**UCSF**
**UC San Francisco Electronic Theses and Dissertations**

**Title**
Unconventional splicing of the HACI mRNA during the unfolded protein response

**Permalink**
https://escholarship.org/uc/item/3g53p60k

**Author**
Gonzalez, Tania Nadin

**Publication Date**
2003

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

# Unconventional splicing of the HAC1 mRNA during the unfolded protein response

by

Tania Nadin Gonzalez

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Cell Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

*This thesis is dedicated to Ludmila and Raul Gonzalez.*

# Acknowledgements

## THANKS TO ALL THE GREAT FOLKS

## IN THE WALTER LAB!!!

MOLECULAR BIOLOGY
4TH EDITION
OF THE CELL
MBoC4

# Abstract

The unfolded protein response (UPR) is an intracellular signaling pathway that, in response to accumulation of misfolded proteins in the lumen of the endoplasmic reticulum (ER), upregulates transcription of ER resident chaperones. In the yeast *Saccharomyces cerevisiae*, a key step in this pathway is the non-conventional, regulated splicing of the mRNA encoding the positive transcriptional activator Hac1p. When the UPR is off, translation of *HAC1* mRNA is blocked by interactions between the *HAC1* 5'UTR and a 252 nucleotide intron. When the UPR is turned on, this block is relieved by splicing out of the intron and translation resumes. The bifunctional transmembrane kinase/endoribonuclease Ire1p cleaves *HAC1* mRNA at the splice sites that flank the intron and tRNA ligase joins the two exons together. We now know that the salient features of the Ire1p-dependent UPR signaling pathway first discovered in *S. cerevisiae* are conserved from yeast, through worms, and onto mammals. This thesis adds to the continuing story of the conserved Ire1p signaling pathway in yeast. We reconstitute *HAC1* mRNA splicing in an efficient *in vivo* reaction and show that, in many ways, the mechanism of *HAC1* mRNA splicing resembles that of pre-tRNA splicing. Ire1p cleavage produces 2',3'-cyclic phosphates, the excised exons remain associated by base pairing, and exon ligation by tRNA ligase follows the same chemical steps as for pre-tRNA splicing. In contrast to the similarities to tRNA splicing, the structural features of the splice junctions recognized by Ire1p differ from those recognized by tRNA endonuclease. Small stem-loop structures predicted to form at both *HAC1* mRNA splice junctions are required and sufficient for Ire1p cleavage. Previous work in our laboratory led to the unexpected discovery that tRNA ligase is required for *HAC1* mRNA splicing by identifying an allele, *rlg1-100*, that is defective for *HAC1* mRNA splicing yet is functional for pre-tRNA splicing. We present data consistent with a model we propose explaining the basis for the *rlg1-100* ligase defect. *HAC1* mRNA splicing takes place in the cytoplasm. We suggest that unlike wild type tRNA ligase, *rlg1-100* ligase does not localize to the site of cytosolic *HAC1* mRNA splicing in sufficient quantities to support robust *HAC1* mRNA splicing. Finally, in order to learn more about *in vivo HAC1* mRNA splicing, we conducted a genetic screen for suppressors of the *rlg1-100* tRNA ligase defect. Unexpected, we found that mutations affecting mRNA turnover can suppress the inability of *rlg1-100* strains to grow on UPR-inducing media. This suppressor phenotype requires cleavage of the *HAC1* mRNA at the 5' splice site and production of a C-terminally truncated Hac1p from the resultant 5' *HAC1* mRNA exon. Cleavage at the 5' splice site likely relieves the translational inhibition which results from base pairing between the *HAC1* 5' UTR and its intron.

# Table of Contents

## Chapter I

Introduction: The unfolding tale of the unfolded protein response .......................... 1

## Chapter II

Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway .......................................................................................... 27

## Chapter III

Preliminary characterization of the *rlg1-100* tRNA ligase .................................. 42

## Chapter IV

In search of *rlg1-100* suppressors: a genetic screen .......................................... 89

## Appendix A

The sequence of tRNA ligase: speculations on structure and function ..................... 166

## Appendix B

The possible role of the NAD-dependent 2'-phosphotransferase, Tpt1p in the unfolded protein response pathway ................................................................. 197

# List of Tables

## Chapter I

Table I-1. Inducers of the unfolded protein response........................................13

## Chapter III

Table III-1. Chapter 3 plasmid list..........................................................64
Table III-2. Chapter 3 yeast strain list........................................................65

## Chapter IV

Table IV-1. Chapter 4 plasmid list..........................................................118
Table IV-2. Chapter 4 yeast strain list......................................................119
Table IV-3. Linkage analysis for C1-Sor$^+$ strains..........................................120
Table IV-4. Linkage analysis for C2-Sor$^+$ strains..........................................121

## Appendix B

Table B-1. Appendix B plasmid list...........................................................208

# List of Figures

## Chapter I

Figure I-1. The UPR pathway in yeast....................14

Figure I-2. The diversity of the UPR in mammals....................16

## Chapter II

Figure II-1. Secondary structure prediction of the *HAC1*$^u$ mRNA intron and splice site stem-loop structures....................30

Figure II-2. Cleavage of stem-loop RNAs by Ire1p(k+t)....................31

Figure II-3. Mutational analysis of 3' splice site loop....................32

Figure II-4. Formation of 3' terminal guanosine 2',3'-cyclic phosphate after Ire1p(k+t) cleavage of stem-loop RNA....................33

Figure II-5. Characterization of termini produced after Ire1p(k+t) cleavage of stem-loop RNA by sequential enzymatic modification....................34

Figure II-6. Ire1p(k+t) cleaves *HAC1*$^u$ mRNA at the same positions as it cleaves the 5' and 3' stem-loops....................34

Figure II-7. Ire1p(k+t) cleaves *HAC1*$^u$ mRNA more efficiently than stem-loop RNAs...35

Figure II-8. *HAC1*$^u$ mRNA exons remain base paired following Ire1p(k+t) cleavage.....36

Figure II-9. The phosphate group at the splice junction derives from the nucleotide triphosphate....................37

Figure II-10. Model of *HAC1*$^u$ mRNA splicing....................38

## Chapter III

Figure III-1. Northern blot analysis of *HAC1* mRNA splicing in *RLG1* and *rlg1-100* strains........................................................................................66

Figure III-2. Growth of *rlg1-100* strains overexpressing *rlg1-100, RLG1, IRE1*, or *HAC1*........................................................................................68

Figure III-3. Northern blot analysis of *rlg1-100* strains overexpressing *rlg1-100*..........70

Figure III-4. Western blot analysis of tRNA ligase protein in *rlg1-100* and *RLG1* strains........................................................................................72

Figure III-5. *In vitro* splicing assay comparing *RLG1* and *rlg1-100* ligase activities......74

Figure III-6. Western blot analysis of ligase in the membrane and soluble cellular fractions of *RLG1* and *rlg1-100* strains........................................................76

Figure III-7. A model for *HAC1* mRNA splicing in yeast...................................78

## Chapter IV

Figure IV-1. C1-Sor$^+$ strain growth on UPR inducing media...............................122

Figure IV-2. C2-Sor$^+$ strain growth on UPR inducing media...............................124

Figure IV-3. C3-Sor$^+$ strain growth on UPR inducing media...............................126

Figure IV-4. Northern blot analysis of *HAC1* mRNA splicing in *RLG1* and *rlg1-100* strains........................................................................................128

Figure IV-5. Northern blot analysis of *HAC1* mRNA splicing in C1-Sor$^+$ strains........130

Figure IV-6. Northern blot analysis of HAC1 mRNA splicing in C2-Sor$^+$ strains........132

Figure IV-7. Northern blot analysis of HAC1 mRNA splicing in C3-Sor$^+$ strains........134

Figure IV-8. Calculated percent *HAC1* mRNA splicing for *RLG1*, *rgl1-100*, C1-Sor$^+$, C2-Sor, and C3-Sor strains..........................................................136

Figure IV-9. Model for *HAC1* mRNA splicing in wild type (*RLG1*) and UPR defective (*rlg1-100*) ligase strains..........................................................138

Figure IV-10. Model for *HAC1* mRNA splicing in C2-Sor$^+$ strains......................140

Figure IV-11. Hac1p proteins capable of being produced from *HAC1* mRNA...........142

Figure IV-12. Northern blot analysis of *HAC1* mRNA splicing in C2-Sor$^+$ strain.......144

Figure IV-13. Western blot analysis of Hac1p production in a C2-Sor$^+$ strain............146

Figure IV-14. Growth of a C2-Sor$^+$ strain carrying wild type, or splice site mutant versions of *HAC1*..........................................................148

Figure IV-15. Northern blot analysis of *HAC1* mRNA splicing in *RLG1*, C2-Sor$^+$ strains..........................................................150

Figure IV-16. C2-Sor$^+$ strains are defective for nonstop mediated mRNA decay........152

Figure IV-17. Model for nonstop mediated mRNA decay in yeast........................154

Figure IV-18. Growth of diploid strains deleted for *SKI2, SKI3,* or *SKI8*..................156

Figure IV-19. Growth of haploid *rlg1-100* strains deleted for *SKI2, SKI3*, or *SKI8*......158

**Appendix A**

Figure A-1. The domains of tRNA ligase..........................................................175

Figure A-2. tRNA ligase genes of yeast..........................................................177

Figure A-3. Secondary structural prediction for the adenylylate synthetase domain of tRNA ligase, a member of the covalent nucleotidyl transferase protein family...........182

Figure A-4. Small islands of homology within the kinase domain of tRNA ligase......184

Figure A-5. Secondary structural prediction for the 2',3'-cyclic phosphodiesterase domain of tRNA ligase, a member of the 2H phosphoesterase protein family.............186

Figure A-6. Primary sequence alignment of tRNA ligase and yeast Ski2p...............188

Figure A-7. A model for binding of cleaved tRNA by T4 polynucleotide kinase........190

**Appendix B**

Figure B-1. Growth of wild type and mutant *tpt1* strains on UPR-inducing media......209

Figure B-2. *HAC1* mRNA splicing and Hac1p production in *tpt1* strains.................210

# Chapter 1

# Introduction:

# The unfolding tale of the unfolded protein response

The endoplasmic reticulum (ER) plays a starring role in the life of all eukaryotic cells. None can live without it. Indeed, one of the most primitive eukaryotes known, Giardia, dispenses with mitochondria and lysosomes, yet it has an ER. As organelles go, one might even argue that the ER is second only in importance to the nucleus itself. What is the ER and why is it so important?

