz, S.E., Hostomsky, Z., and Abelson, J.N. (1990) *Methods Enzymol.***181**, 97-121

Volckaert, G., and Fiers, W. (1976) *Anal. Biochem.* **72**, 433-446

Westaway, S.K., Phizicky, E.M., and Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171-3176

Westaway, S.K., Belford, H.G., Abelson, J., Apostol, B.L., and Greer, C.L. (1992) J. Biol. Chem., accompanying manuscript

Xu, Q., Teplow, D., Lee, T.D., and Abelson, J. (1990) *Biochem.* **29**, 6132-6138

FIGURE LEGENDS

FIG. 1. **Labeling the splice junction phosphodiester bond with [γ-^{32}P]GTP.** *A*. Biotinylated pre-tRNA^{Phe} (620 fmol, 9.6 dpm/fmol) was cut with tRNA endonuclease, subsequently incubated with a combination of the DAKC ligase, GTP and ATP. Products were resolved by polyacrylamide gel electrophoresis and visualized by autoradiography as described in *Experimental Procedures*. From the *left*, lanes 1-3 also contained 6.2 fmol pre-tRNA of high specific activity (4.8 x 10^3 dpm/fmol). The mobilities of pre-tRNA substrate, halves produced by endonuclease cleavage and spliced tRNA product are indicated at left edge. Lane 4 contained only low specific activity substrate incubated with DAKC, GTP and ATP. The final lanes contained low specific activity substrate and a combination of DAKC, ATP and [γ-^{32}P]GTP (5000 Ci/mmol; 100 μCi/assay, equivalent to 2 μM GTP in the assay) as specified. With the exception of reactions containing pre-tRNA of high specific activity, splicing products were purified prior to electrophoresis by binding to avidin affinity resin as described in *Experimental Procedures. B.* Schematic of the tRNA^{Phe} splice junction structure and nuclease digestion products. Shown are the 2'-phosphate and 3',5'-phosphodiester bond and the digestion products expected with P1 nuclease and SVP. The asterisk indicates ^{32}P-label position. *C.* P1 nuclease analysis of ^{32}P-labeled *tRNA* (*A*, lane 6) and *halves* (*A*, lanes 5 and 6). *D.* SVP analysis of the tRNA P1 digestion product (*C*, *tRNA*). RNAs were eluted, digested, chromatographed, and visualized by autoradiography and UV-shadowing as described in *Experimental Procedures*. Dotted lines show the position of unlabeled nucleotide monophosphate and diphosphate digestion products. The position of the labeled junction is also indicated.

FIG. 2. **Effects of combined cofactor use in tRNA joining**. *A*. Autoradiograph of a joining titration. Each joining assay contained 6.2 fmol pre-tRNA, 10 μM ATP, 0.4 μM DAKC ligase, and 0-500 μM GTP as described in *Experimental Procedures* and indicated

in the figure. The position of pre-tRNA, tRNA halves and mature tRNA product are indicated on the left. *B.* GTP stimulation of ATP-dependent joining. Joining assays like that shown in *A* were quantitated by Cerenkov cpm in gel slices. *% Joined* was determined as [CPM tRNA/(CPM tRNA + CPM halves)] x 100; joining due to the presence of ATP alone has not been subtracted. Reactions contained GTP as specified, and 10 μM ATP and 0.02 nM DAKC (*squares*), 2 μM ATP and 0.08 nM DAKC (*circles*), or 10 μM ATP and 1.9 nM yeast ligase (*triangles*). *C.* ATP stimulation of GTP-dependent joining. Reactions contained ATP as specified, 1 μM GTP and 0.02 nM DAKC (*squares*), 0.08 nM DAKC (*circles*), or 1.9 nM yeast ligase (*triangles*). *D.* CTP and UTP stimulation of ATP- or GTP-dependent joining. Titrations contained the following; *filled circles*, varied CTP concentration, 10 μM ATP, and 0.04 nM DAKC; *filled squares*, varied UTP, 10 μM ATP, 0.14 nM DAKC; *open circles*, varied CTP, 1 mM GTP, 0.02 nM DAKC; *open squares*, varied UTP, 1 mM GTP, 0.14 nM DAKC. Background cpm in gel slices have not been subtracted and account for initial apparent low levels of joining in the absence of ATP. Solid lines in panels *B*, *C*, and *D* were generated by nonlinear regression analysis of the data using SigmaPlot™ (Jandel Corp.).