The ER is a large, membrane-enclosed network within the cell. Many essential cellular functions take place within its lumen or on its surface. The ER makes most of the lipids used by the cell; sequesters $Ca^{++}$ for storage and release; folds and covalently modifies all secretory, intraorganelle, and membrane proteins; and functions as the entry and quality control point for the secretory pathway. The ER network is composed of distinct morphological and functional domains. Whereas the ribosome-studded rough ER is the site of cotranslational protein translocation into the ER, the smooth ER is the main site of lipid biosynthesis. There are transport vesicle exit sites where vesicles bound for the cis-Golgi network bud off of the ER. In addition, a portion of the ER membrane forms the nuclear envelope. And finally there is the specialized environment of the ER lumen, a compartment that maintains an oxidizing redox potential and high $Ca^{++}$ concentration relative to the cytosol. The ER lumen is the site where proteins are core glycosylated, their disulfide bonds are formed, and finally, where they are folded.

With such a central cellular role, it is not surprising that disruption of ER functions can lead to devastating disease (Aridor and Balch, 1999; Kuznetsov and Nigam, 1998; Rutishauser and Spiess, 2002) and that cells have evolved means to maintain ER homeostasis. In the face of changing environmental demands, multicellular and single-celled eukaryotes have evolved signaling pathways that assess ER function and respond in ways that, for the most part, reestablish ER function. The unfolded protein response (UPR) is one such pathway that monitors and adjusts the protein folding capacity of the ER.

The protein folding capacity of the ER depends upon the activities of specialized ER-resident chaperone proteins. These chaperones function by promoting the proper folding of newly synthesized proteins and in the process, also prevent their aggregation (Fewell et al., 2001; Stevens and Argon, 1999). The folding capacity of the ER can be decreased by overwhelming the chaperones with a large bolus of improperly folded proteins or by reducing the catalytic function of the chaperones themselves (Table I-1). The UPR increases the folding capacity of the ER by up-regulating production of ER chaperones. The UPR also decreases the folding load on the ER by slowing down the production of new proteins and increasing degradation of misfolded proteins. Yeast and metazoans accomplish this task in a variety of ways.

Three proteins form the core of the UPR signaling pathway in the yeast *Saccharomyces cerevisiae*: Ire1p, a transmembrane serine/threonine kinase and endoribonuclease; tRNA ligase, a protein required for pre-tRNA splicing; and Hac1p, a bZIP transcription factor (Figure I-1). Hac1p stimulates transcription of ER chaperone proteins such as Hsp178 (Bip/Kar2) and protein disulfide isomerase (PDI) by binding to a common regulatory sequence in their promoters, the UPR element (UPRE) (Cox and Walter, 1996). When the UPR is off, translation of the *HAC1* mRNA is blocked by interactions between the *HAC1* 5'UTR and a 252 nucleotide intron located at the 3' end of the Hac1p coding region (Ruegsegger et al., 2001). When the UPR is turned on, this block is relieved by the splicing out of the inhibitory intron and translation resumes (Chapman and Walter, 1997). Ire1p is the sensor and transducer of the unfolded protein signal across the ER membrane. When the UPR is off, the ER chaperone Bip binds to the lumen domain of Ire1p, preventing it from self-associating (Okamura et al., 2000). During the UPR, Bip is thought to be titrated away from Ire1p by binding to the misfolded proteins accumulating in the ER lumen. As a consequence, Ire1p oligomerizes, causing activation of its cytoplasmic kinase and endoribonuclease domains (Shamu and Walter, 1996; Sidrauski and Walter, 1997). Activated Ire1p initiates splicing by cleaving *HAC1*$^{u}$ (u for uninduced) mRNA to liberate the two exons and inhibitory intron. The two exons are joined by tRNA ligase (Sidrauski and Walter, 1997) to produce *HAC1*$^{i}$ (i for induced) mRNA, which is efficiently translated to produce Hac1$^{i}$p. Splicing replaces the last 10 amino acids of Hac1$^{u}$p with a new 18 amino acid tail in Hac1$^{i}$p, creating a more potent transcriptional activator in the process (Mori et al., 2000). Amazingly, splicing takes place on polysome associated *HAC1* mRNA in the cytoplasm (Ruegsegger et al., 2001). Thus unlike any other mRNA in yeast, *HAC1* mRNA is spliced by a non-spliceosomal mechanism that takes place in the cytoplasm.

In addition to stimulating ER chaperone protein synthesis, the yeast UPR also induces expression of the ER associated degradation (ERAD) machinery (Casagrande et al., 2000; Travers et al., 2000). During ERAD, misfolded proteins are retrotranslocated out of the ER and into the cytoplasm, where they are degraded by the cytoplasmic proteosome. By increasing the rate at which misfolded ER proteins are degraded, the protein-folding load on the ER is reduced. The intimate connection between ERAD and chaperone production during the UPR is demonstrated by the synthetic lethal phenotype of yeast simultaneously defective for Ire1p signaling and ERAD. If a cell cannot degrade the excess misfolded proteins, nor fold them properly, it dies.

In the evolutionary jump from free-living, single celled organisms such as yeast to multicellular organisms such as mammals, the UPR became much more complex, both in terms of the number of sensors and the output of the response (Figure I-2). In mammals, four sensors of ER protein folding capacity have been identified: PERK, an eIF2α kinase; ATF6, a bZIP-transcription factor; and IRE1α and IRE1β, paralogues homologous to yeast Ire1p. Each of these proteins senses the folding state of the ER lumen via interactions with Bip. Activation of these sensors results in increased chaperone production, decreased global translation, and cell cycle arrest. Prolonged activation can result in cell death by apoptotic pathways. In these multicellular organisms, the UPR pathway is also required for the differentiation and function of professional secretory cells (DeGracia et al., 2002; Harding et al., 2002; Kaufman, 2002; Kaufman et al., 2002; Ryu et al., 2002; Urano et al., 2000a).

PERK is an integral ER membrane protein, with its C-terminal kinase domain in the cytoplasm and its N-terminal domain in the ER lumen, where it interacts with Bip. On UPR activation, Bip is sequestered away from PERK by the accumulating misfolded proteins, and PERK oligomerizes, activating its kinase domain (Bertolotti et al., 2000). Activated PERK down-regulates global cellular translation by phosphorylating the translation initiation factor eIF4α (Harding et al., 2000b). This has the effect of greatly reducing the protein-folding load on the ER. For some mRNAs, such as that encoding the transcription factor ATF4, global translational down-regulation actually stimulates translation of its mRNA (Harding et al., 2000a). ATF4 in turn activates transcription of the pro-apoptotic transcription factor CHOP, whose mRNA can also be translated under these cellular conditions (Harding et al., 2000a; Wang et al., 1998b; Zinszner et al., 1998). Lastly, PERK activation also down regulates translation of cyclin D1, resulting in cell cycle arrest (Brewer and Diehl, 2000; Brewer et al., 1999). It is thought that the cell's decision between life (pro-survival) or death (pro-apoptosis) is made during this cell cycle arrest (Patil and Walter, 2001).

ATF6 is a bZip transcription factor that also carries a transmembrane domain that tethers it to the ER membrane. The N-terminal domain of ATF6 resides in the ER lumen, where it binds to Bip. Accumulation of unfolded proteins leads to Bip dissociation and ATF6 vesicular transport to the Golgi apparatus (Shen et al., 2002). In the golgi, ATF6 undergoes two sequential proteolytic cleavages, the first by site-1 protease (S1P) and the second by site-2 protease (S2P), liberating the ATF6 transcriptional domain from the golgi membrane and freeing it to travel to the nucleus (Haze et al., 1999; Ye et al., 2000). In the nucleus, ATF6 activates transcription of genes, such as the ER chaperones Bip and GRP94, which carry ER stress response elements (ERSE) in their promoters (Yoshida et al., 1998; Yoshida et al., 2000). Thus ATF6 increases production of ER chaperones, thereby increasing the protein folding capacity of the ER. The S1P and S2P proteases responsible for regulated intramembrane proteolysis (RIP) of ATF6 were originally discovered during RIP of the sterol regulatory binding proteins (SREBPs), transcription factors that upregulate expression of cholesterol synthesis genes in response to membrane sterol depletion. Interestingly, activation of the UPR does not lead to SREBP proteolysis and activation nor does sterol depletion lead to ATF6 proteolysis and activation .

Mammalian IRE1α and IRE1β are orthologues of yeast Ire1p and were originally identified in mice and humans (Niwa et al., 1999; Tirasophon et al., 1998; Wang et al., 1998a). Overexpression of either mammalian IRE1 leads to transcriptional upregulation of a reporter gene carrying an ERSE in its promoter, suggesting that, as is the case in yeast, mammalian IRE1 is important for induction of chaperone gene expression and reestablishment of ER folding capacity during the UPR (Tirasophon et al., 2000). The tissue distribution of the IRE1α and β differ, implying that each performs a different role in the body. This is supported by the phenotypes of IRE1 null mice. IRE1α is expressed through out the body and when knocked-out, results in embryonic lethality (Tirasophon et al., 1998; Urano et al., 2000a; Urano et al., 2000b). IRE1β is expressed predominantly in the epithelium of the gut, and when it is knocked-out, results in viable mice with intestinal problems (Bertolotti et al., 2001; Wang et al., 1998a). The linker region (just N-terminal to the transmembrane domain) and the lumen domain of the IRE1α and β diverge the most in sequence (Niwa et al., 1999). It is likely that these differences are functionally significant within the context of the tissues in which each IRE1 variant is expressed.

IRE1α plays a major role in inducing the pro-apoptotic pathway during the UPR (FigI-2). When activated, IRE1α recruits the cytosolic protein TRAF2 (tumor necrosis factor receptor-associated factor-2) to the ER membrane. This, in turn, leads to activation of the ER caspase-12 as well as the protein kinase JNK (cJUN N-terminal kinase, also known as stress-activated protein kinase (SAPK)). JNK goes on to activate the mitochondrial caspase cascade (Kaufman, 2002). Thus, activation of IRE1α can simultaneously set two caspase cascades in motion, propelling the cell toward selfdestruction. IRE1β may also have a part to play in apoptosis. Translation is significantly reduced in IRE1β overexpressing cells as a consequence of IRE1β mediated cleavage of 28S rRNA. Apoptosis sets in following 28S rRNA cleavage; however, a direct causal relationship has not demonstrated between the two events. (Iwawaki et al., 2001).

When mammalian IRE1 was cloned and its characterization began, the search was renewed with great intensity for a metazoan mRNA that is spliced in an UPR-dependent manner akin to *HAC1* in yeast. Especially provocative were data showing that yeast *HAC1* mRNA is accurately spliced when expressed ectopically in human cells as well as *in vitro* work demonstrating that both forms of human IRE1 accurately cleave *HAC1* mRNA substrates (Niwa et al., 1999). The elusive metazoan mRNA was recently identified as the XBP1 mRNA. Induction of the UPR results in splicing of the XBP1 mRNA in an IRE1-dependent manner (Calfon et al., 2002; Lee et al., 2002; Yoshida et al., 2001). Furthermore, IRE1α cleaves XBP1 mRNA in vitro (Calfon et al., 2002; Shen et al., 2001). The XBP1 splice junction sequences match those identified in *HAC1* mRNA as essential for splicing (Gonzalez et al., 1999). Unexpectedly, whereas the *HAC1* intron is 252 nucleotides long, the XBP1 intron is only 23 nucleotides long. It now appears, though, that the *S.cerevisiae HAC1* intron may be the exception. Two *HAC1* homologues have recently been identified in the filamentous fungi *Trichoderma reesei* and *Aspergillus nidulans*; both *HAC1* mRNAs have 20 nucleotide long introns that are spliced out upon UPR induction (Saloheimo et al., 2003).

Splicing out of the XBP1 intron accomplishes several things. It shifts the reading frame so that sequences within the 3'UTR of the unspliced mRNA now code part of the new protein, XBP1-s (s for spliced). Thus, the C-terminal 97 amino acids of the XBP-u (u for unspliced) protein are replaced by a different set of 212 amino acids in the XBP1-s protein. Splicing of the XBP1 mRNA appears to increase both the translation of the spliced message and the stability of the XBP1-s protein (Calfon et al., 2002). In addition, XBP1-s is a more potent transcriptional activator than XBP1-u (Lee et al., 2002; Yoshida et al., 2001). The increased translation of the spliced XBP1 mRNA and the increased transcriptional activation potential of the XBP-s protein mirror the effects of *HAC1* mRNA splicing in yeast.

XBP-1 is a bZip transcription factor that stimulates transcription of reporter genes carrying the ERSE element in their promoter regions (Lee et al., 2002). Thus, like the UPR transcription factor ATF6, XBP1 likely stimulates the expression of chaperone genes during the UPR. Both ATF6 and IRE1 function upstream of XBP1 as both are required to induce production of XBP1 protein (Yoshida et al., 2001). When the UPR is turned on, ATF6 activates transcription of XBP1 to produce the XBP1 mRNA that IRE1 then splices (Figure I-2) (Lee et al., 2002). Both ATF6 and XBP1-s go on to stimulate further transcription of XBP1 as well as transcription of ER chaperone proteins.

In the worm *C. elegans*, both XBP1 and IRE1 are required for the organism to survive prolonged ER stress induced by unfolded proteins (Shen et al., 2001). However, in mammalian tissue culture, neither is required. In mammalian cells, downstream transcriptional targets of the UPR are induced in the absense of IRE1 function, and cells survive the insult (Lee et al., 2002; Urano et al., 2000a; Urano et al., 2000b). Here, it appears that ATF6 induction of chaperones is sufficient to promote cell survival. If XBP1 and IRE1 are not essential for the UPR in mammalian cells, then what are they doing?

An exciting clue came from the fact that XBP1 function is absolutely essential for terminal differentiation of B-cells into the antibody secreting factories that are plasma cells (Reimold et al., 2001). During this developmental switch, the ER expands by 5-fold to accommodate the intense secretory demands of the plasma cell, estimated to release thousands of antibodies per second into its environment (de StGroth and Scheidegger, 1980; Kaufman, 2002). A massive increase in secretory output, such as that seen in B-cell differentiation, would increase the folding burden on the ER, likely inducing the UPR. IRE1 and spliced XBP1 thus might be specifically required for cells to cope with this developmentally induced type of physiological stress. Recent experiments bear this out. In tissue culture models of B-cell differentiation, XBP1 mRNA splicing is induced by immunoglobin (Ig) production and cellular differentiation. ER chaperone expression increases as well (Gass et al., 2002; Iwakoshi et al., 2003). Moreover, the inability of XBP1 deleted splenic B-cells to secrete Ig can be rescued by ectopic expression from a XBP1 gene carrying the sequence for the spliced form of the protein but not by expression from a XBP1 gene unable to splice out the intron (Iwakoshi et al., 2003). Though plasma cell differentiation has not been investigated using IRE1 deleted cell lines, given the importance of spliced XBP1 in this process, it is highly probable that IRE1 will likewise be required for plasma cell differentiation.

The unfolding tale of the unfolded protein response began back in 1988, with the discovery by Mary-Jane Gething and colleagues that the presence of misfolded proteins in the ER induces the expression of ER chaperones Bip (GRP78) and GRP94 (Kozutsumi et al., 1988). The identification of the yeast Bip homologue, Kar2p (Normington et al., 1989), along with the promoter elements required to upregulate its expression during the

UPR (Mori et al., 1992) soon led to the discovery of the first two components of the UPR signaling pathway in yeast: the integral ER membrane protein, Ire1p (Cox et al., 1993; Mori et al., 1993); and the UPR specific transcription factor, Hac1p (Cox and Walter, 1996). We now know that the salient features of the Ire1-dependent UPR signaling pathway first discovered in *S.cerevisiae* are conserved from yeast, through worms, and on to mammals. The signal sensor and transducer, Ire1p, oligomerizes upon accumulation of misfolded proteins in the ER, activating its endoribonuclease domain. Activated Ire1p initiates splicing of an mRNA encoding a UPR-specific, bZip transcription factor. The spliced mRNA is translated to produce a potent transcriptional activator that upregulates transcription of ER chaperone genes.

The work presented in this thesis adds to the continuing story of the conserved Ire1p signaling pathway in yeast. In Chaper II, we define in detail the minimal primary sequence and secondary structure requirements for Ire1p-mediated cleavage of *HAC1* RNA substrates. This information was vital to the recent identification of XBP1 as the mRNA splicing substrate in metazoans. We also demonstrate that the process by which *HAC1* mRNA is spliced by Ire1p and tRNA ligase is mechanistically equivalent to pre-tRNA splicing (Gonzalez et al., 1999). The unexpected discovery that tRNA ligase is required for *HAC1* mRNA splicing was made when the *rlg1-100* ligase allele was identified in a genetic screen (Sidrauski et al., 1996). Though functional for pre-tRNA splicing, during the UPR, the *rlg1-100* ligase is unable to complete *HAC1* mRNA splicing *in vivo*. In Chapters III and IV, we further explore the role tRNA ligase plays in *HAC1* mRNA splicing by investigating the basis for the UPR defect of the *rlg1-100* allele utilizing biochemical, molecular, and genetic approaches. tRNA ligase is a compelling, multifunctional protein, yet little is understood about how it coordinates the functions of its three catalytic domains. In Appendix A, we take a closer look at the primary sequence of tRNA ligase and compile information from the literature regarding potential functional and structural sequence homologies between tRNA ligase and other, more highly characterized proteins. Lastly, in Appendix B, we describe experiments aimed at determining if the NAD-dependent 2'-phosphotransferase that is required to complete pre-tRNA splicing is also required to produce spliced HAC1 mRNA competent for translation.

*Table I-1. Inducers of the Unfolded Protein Response*

| **Inducer** | **Mechanism** |
|---|---|
| Thapsagargin | *Inhibits ER $Ca^{++}$-ATPase pumps, causing reduction in ER $Ca^{++}$<br>*Reduces function of $Ca^{++}$ dependent ER chaperones |
| Tunicamycin | *Inhibits N-linked glycosylation by inhibiting glycosaminyl-1-phosphate transferase<br>*Unglycosylated proteins don't fold properly |
| DTT | *Reduces disulfide bonds in proteins, causing already folded proteins to unfold and newly translated ER proteins to misfold<br>*Inhibits function of ER chaperones PDI and EroI |
| Reduced cytoplasmic protein degradation | *Reduces rate of ER associated protein degradation (ERAD) by overwhelming the cytoplasmic protein degradation machinery |
| Reduced ERAD | *Mutations in ERAD components and other conditions which decrease ERAD rates lead to increases in misfolded proteins in the ER |
| Suddenly increased secretory load | *Sudden increase in secretory output, transiently overwhelms ER chaprones and thus ER protein folding capacity<br>*Example: B-cell/plasma cell differentiation & antibody secretion |
| Production of misfolded mutant proteins | *Increased demand on ER chaperones and thus ER protein folding capacity |
| Disruption of lipid synthesis | *Possible mechanism is reduction of ER lumenal volume due decreased production of ER membranes<br>*Example: inositol starvation in *S. cerevisiae* |
| Glucose starvation | *reduction in ATP needed for proper ER chaperone function<br>*reduction in raw material used to glycosylate ER proteins |
| Lead | *mechanism unknown<br>*speculation that inhibits ER $Ca^{++}$-ATPase pumps<br>*speculation that inhibits chaperone Bip/GRP78 function by binding it |

**References:** Kaufmann 2002; Treiman 2002; Travers et al. 2000; Stevens and Argon 1999; Iwakoshi et al. 2003; Fewell et al. 2001; Paschen and Frandsen 2001; Qian and Tiffany-Castiglioni 2003; Michalak et al. 2002; Cox et al. 1997.