FIG. 3. **Nucleotide triphosphate binding by DAKC ligase.** Binding experiments were performed as described in *Experimental Procedures*. Bound NTP was calculated as dpm retained by filters in the presence of DAKC minus the background dpm retained in the presence of BSA. *Bound* dpm is plotted versus increasing concentration of nonradioactive *NTP* competitor added. Panel *A* shows the reduction of bound [α-^{32}P]ATP by unlabeled ATP (*filled circles*), GTP (*hollow circles*), UTP (*filled squares*), or CTP (*hollow squares*). The inset graph expands the scale for the [α-^{32}P]ATP versus ATP curve in the μM range. Panel *B* shows the reduction of bound [α-^{32}P]GTP by increasing addition of unlabeled GTP (*filled circles*), ATP (*hollow circles*), UTP (*filled squares*) and CTP (*hollow squares*). Inset graph expands the [α-^{32}P]GTP versus GTP curve in the μM range. Solid

lines in panels *A* and *B* represent curves generated by nonlinear regression analysis of the data using SigmaPlot™ (Jandel Corp.).

FIG. 4. **Model for interactions among nucleotide binding and enzyme active sites in tRNA ligase.** *Bold arrows* indicate preferred interactions; *dotted arrows*, interactions prevented by the presence of amino acids 397-595. Symbols are *C*, the cyclic phosphodiesterase active site; *Ad*, the adenylylate synthetase active site; *K*, the kinase active site. Nucleotide binding sites are indicated as boxes.

TABLE I
Summary of activity and binding constants

Activity Assay[a]	ATP	GTP	UTP	CTP
Joining[b]	900 μM	0	0	0
Joining with constant GTP[b]	2μM	N/A[d]	0	0
Joining with constant ATP[b]	N/A[d]	0.8 μM	0.1 mM	0.2 mM
Kinase[b]	200 μM	0.25 μM	25 μM	60 μM
NTP Binding[c], *ATP	4 μM	5 mM	7 mM	3 mM
NTP Binding[c], *GTP	1 mM	0.15 μM	2 mM	6 mM

[a]Ligase enzymatic activity assays were performed as described in *Experimental Procedures* with the exceptions of single nucleotide-dependent joining and kinase assays which were reported elsewhere (Westaway *et al.*, accompanying paper) and summarized here. Data were analyzed by nonlinear regression using SigmaPlot™ (Jandel Corp.).

[b]Results shown for joining and kinase reactions are K_{app} values.

[c]Values shown for NTP binding assays are binding constants (K_D, $K_D{}^C$) determined as the reduction of bound radiolabelled ATP or GTP (indicated by *) by added unlabelled NTP as specified.

[d]N/A, not applicable.

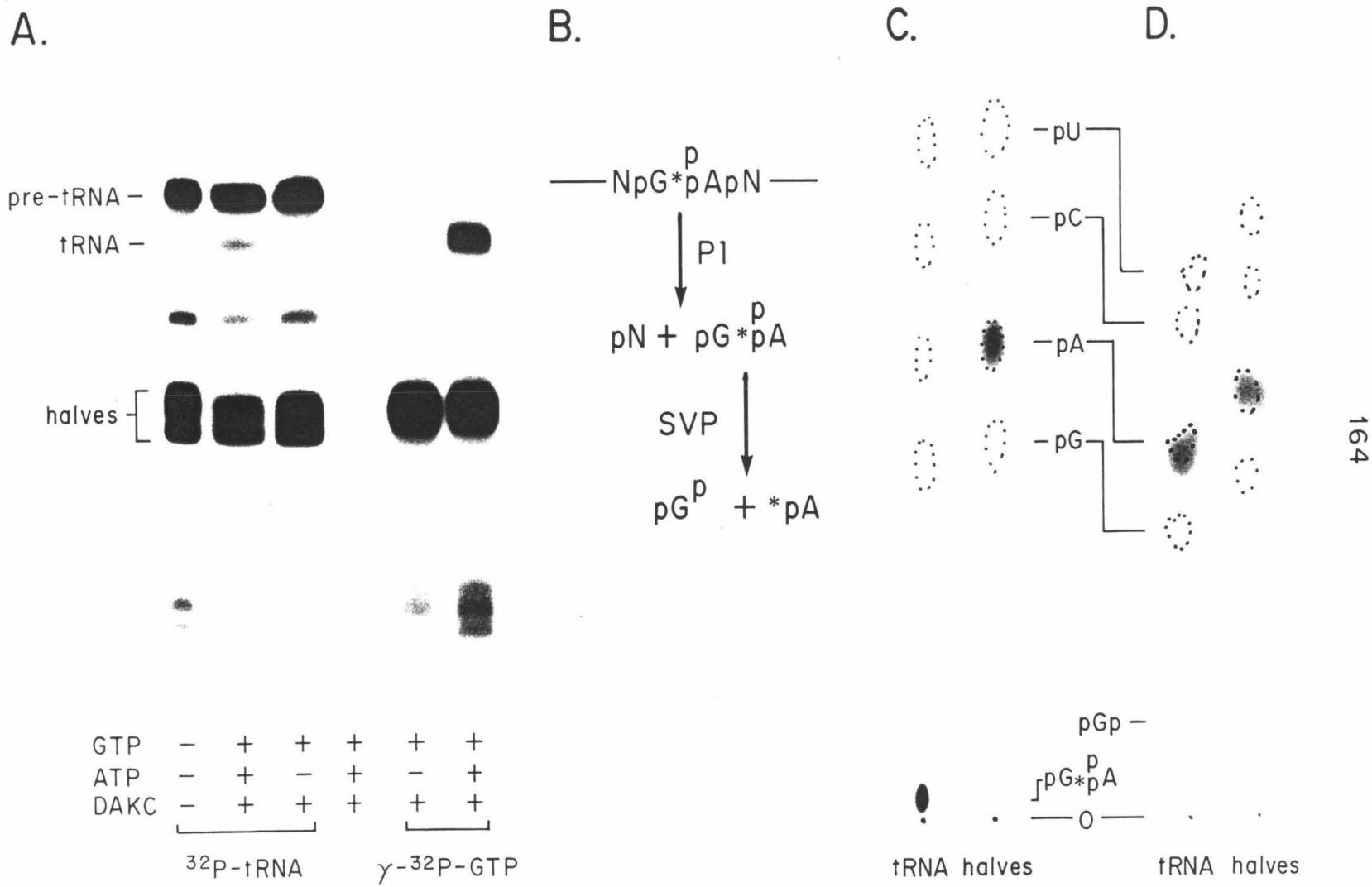

A.
B.
C.
D.
pre-tRNA
tRNA
halves
GTP
ATP
DAKC
32P-tRNA
γ-32P-GTP
NpG*pApN
P1
pN + pG*pA
SVP
pG^p + *pA
pU
pC
pA
pG
pGp
pG*pA
O
tRNA halves
tRNA halves