***Figure I-1 The UPR pathway in yeast***

1. As the newly transcribed *HAC1* mRNA leaves the nucleus and enters the cytoplasm, ribosomes load onto it.
2. Once in the cytoplasm, the *HAC1* 5' UTR and intron basepair, inhibiting translation.
3. Unfolded proteins accumulate in the ER lumen, titrating the ER chaperone Bip away from the lumen domain of Ire1p. As a result, Ire1p oligomerizes, leading to activation of its kinase and endonuclease domains.
4. Activated Ire1p initiates splicing by cleaving *HAC1* mRNA at the two splice sites. tRNA ligase joins the resulting exons together.
5. Splicing out of the intron relieves the translational block, and the ribosomes commense to make Hac1p.
6. Hac1p travels to the nucleus and activates transcription of genes encoding ER chaperone proteins such as Bip that carry a unfolded protein response element (UPRE) in their promoter regions.
7. We thank Jason Brickner for the kind use of this figure.

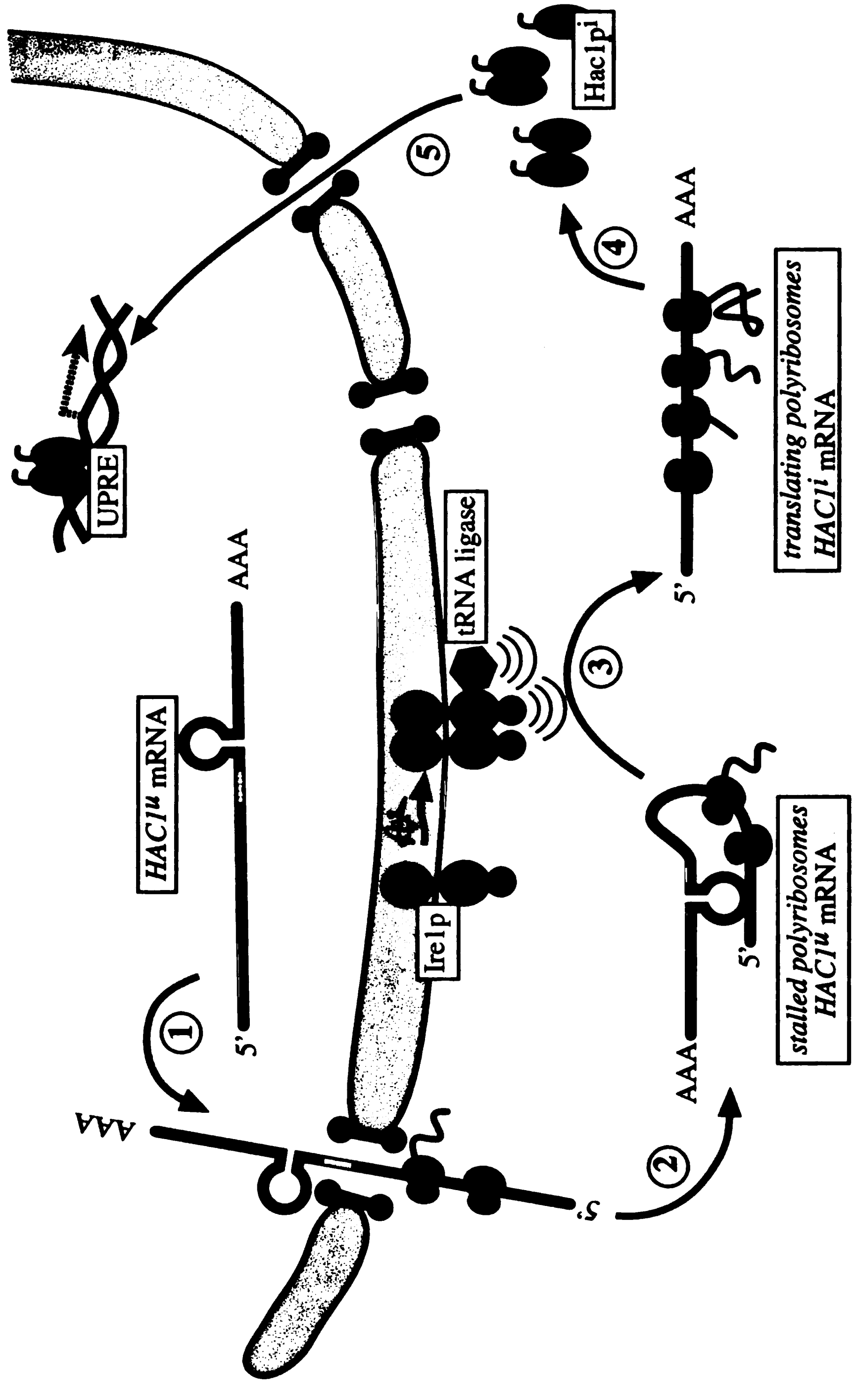
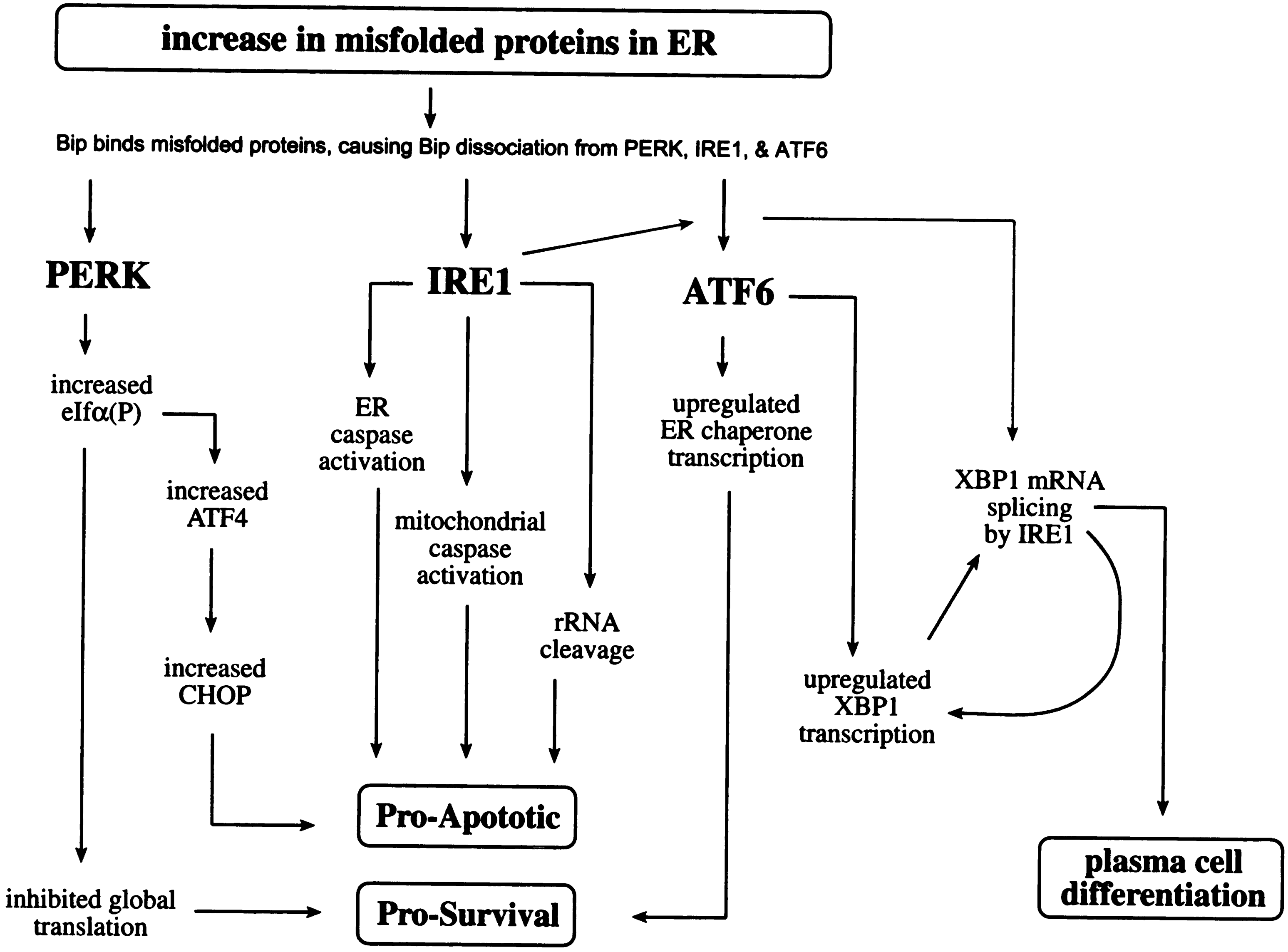

①
②
③
④
⑤
AAA
5'
HAC1$^{u}$ mRNA
AAA
5'
Ire1p
tRNA ligase
UPRE
Hac1p$^{i}$
AAA
5'
stalled polyribosomes
HAC1$^{u}$ mRNA
AAA
5'
translating polyribosomes
HAC1$^{i}$ mRNA

***Figure I-2 The diversity of the UPR in mammals***

PERK, IRE1, ATF6: the three sensors/transducers of unfolded proteins in the ER and their downstream outputs. For clarity, splicing of the XBP1 mRNA by IRE1 is shown in red. Details of these pathways are discussed in the text. Adapted from (DeGracia et al., 2002).


increase in misfolded proteins in ER
Bip binds misfolded proteins, causing Bip dissociation from PERK, IRE1, & ATF6
PERK
IRE1
ATF6
increased eIfα(P)
increased ATF4
increased CHOP
inhibited global translation
ER caspase activation
mitochondrial caspase activation
rRNA cleavage
upregulated ER chaperone transcription
upregulated XBP1 transcription
XBP1 mRNA splicing by IRE1
Pro-Apototic
Pro-Survival
plasma cell differentiation

## *REFERENCES*

Aridor, M., and Balch, W. E. (1999). Integration of endoplasmic reticulum signaling in health and disease. Nat Med *5*, 745-751.

Bertolotti, A., Wang, X., Novoa, I., Jungreis, R., Schlessinger, K., Cho, J. H., West, A. B., and Ron, D. (2001). Increased sensitivity to dextran sodium sulfate colitis in IRE1beta-deficient mice. J Clin Invest *107*, 585-593.

Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P., and Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nat Cell Biol *2*, 326-332.

Brewer, J. W., and Diehl, J. A. (2000). PERK mediates cell-cycle exit during the mammalian unfolded protein response. Proc Natl Acad Sci U S A *97*, 12625-12630.

Brewer, J. W., Hendershot, L. M., Sherr, C. J., and Diehl, J. A. (1999). Mammalian unfolded protein response inhibits cyclin D1 translation and cell-cycle progression. Proc Natl Acad Sci U S A *96*, 8505-8510.

Calfon, M., Zeng, H., Urano, F., Till, J. H., Hubbard, S. R., Harding, H. P., Clark, S. G., and Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. Nature *415*, 92-96.

Casagrande, R., Stern, P., Diehn, M., Shamu, C., Osario, M., Zuniga, M., Brown, P. O., and Ploegh, H. (2000). Degradation of proteins from the ER of S. cerevisiae requires an intact unfolded protein response pathway. Mol Cell *5*, 729-735.

Chapman, R. E., and Walter, P. (1997). Translational attenuation mediated by an mRNA intron. Curr Biol *7*, 850-859.

Cox, J. S., Chapman, R. E., and Walter, P. (1997). The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Mol Biol Cell *8*, 1805-1814.

Cox, J. S., Shamu, C. E., and Walter, P. (1993). Transcriptional Induction of Genes Encoding Endoplasmic Reticulum Resident Proteins Requires a Transmembrane Protein Kinase. Cell *73*, 1197-1206.

Cox, J. S., and Walter, P. (1996). A novel mechanism for regulating activity of a trancription factor that controls the unfolded protein response. Cell *87*, 391-404.

de StGroth, S. F., and Scheidegger, D. (1980). Production of monoclonal antibodies: strategy and tactics. J Immunol Methods *35*, 1-21.

DeGracia, D. J., Kumar, R., Owen, C. R., Krause, G. S., and White, B. C. (2002). Molecular pathways of protein synthesis inhibition during brain reperfusion: implications for neuronal survival or death. J Cereb Blood Flow Metab *22*, 127-141.

Fewell, S. W., Travers, K. J., Weissman, J. S., and Brodsky, J. L. (2001). The action of molecular chaperones in the early secretory pathway. Annu Rev Genet *35*, 149-191.

Gass, J. N., Gifford, N. M., and Brewer, J. W. (2002). Activation of an unfolded protein response during differentiation of antibody-secreting B cells. J Biol Chem *277*, 49047-49054.

Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. Embo J *18*, 3119-3132.

Harding, H. P., Calfon, M., Urano, F., Novoa, I., and Ron, D. (2002). Transcriptional and translational control in the Mammalian unfolded protein response. Annu Rev Cell Dev Biol *18*, 575-599.

Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., and Ron, D. (2000a). Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol Cell *6*, 1099-1108.

Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H., and Ron, D. (2000b). Perk is essential for translational regulation and cell survival during the unfolded protein response. Mol Cell *5*, 897-904.

Haze, K., Yoshida, H., Yanagi, H., Yura, T., and Mori, K. (1999). Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. Mol Biol Cell *10*, 3787-3799.

Iwakoshi, N. N., Lee, A. H., Vallabhajosyula, P., Otipoby, K. L., Rajewsky, K., and Glimcher, L. H. (2003). Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. Nat Immunol. *4*, 321-329.

Iwawaki, T., Hosoda, A., Okuda, T., Kamigori, Y., Nomura-Furuwatari, C., Kimata, Y., Tsuru, A., and Kohno, K. (2001). Translational control by the ER transmembrane kinase/ribonuclease IRE1 under ER stress. Nat Cell Biol *3*, 158-164.

Kaufman, R. J. (2002). Orchestrating the unfolded protein response in health and disease. J Clin Invest *110*, 1389-1398.

Kaufman, R. J., Scheuner, D., Schroder, M., Shen, X., Lee, K., Liu, C. Y., and Arnold, S. M. (2002). The unfolded protein response in nutrient sensing and differentiation. Nat Rev Mol Cell Biol *3*, 411-421.

Kozutsumi, Y., Segal, M., Normington, K., Gething, M. J., and Sambrook, J. (1988). The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. Nature *332*, 462-464.

Kuznetsov, G., and Nigam, S. K. (1998). Folding of secretory and membrane proteins. N Engl J Med *339*, 1688-1695.

Lee, K., Tirasophon, W., Shen, X., Michalak, M., Prywes, R., Okada, T., Yoshida, H., Mori, K., and Kaufman, R. J. (2002). IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response. Genes Dev *16*, 452-466.

Michalak, M., Parker, J. M. R., and Opas, M. (2002). $Ca^{2+}$ signaling and calcium binding chaperones of the endoplasmic reticulum. Cell Calcium *32*, 269-278.

Mori, K., Ma, W., Gething, M. J., and Sambrook, J. (1993). A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus. Cell *74*, 743-756.

Mori, K., Ogawa, N., Kawahara, T., Yanagi, H., and Yura, T. (2000). mRNA splicing-mediated C-terminal replacement of transcription factor Hac1p is required for efficient activation of the unfolded protein response. Proc Natl Acad Sci U S A *97*, 4660-4665.

Mori, K., Sant, A., Kohno, K., Normington, K., Gething, M. J., and Sambrook, J. F. (1992). A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. Embo J *11*, 2583-2593.

Niwa, M., Sidrauski, C., Kaufman, R. J., and Walter, P. (1999). A role for presenilin-1 in nuclear accumulation of Ire1 fragments and induction of the mammalian unfolded protein response. Cell *99*, 691-702.

Normington, K., Kohno, K., Kozutsumi, Y., Gething, M. J., and Sambrook, J. (1989). S. cerevisiae Encodes an Essential Protein Homologous in Sequence and Function to Mammalian BiP. Cell *57*, 1223-1236.

Okamura, K., Kimata, Y., Higashio, H., Tsuru, A., and Kohno, K. (2000). Dissociation of Kar2p/BiP from an ER sensory molecule, Ire1p, triggers the unfolded protein response in yeast. Biochem Biophys Res Commun *279*, 445-450.

Paschen, W., and Frandsen, A. (2001). Endoplasmic reticulum dysfunction - a common denominator for cell injury in acute and degenerative diseases of the brain? J Neurochem *79*, 719-725.

Patil, C., and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Curr Opin Cell Biol *13*, 349-355.

Qian, Y., and Tiffany-Castiglioni, E. (2003). Lead-induced endoplasmic reticulum (ER) stress reponses in the nervous system. Neurochem Res *28*, 153-162.

Reimold, A. M., Iwakoshi, N. N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallese, E. M., Friend, D., Grusby, M. J., Alt, F., and Glimcher, L. H. (2001). Plasma cell differentiation requires the transcription factor XBP-1. Nature *412*, 300-307.

Ruegsegger, U., Leber, J. H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103-114.

Rutishauser, J., and Spiess, M. (2002). Endoplasmic reticulum storage diseases. Swiss Med Wkly *132*, 211-222.

Ryu, E. J., Harding, H. P., Angelastro, J. M., Vitolo, O. V., Ron, D., and Greene, L. A. (2002). Endoplasmic reticulum stress and the unfolded protein response in cellular models of Parkinson's disease. J Neurosci *22*, 10690-10698.

Saloheimo, M., Valkonen, M., and Penttila, M. (2003). Activation mechanisms of the HACI-mediated unfolded protein response in filamentous fungi. Mol Microbiol *47*, 1149-1161.

Shamu, C. E., and Walter, P. (1996). Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. Embo J *15*, 3028-3039.

Shen, J., Chen, X., Hendershot, L., and Prywes, R. (2002). ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of Golgi localization signals. Dev Cell *3*, 99-111.

Shen, X., Ellis, R. E., Lee, K., Liu, C. Y., Yang, K., Solomon, A., Yoshida, H., Morimoto, R., Kurnit, D. M., Mori, K., and Kaufman, R. J. (2001). Complementary signaling pathways regulate the unfolded protein response and are required for C. elegans development. Cell *107*, 893-903.

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response [see comments]. Cell *87*, 405-413.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell *90*, 1-20.

Stevens, F. J., and Argon, Y. (1999). Protein folding in the ER. Semin Cell Dev Biol *10*, 443-454.

Tirasophon, W., Lee, K., Callaghan, B., Welihinda, A., and Kaufman, R. J. (2000). The endoribonuclease activity of mammalian IRE1 autoregulates its mRNA and is required for the unfolded protein response. Genes Dev *14*, 2725-2736.

Tirasophon, W., Welihinda, A. A., and Kaufman, R. J. (1998). A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells. Genes Dev *12*, 1812-1824.

Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249-258.

Treiman, M. (2002). Regulation of endoplasmic reticulum calcium storage during the unfolded protein response - significance in tissue ischemia? TCM *12*, 57-62.

Urano, F., Bertolotti, A., and Ron, D. (2000a). IRE1 and efferent signaling from the endoplasmic reticulum. J Cell Sci *113 Pt 21*, 3697-3702.

Urano, F., Wang, X., Bertolotti, A., Zhang, Y., Chung, P., Harding, H. P., and Ron, D. (2000b). Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. Science *287*, 664-666.

Wang, X. Z., Harding, H. P., Zhang, Y., Jolicoeur, E. M., Kuroda, M., and Ron, D. (1998a). Cloning of mammalian Ire1 reveals diversity in the ER stress responses. Embo J *17*, 5708-5717.

Wang, X. Z., Kuroda, M., Sok, J., Batchvarova, N., Kimmel, R., Chung, P., Zinszner, H., and Ron, D. (1998b). Identification of novel stress-induced genes downstream of chop. Embo J *17*, 3619-3630.