Figure 1

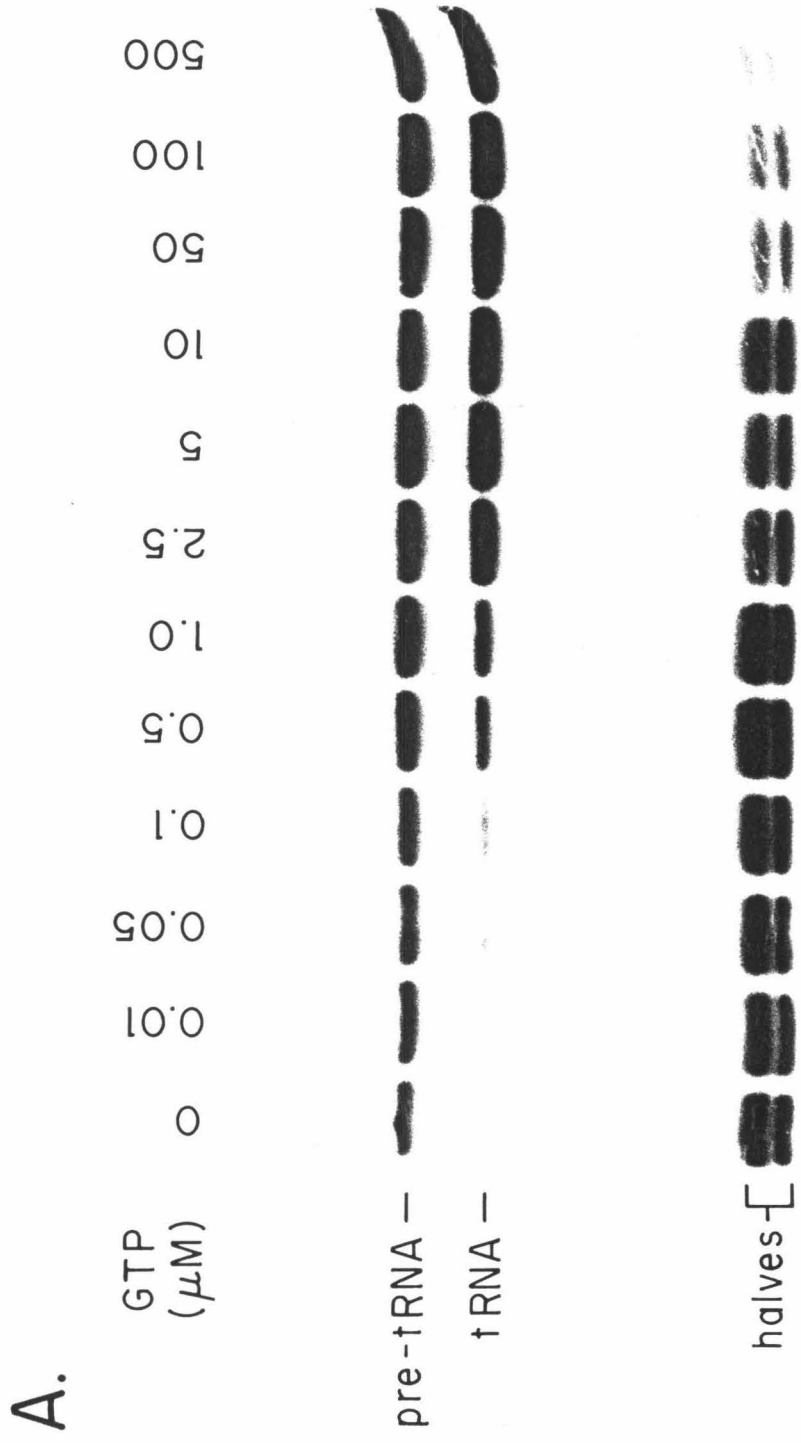


Figure 2A

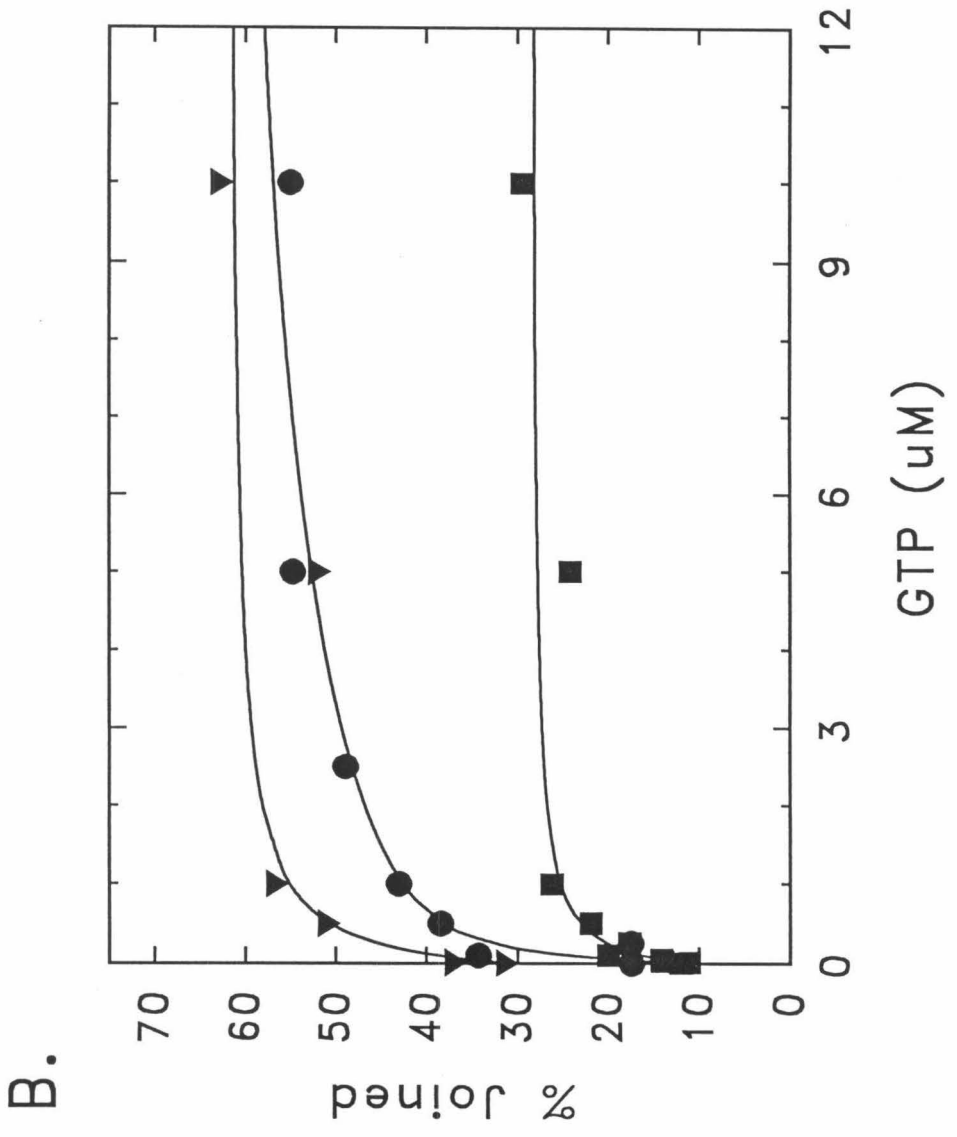


Figure 2B

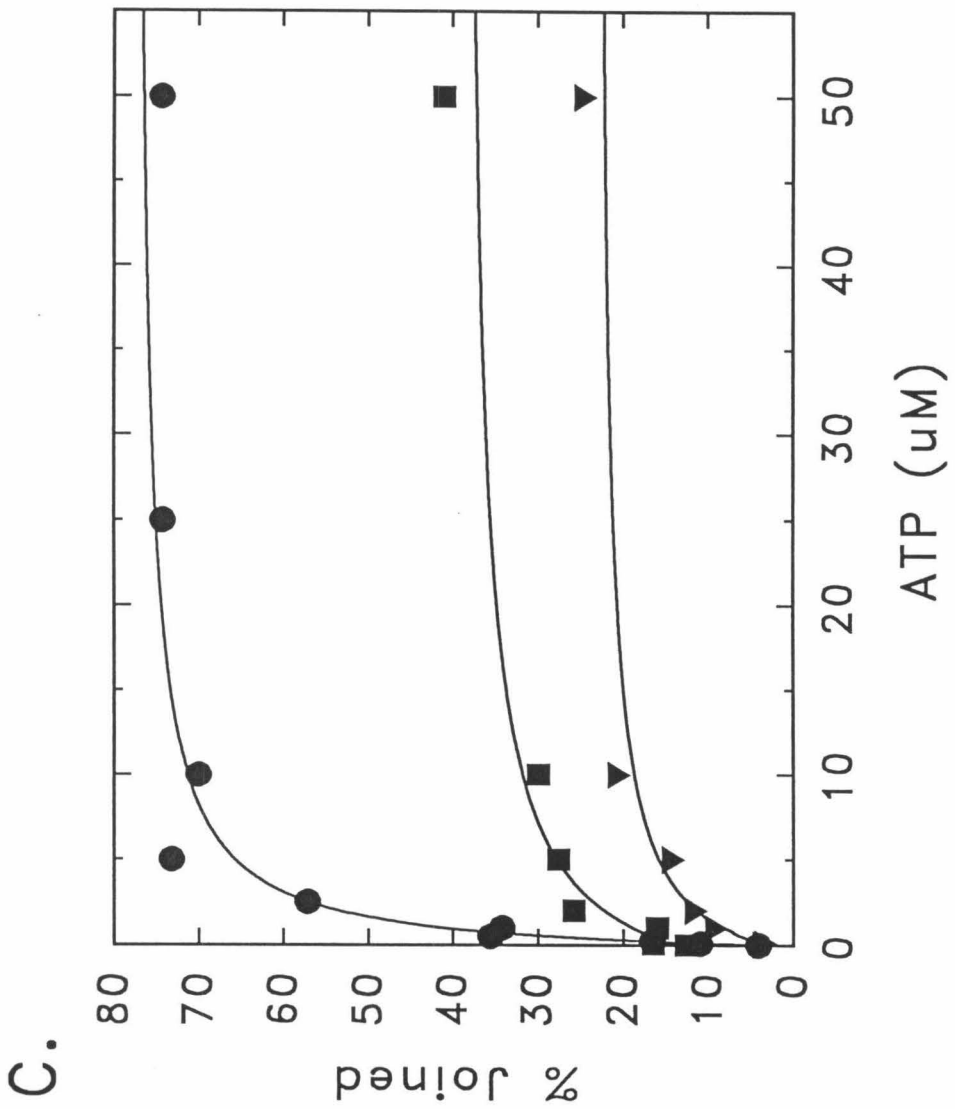