Ye, J., Rawson, R. B., Komuro, R., Chen, X., Dave, U. P., Prywes, R., Brown, M. S., and Goldstein, J. L. (2000). ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. Mol Cell *6*, 1355-1364.

Yoshida, H., Haze, K., Yanagi, H., Yura, T., and Mori, K. (1998). Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. J Biol Chem *273*, 33741-33749.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., and Mori, K. (2001). XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell *107*, 881-891.

Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., and Mori, K. (2000). ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response. Mol Cell Biol *20*, 6755-6767.

Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R. T., Remotti, H., Stevens, J. L., and Ron, D. (1998). CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. Genes Dev *12*, 982-995.

## Chapter 2

# Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway

**The EMBO Journal** Vol.18 No.11 pp.3119–3132, 1999

# Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway

**Tania N.Gonzalez, Carmela Sidrauski, Silke Dörfler and Peter Walter**[1]

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448, USA

[1]Corresponding author
e-mail: walter@cgl.ucsf.edu

**The unfolded protein response is an intracellular signaling pathway that, in response to accumulation of misfolded proteins in the lumen of the endoplasmic reticulum (ER), upregulates transcription of ER resident chaperones. A key step in this pathway is the nonconventional, regulated splicing of the mRNA encoding the positive transcriptional regulator Hac1p. In the yeast *Saccharomyces cerevisiae*, the bifunctional transmembrane kinase/endoribonuclease Ire1p cleaves *HAC1* mRNA at both splice junctions and tRNA ligase joins the two exons together. We have reconstituted *HAC1* mRNA splicing in an efficient *in vitro* reaction and show that, in many ways, the mechanism of *HAC1* mRNA splicing resembles that of pre-tRNA splicing. In particular, Ire1p endonucleolytic cleavage leaves $2',3'$-cyclic phosphates, the excised exons remain associated by base pairing, and exon ligation by tRNA ligase follows the same chemical steps as for pre-tRNA splicing. To date, this mechanism of RNA processing is unprecedented for a messenger RNA. In contrast to the striking similarities to tRNA splicing, the structural features of the splice junctions recognized by Ire1p differ from those recognized by tRNA endonuclease. We show that small stem–loop structures predicted to form at both splice junctions of *HAC1* mRNA are required and sufficient for Ire1p cleavage.**

*Keywords*: endoplasmic reticulum/Ire1p/mRNA splicing/tRNA ligase/unfolded protein response

## Introduction

The lumen of the endoplasmic reticulum (ER) is a highly specialized compartment in eukaryotic cells. Secretory and most membrane proteins are folded, covalently modified and oligomerized in this compartment with the assistance of specialized ER resident proteins (Gething and Sambrook, 1992). It is only after proper folding and oligomeric assembly that proteins are able to continue their journey through the secretory pathway to the cell surface or to ER-distal compartments, such as the Golgi apparatus or lysosomes (Hammond and Helenius, 1995). Perturbations in the ER lumenal environment can thus be highly detrimental, as they can block production of many essential cellular components. Regulatory networks have evolved to detect and respond to changes in the ER, thus enabling cells to maintain an optimal folding environment in the ER. An example of this is the unfolded protein response (UPR), an ER to nucleus signaling pathway found in all eukaryotic cells, which is induced by accumulation of unfolded proteins in the ER. The UPR enables cells to increase the folding capacity of the ER lumen by increasing the transcription of genes encoding ER resident proteins that mediate protein folding, such as BiP, an hsp70-like chaperone, or PDI, an enzyme that catalyzes disulfide bond formation (reviewed in Shamu *et al.*, 1994; Chapman *et al.*, 1998; Sidrauski *et al.*, 1998).

Four factors involved in UPR signaling have been identified in the yeast *Saccharomyces cerevisiae*: Ire1p, a transmembrane serine/threonine kinase that also exhibits site-specific endoribonuclease activity (Cox *et al.*, 1993; Mori *et al.*, 1993; Sidrauski and Walter, 1997); Ptc2p, a serine/threonine phosphatase thought to modulate the activity of Ire1p (Welihinda *et al.*, 1998); tRNA ligase, an RNA processing enzyme required for splicing of tRNA precursors (Greer *et al.*, 1983; Phizicky *et al.*, 1992; Sidrauski *et al.*, 1996; Sidrauski and Walter, 1997); and Hac1p, a member of the leucine zipper family of transcription factors (Cox and Walter, 1996; Mori *et al.*, 1996; Nikawa *et al.*, 1996).

Ire1p is localized to the ER and/or inner nuclear membranes, which are contiguous with one another. The N-terminal half of Ire1p lies in the ER lumen (Mori *et al.*, 1993), where it senses by an unknown mechanism increases in the concentration of misfolded proteins. The kinase and endoribonuclease domains of Ire1p map to its C-terminal half and are located in either the cytoplasm or nucleus, where they induce downstream events in the UPR pathway. Like many other transmembrane kinases, oligomerization of Ire1p in the plane of the membrane induces its kinase activity and leads to Ire1p phosphorylation (Shamu and Walter, 1996; Welihinda and Kaufman, 1996). Concomitantly, the Ire1p endoribonucleolytic activity is induced (Sidrauski and Walter, 1997). The only substrate of Ire1p endoribonuclease known to date is the mRNA encoding the UPR-specific transcription factor Hac1p.

Hac1p upregulates transcription of genes encoding ER resident proteins by binding to a common regulatory sequence in their promoters, the UPR element (UPRE) (Mori *et al.*, 1992, 1996; Kohno *et al.*, 1993; Cox and Walter, 1996; Nikawa *et al.*, 1996). Interestingly, Hac1p is regulated by changes in its abundance (Cox and Walter, 1996; Kawahara *et al.*, 1997); the level of Hac1p, in turn, is controlled by the regulated splicing of its mRNA. In the absence of splicing, *HAC1* mRNA translation is inhibited due to the presence of a 252 nucleotide intron (Chapman and Walter, 1997; Kawahara *et al.*, 1997). The mechanism by which the presence of the intron attenuates translation of *HAC1* mRNA is currently unknown.

© European Molecular Biology Organization

Removal of the intron occurs by a non-conventional reaction mechanism that is catalyzed by Ire1p and tRNA ligase (Sidrauski and Walter, 1997). Upon accumulation of unfolded proteins in the ER, Ire1p initiates splicing by cleaving *HAC1*$^u$ mRNA (*u* for *uninduced*) to liberate the intron, and the 5′ and 3′ exons. The two exons are subsequently joined by tRNA ligase to produce *HAC1*$^i$ mRNA (*i* for *induced*), which is efficiently translated to produce Hac1p. Thus, the regulated splicing of *HAC1* mRNA is a key regulatory step in the UPR signaling pathway.

Although best understood in yeast, the salient features of this unusual signaling pathway are likely to be conserved in all eukaryotic cells. Homologues of Ire1p have been identified in *Caenorhabditis elegans* and in mammalian cells. The mammalian homologues are likely to function in a corresponding UPR pathway as is suggested from the phenotype of dominant negative mutants in these components (Tirasophon *et al.*, 1998; Wang *et al.*, 1998). Moreover, we have recently shown that mammalian cells accurately splice the yeast *HAC1* mRNA intron (M.Niwa and P.Walter, unpublished results). Thus, both the signal transduction components and mechanism appear phylogenetically conserved from yeast to mammals. This conservation makes it likely that our understanding of the molecular mechanism of the UPR in yeast will be directly applicable to our understanding of the corresponding pathway in human cells, where defects in protein folding in the ER can lead to devastating diseases (Kuznetsov and Nigam, 1998).

The splicing of *HAC1* mRNA by Ire1p and tRNA ligase is unprecedented, and little is known regarding its mechanism. Significantly, tRNA ligase, an enzyme previously thought to function exclusively in the splicing of pre-tRNAs, has been shown both *in vivo* and *in vitro* to participate in this reaction (Sidrauski *et al.*, 1996; Sidrauski and Walter, 1997). In fact, we have previously shown that Ire1p and tRNA ligase are sufficient to splice *HAC1*$^u$ mRNA *in vitro*, supporting the notion that these are the only two components absolutely required for the reaction. This is in striking contrast to splicing of all other pre-mRNAs which require >100 proteins and small nuclear RNAs constituting the spliceosome and its associated components (Moore *et al.*, 1993). Pre-tRNA splicing, on the other hand, is more akin to *HAC1* mRNA splicing in that it also requires only two components for cleavage and ligation, tRNA endonuclease and tRNA ligase (Greer *et al.*, 1983; Peebles *et al.*, 1983; Belford *et al.*, 1993). Thus, we previously speculated that pre-tRNA and *HAC1*$^u$ mRNA splicing occur by similar mechanisms.

The mechanism by which pre-tRNAs are spliced is well understood (reviewed in Westaway and Abelson, 1995; Abelson *et al.*, 1998). tRNA endonuclease cleaves precursor tRNA, releasing base-paired 5′ and 3′ exons and a linear intron. Cleavage generates a 2′,3′-cyclic phosphate terminus at the 3′ end of the 5′ exon, and a 5′-OH terminus at the 5′ end of the 3′ exon (Knapp *et al.*, 1979; Peebles *et al.*, 1983). The two exons are subsequently joined in a series of reactions catalyzed by the multifunctional enzyme tRNA ligase (Greer *et al.*, 1983; Xu *et al.*, 1990; Belford *et al.*, 1993). The first step involves opening of the terminal 2′,3′-cyclic phosphate to leave a 2′-phosphate at the 3′ end of the 5′ exon. The 3′ exon is phosphorylated at its 5′ terminus after transfer of γ-phosphate from either GTP or ATP. Thus, the phosphate group derived from the γ-position of a nucleotide triphosphate ultimately links the two exons in the spliced mRNA. The 5′ terminal phosphate is next activated by the transfer of AMP from tRNA ligase to form a high energy 5′–5′ phosphoanhydride bond. Ligation occurs with the concomitant release of the AMP activating group. The 2′ phosphate remaining at the splice junction is later removed from spliced tRNA by a third enzyme, nicotinamide adenine dinucleotide (NAD)-dependent 2′-phosphotransferase (McCraith and Phizicky, 1990; Spinelli *et al.*, 1997).