Figure 2C

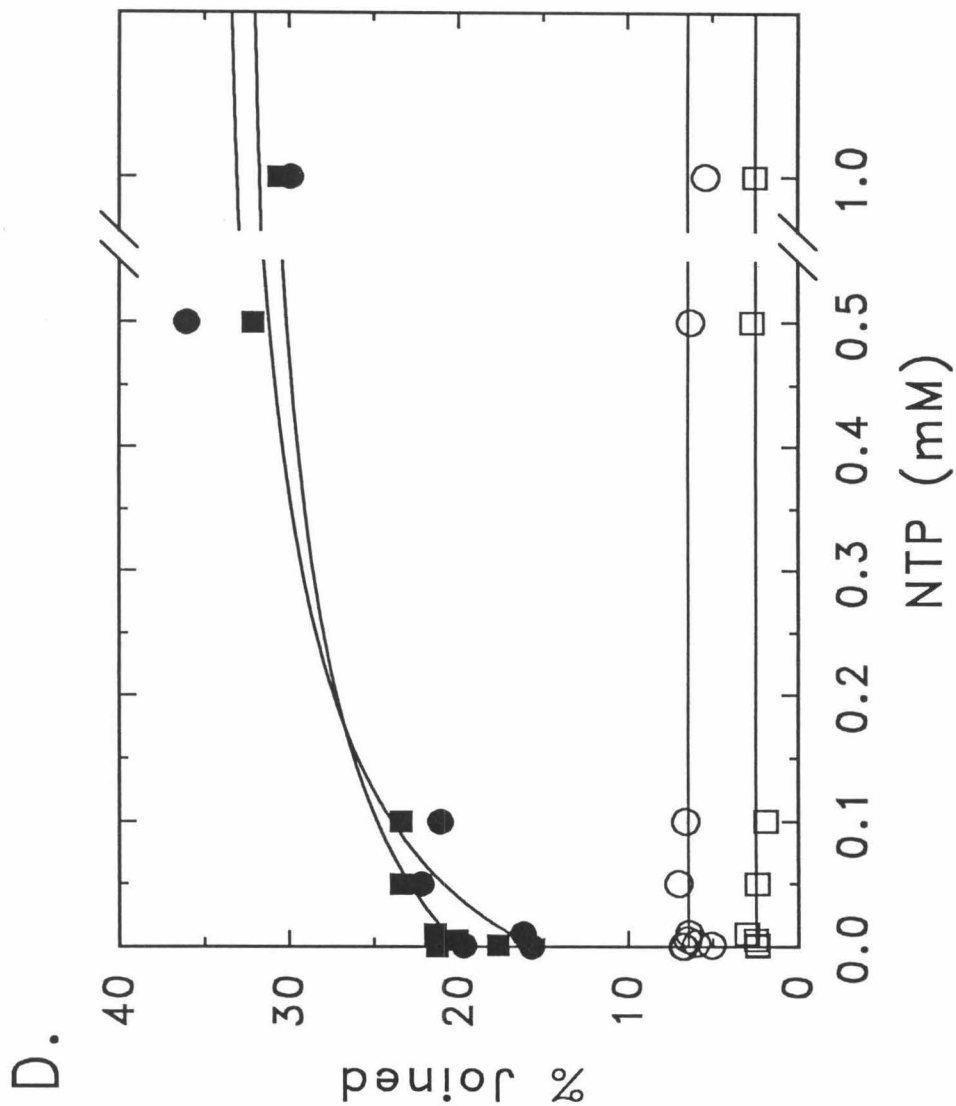


Figure 2D

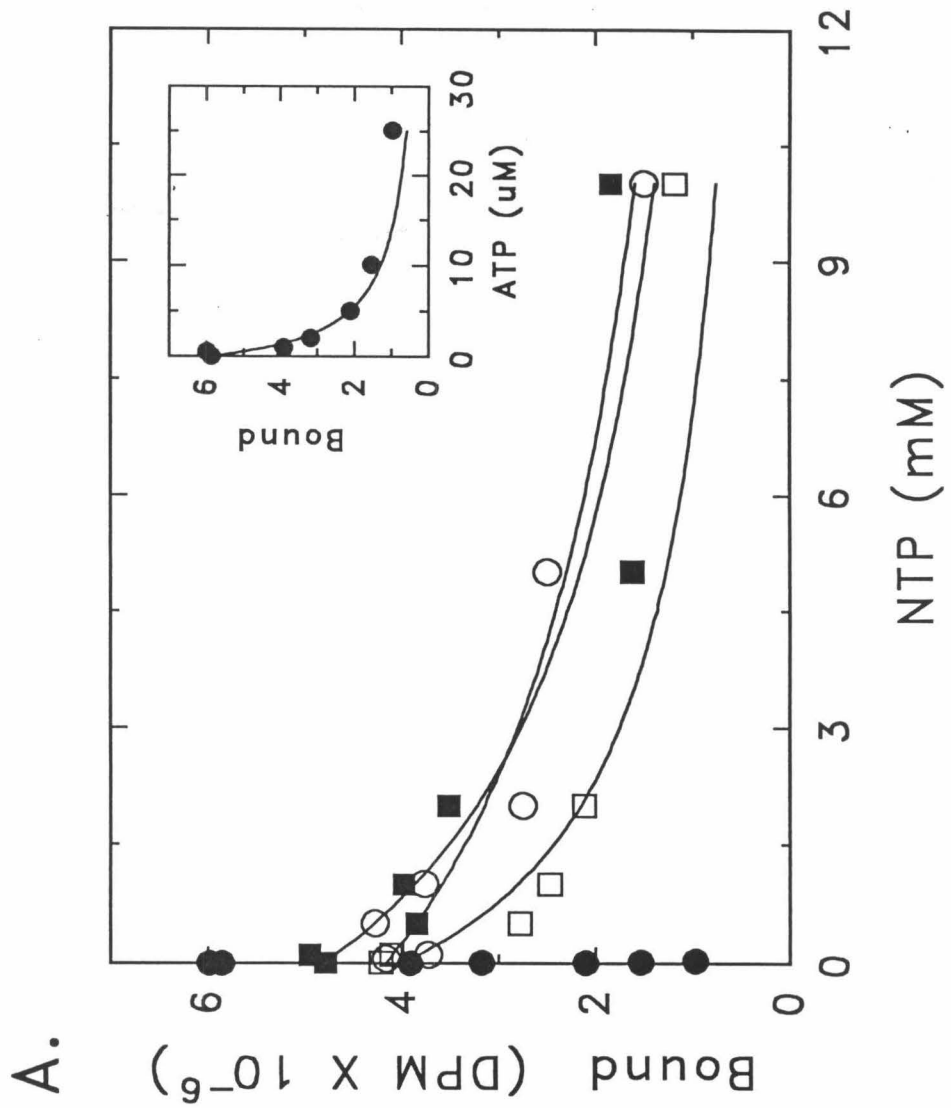


Figure 3A

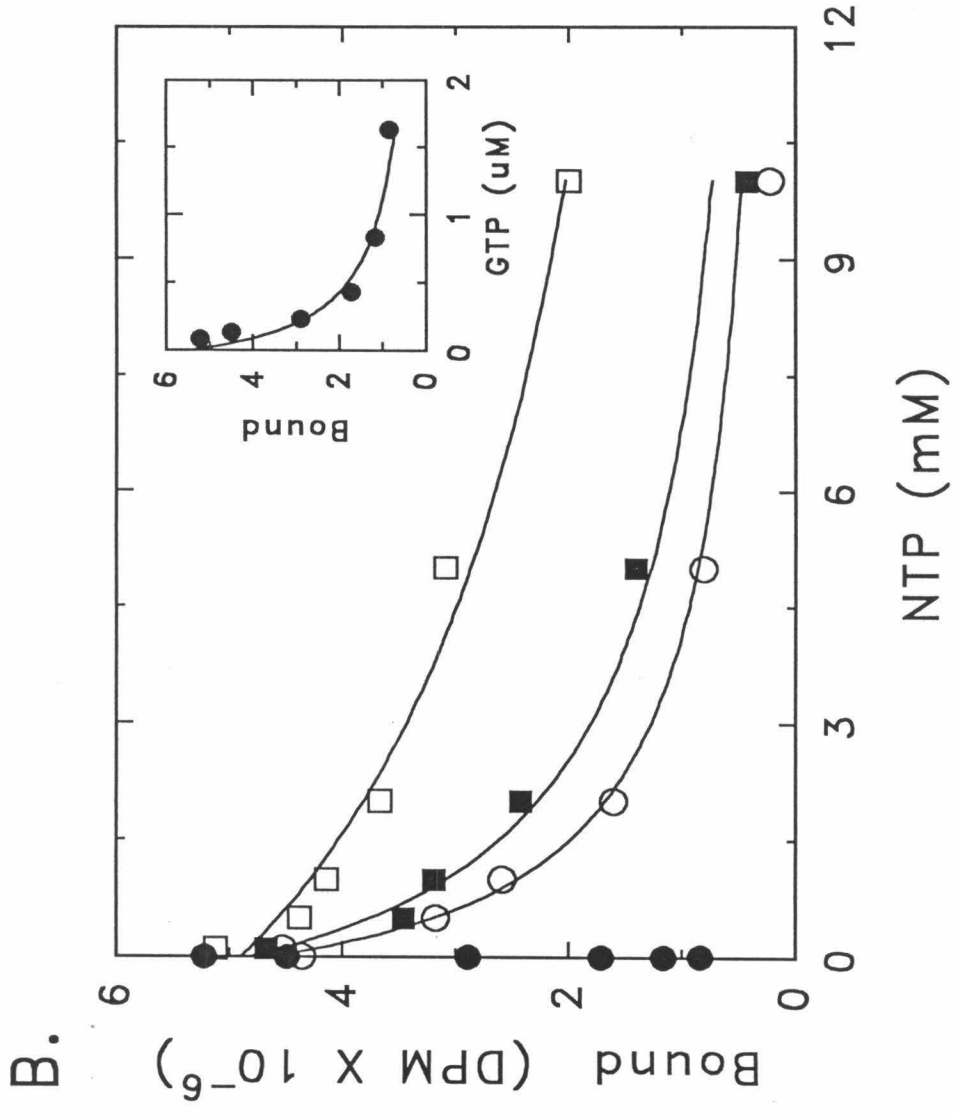