By extension, processing of *HAC1*$^u$ mRNA may proceed in a manner similar to that outlined above. Indeed, both reactions are initiated by highly substrate-specific endoribonucleases (Peebles *et al.*, 1983; Di Nicola Negri *et al.*, 1997; Sidrauski and Walter, 1997). In addition, both endonucleases cleave their substrate 5′ and 3′ splice junctions independently of one another and in no obligate order (Miao and Abelson, 1993; Sidrauski and Walter, 1997; Kawahara *et al.*, 1998). Therefore, they share a splicing chemistry that is strictly incompatible with the spliceosome-catalyzed mechanism, where cleavage of the 3′ splice junction must always follow the cleavage of the 5′ splice junction (Moore *et al.*, 1993). However, Ire1p and tRNA endonuclease lack any significant similarity in amino acid sequence or subunit composition. Whereas tRNA endonuclease is a tetramer composed of four different polypeptide chains (Trotta *et al.*, 1997) that constitutively splices precursor tRNA, Ire1p is composed of one subunit that is thought to homo-oligomerize (Shamu and Walter, 1996; Welihinda and Kaufman, 1996) and its endoribonucleolytic activity is tightly regulated by conditions in the ER lumen. Furthermore, Ire1p and tRNA endonuclease differ significantly with respect to their substrate recognition. tRNA endonuclease recognizes the folded tertiary structure of pre-tRNA, as well as local structures at the intron–exon boundaries, and pays little attention to the sequence at or near the splice sites (Mattoccia *et al.*, 1988; Reyes and Abelson, 1988; Fabbri *et al.*, 1998). In contrast, point mutations generated near the splice junctions of *HAC1* mRNA have been shown to block its splicing both *in vivo* and *in vitro* (Sidrauski and Walter, 1997; Kawahara *et al.*, 1998). To address questions regarding both the similarities and differences in the mechanism of *HAC1* mRNA and pre-tRNA splicing experimentally, we have developed efficient *in vitro* reactions which have allowed us to characterize the mechanism and the substrate requirements for *HAC1* mRNA splicing.

## Results

### *The predicted HAC1 mRNA splice junction stem–loops act as Ire1p minisubstrates*

As previously shown, Ire1p is a site-specific endoribonuclease that cleaves *HAC1*$^u$ mRNA at both splice junctions (Sidrauski and Walter, 1997). Cleavage at either junction can occur independently of cleavage at the other junction; cleavage specificity, therefore, must result from either sequence and/or structural motifs that are common to both splice junctions. Indeed, the predicted secondary structure of the *HAC1*$^u$ intron (Figure 1) reveals similar stem–loop

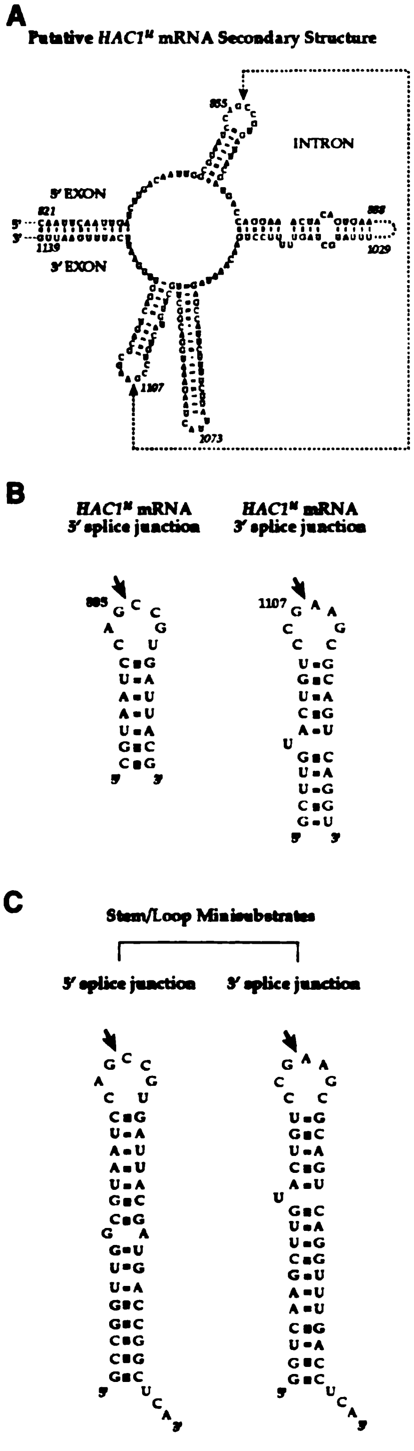


**Fig. 1.** Secondary structure prediction of the *HAC1*$^u$ mRNA intron and splice site stem–loop structures. The secondary structure prediction of the *HAC1*$^u$ mRNA intron and flanking 5′ and 3′ exon regions is shown (**A**) together with a close-up view of predicted stem–loop structures at the 5′ and 3′ splice sites (**B**). Ire1p(k+t) cleavage sites are indicated by arrows and were confirmed in Figure 6. (**C**) Sequence and predicted structures of the minisubstrate stem–loop RNAs used in this study.

structures at the 5′ and 3′ splice site junctions (Sidrauski and Walter, 1997; Kawahara *et al.*, 1998). At either site, the RNA is predicted to fold into short stems with 7-nucleotide loops, each containing a G residue in the third position. Previous work suggested that nucleolytic cleavage by Ire1p occurs at this invariant G, but conflicting models have been proposed as to whether Ire1p cleaves at its 5′ or 3′ side (Kawahara *et al.*, 1997; Sidrauski and Walter, 1997).

To analyze the substrate specificity and the chemistry of the Ire1p-mediated cleavage reaction, we first asked whether the stem–loop structure that is proposed in Figure 1B would be sufficient to direct Ire1p cleavage. To this end, we designed 'minisubstrates', two short RNA molecules shown in Figure 1C that are predicted to fold into structures similar to those predicted at either splice site for authentic *HAC1*$^u$ mRNA. In these minisubstrate stem–loop RNAs, we extended the lengths of the stems by five or six base pairs with unrelated sequences to stabilize the stems. A 3-nucleotide 3′ extension was also added so that the anticipated cleavage products could be more easily distinguished by their size. Stem–loop RNAs were made by *in vitro* transcription with T7 polymerase using synthetic DNA oligonucleotide templates. Transcripts were uniformly labeled during transcription by incorporation of [$^{32}$P]UTP and gel purified.

When incubated with a fragment of Ire1p termed Ire1p(k+t) (Sidrauski and Walter, 1997), which contains the Ire1p kinase domain and C-terminal tail domain and which has been proposed to contain the nuclease active site, the stem–loop RNAs corresponding to either splice site were efficiently cleaved to produce two discrete fragments (Figure 2A). The size of the cleavage products suggested that cleavage occurred at or in close proximity to the predicted site. Cleavage of RNA that was either internally labeled with [$\alpha$-$^{32}$P]UTP during transcription (Figure 2B, lane 4), or that was transcribed in the absence of labeled nucleotide and then labeled at the 5′ terminus with [$\gamma$-$^{32}$P]ATP and polynucleotide kinase (Figure 2B, lane 3), confirmed that the faster migrating band corresponded to the 5′ end of the RNA.

To ascertain that the stem–loop RNAs behaved similarly to authentic *HAC1*$^u$ RNA, we tested the ability of Ire1p(k+t) to cleave our minisubstrates carrying single nucleotide substitutions in loop residues. Previous studies had identified mutations in the predicted loops that prevent Ire1p from cleaving *HAC1*$^u$ mRNA *in vitro* or *in vivo*. Indeed, as shown in Figure 2B (lane 6), a G→C mutation in the loop corresponding to position 1107 in the full-length *HAC1*$^u$ mRNA completely abolished cleavage of the 3′ stem–loop RNA as previously shown for authentic *HAC1*$^u$ mRNA (Sidrauski and Walter, 1997). Likewise, mutations at two other loop positions (C1105G and G1110C), each representing additional invariant bases found at both splice sites, also prevented cleavage by Ire1p(k+t) (Figure 2B, lanes 5 and 7), consistent with previous *in vivo* splicing data (Kawahara *et al.*, 1998). Thus all three residues in the loop whose positions are invariant between both splice junctions are critical for Ire1p(k+t) cleavage.

To obtain a complete picture of the constraints that loop residues put on the Ire1p(k+t) cleavage reaction, we tested a series of mutant 3′ stem–loop minisubstrates in which