Figure 3B

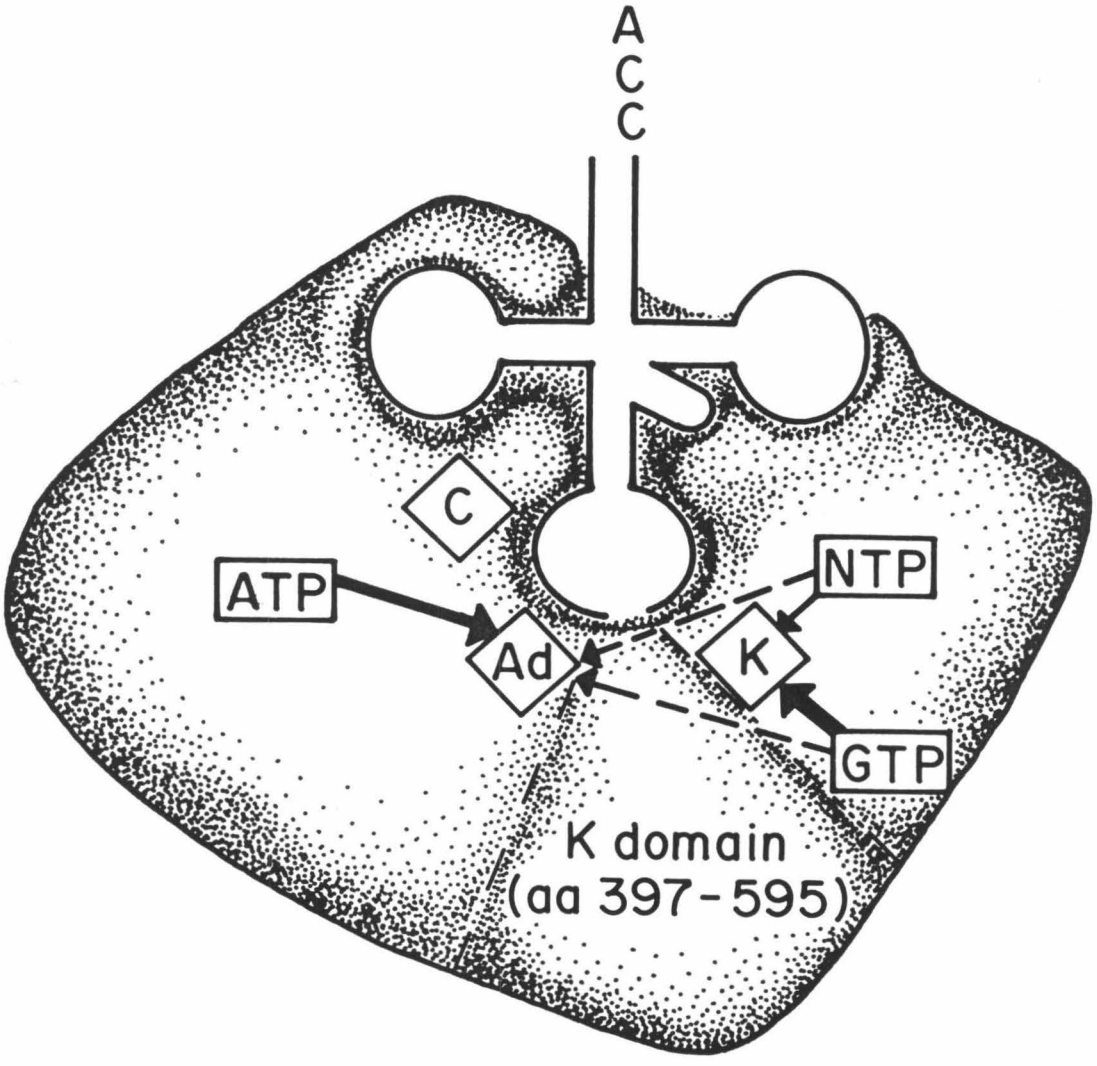


Figure 4