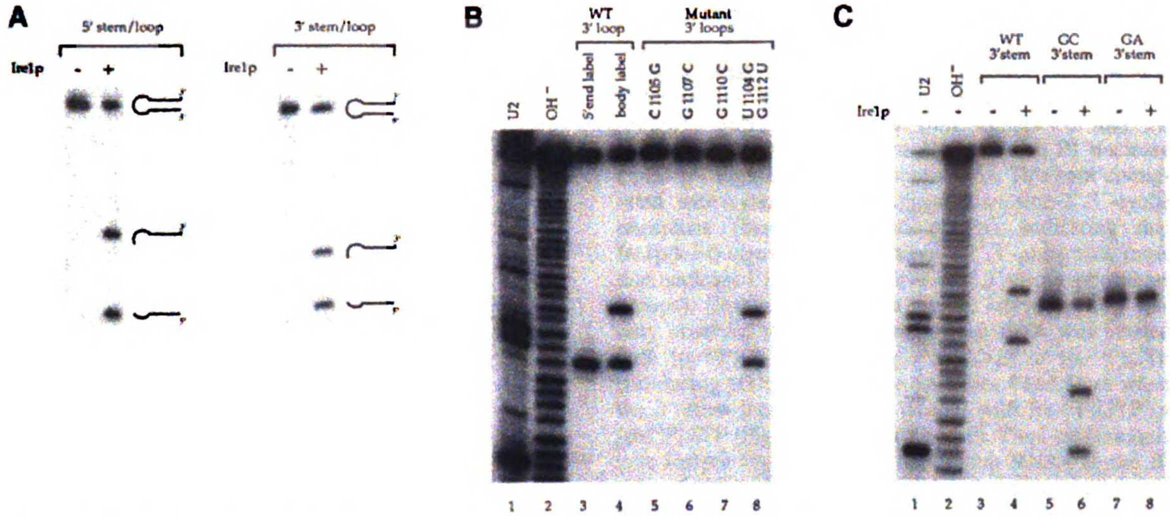


**Fig. 2.** Cleavage of stem–loop RNAs by Ire1p(k+t). **(A)** Radiolabeled *in vitro* transcribed 5′ stem–loop RNA (2000 c.p.m.) (lanes 1 and 2) and 3′ stem–loop RNA (2000 c.p.m.) (lanes 3 and 4) as shown in Figure 1C, were incubated in cleavage buffer (see Materials and methods) in the presence or absence of 1 µg Ire1p(k+t) at 30°C for 2 h. The reaction volume was 20 µl. The reaction products were separated on denaturing polyacrylamide gel and visualized by autoradiography. **(B)** The results of incubation of various mutant loop RNAs with Ire1p(k+t) are shown in lanes 5–8, with the specific loop mutation given above each lane. Refer to Figure 1 to place these in the context of the wild-type 3′ stem–loop. Lanes 1 and 2 carry markers produced by nuclease U2 digestion (lane 1) or alkaline hydrolysis (lane 2) of 5′ end-labeled wild-type 3′ stem–loop RNA. Lanes 3 and 4 contain 5′ end-labeled (lane 3) and internally labeled (lane 4) wild-type 3′ stem–loop RNAs that were cleaved with Ire1p(k+t), respectively. **(C)** The following RNAs were incubated in the presence or absence of Ire1p(k+t) as indicated: wild-type 3′ stem–loop RNA (lanes 3 and 4), a 3′ stem–loop RNA in which the wild-type stem sequence was replaced by a stem of GC base pairs (see Materials and methods) (lanes 5 and 6), and a 3′ stem–loop where every C of the GC stem was replaced by an A (lanes 7 and 8) to abolish base pairing. Note that the substrates in lanes 5–8 are 10 nucleotides shorter than the wild-type 3′ stem–loop RNA in lanes 3 and 4. Lanes 1 and 2 are as in (B).

each loop nucleotide was individually changed to every other ribonucleotide. These data, summarized in Figure 3, show that there is a hierarchy of importance. Most notably, no base substitution was tolerated at the three invariant bases (C1105, G1107 and G1110) that are common to both splice sites. As indicated in Figure 3, these residues are located at positions –3, –1 and +3 with respect to the splice site. Loop position +1 could only be an A or a C, which corresponds to the bases found at the 3′ and 5′ splice junction. Both A and C provide an H-bond donor (a primary amine) as well as an H-bond acceptor (a tertiary amine) at the same position on the 6-membered ring of each nucleotide. This may indicate a role for these functional groups for structuring the loop and/or for forming contacts with Ire1p(k+t). In position –2 all four bases worked equally well, whereas the other two positions (+2 and +4) showed some, albeit weak preference for the bases found at the wild-type 3′ splice site. Interestingly, mutating these residues to those found at the 5′ splice junction reduced cleavage, suggesting that these positions may be sensitive to the sequence context present in either loop. Overall, the effects of these mutations on Ire1p(k+t) cleavage could reflect sequence-specific and/or structural constraints required for Ire1p(k+t) recognition of these minisubstrates.

Mutations predicted to allow pairing of the bases at positions –3, +4 (C–3 to G, or C+4 to G), and hence predicted to create a smaller 5-base loop, abolished cleavage. Likewise, a mutation changing U at position –4 to an A and hence predicted to open the ultimate base pair of the stem impaired cleavage. Taken together, these results suggest that the loop must be constrained to its wild-type length of 7 nucleotides to fold into a defined tertiary structure. Additional support for folding of the bases into a structured loop was obtained by S1 nuclease probing experiments. S1 nuclease cleaves single-stranded RNA without regard to primary sequence. Only the bonds following the nucleotides at the –1 and +1 positions were sensitive to S1 nuclease cleavage (data not shown), suggesting that the remaining 5-loop residues are inaccessible to the S1 nuclease.

With the exception of positions +1 and –4, our results are consistent with data from a previous study investigating *HAC1* mRNA splicing *in vivo*. In this study, changing position +1 from A to U or changing position –4 from U to A reduced splicing by 25% or less (Kawahara *et al.*, 1998). This discrepancy most likely reflects the fact that Ire1p interactions with *HAC1*$^u$ mRNA are stronger than those with the stem–loop minisubstrates.

We next confirmed the importance of the stem for Ire1p(k+t) cleavage. Switching the bases that form the proposed U–G pair closing the loop did not diminish cleavage (Figure 2B, lane 8), suggesting that the nucleotide sequence is not important as long as the regions flanking the 7-membered loop can form a stable stem. To test this notion further, we made a stem–loop in which all base pairs in the stem, with the exception of the UG pair that closes off the loop, were changed to GC or CG pairs. Due to the high melting temperature of GC and CG base pairs, we were also able to shorten the stem down to 11 bp from the 16 bp used in our original 3′ stem–loop minisubstrate (Figure 1C). Ten base pairs is the predicted length of the 3′ stem–loop within the intact *HAC1*$^u$ mRNA (Figure 1A and B). As shown in Figure 2C, lanes 5 and 6, cleavage

UCSF LIBRARY

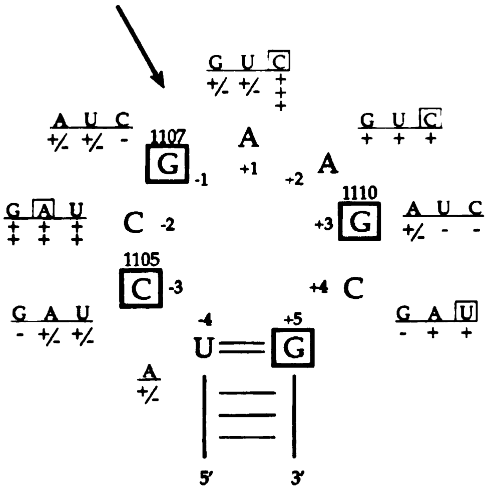


**Fig. 3.** Mutational analysis of 3′ splice site loop. A series of twenty-two 3′ stem–loop mutants, each carrying a single loop point mutation, was transcribed and radiolabeled as in Figure 2. The wild-type 3′ loop sequence is shown with the 4 nucleotides whose positions are invariant between 5′ and 3′ splice junction stem–loops boxed in bold and their nucleotide position indicated above (refer to Figure 1A and B). In addition, each loop residue has been assigned a (+) or (−) number relative to the Ire1p(k+t) cleavage site indicated by an arrow. The sequence of the various point mutants is indicated next to each loop residue. Of these, those which are boxed indicate the equivalent residue at that position in the 5′ loop (refer to Figure 1). Indicated below these sequences is the extent to which each was cleaved by Ire1p(k+t). Values are reported as the fraction (mutant cleavage/wild-type cleavage) where: (−) no cleavage; (+/−) 0–10%; (+) 10–60%; (++) 60–120%; (+++) ≥120%. Ire1p(k+t) reactions were performed as described in Figure 2.

occurred at about the same level as found for stem–loops with the authentic *HAC1*$^u$ mRNA sequence (lanes 3 and 4). We next changed all Cs of the stem to As. This drastic change, predicted to disrupt the stem completely, abolished cleavage (Figure 2C, lanes 7 and 8). Thus we conclude that an intact stem is important, but that, in contrast to the loop, the nucleotide sequence of the residues that form the stem is irrelevant for cleavage by Ire1p(k+t). Taken together, these results show that the stem–loop minisubstrates represent the minimal RNA element that is required and sufficient for Ire1p(k+t) cleavage.

### *Ire1p(k+t) cleavage generates 2′,3′-cyclic phosphodiester termini*

Having confirmed that Ire1p(k+t) processes our minisubstrates with specificity indistinguishable from *HAC1*$^u$ mRNA, we used these simplified substrates to explore the chemistry of the Ire1p(k+t) cleavage reaction. To this end, we employed nearest neighbor analysis to determine the nature of the RNA termini produced upon Ire1p(k+t) cleavage. Stem–loop RNAs corresponding to either the 5′ (Figure 4A) or 3′ splice junction (Figure 4B) were labeled by incorporation of [α-$^{32}$P]CTP, [α-$^{32}$P]GTP or [α-$^{32}$P]ATP as indicated above the lanes. Digestion of the transcription products with P1 nuclease (which cleaves RNA to leave 5′-phosphate and 3′-OH groups) generated the expected 5′-monophosphorylated nucleotides as shown by thin layer chromatography (TLC) (Figure 4A and B, lanes 1–3), confirming that the labeled stem–loop RNAs contained labeled phosphate groups only in the appropriate positions.

In contrast, a new labeled nucleotide species was observed on thin layer plates when the transcription products were first cleaved with Ire1p(k+t) and the isolated 5′ fragment was then digested with P1 nuclease (‘pG>p’, Figure 4A and B, lanes 4–6). This spot comigrated with synthetic 5′-phosphate-guanosine-2′,3′-cyclic phosphate (here abbreviated pG>p), indicating that Ire1p(k+t) cleaves RNA to leave a 2′,3′-phosphate (note that nuclease P1 does not cleave at 2′,3′-cyclic phosphate groups). Consistent with this notion, labeled pG>p was only observed when the 5′ stem–loop RNA was labeled with [α-$^{32}$P]GTP (producing [$^{32}$P]G>p) or [α-$^{32}$P]CTP (producing pG>[$^{32}$P]) (Figure 4A, lanes 4 and 5), or when the 3′ stem–loop RNA was labeled with [α-$^{32}$P]GTP or [α-$^{32}$P]ATP (Figure 4B, lanes 5 and 6). Thus we conclude that Ire1p(k+t) cleaves the stem–loop RNAs at the 3′ side of the invariant G found in the third position of the loop. Furthermore, unlike any other mRNA splicing intermediate, Ire1p(k+t) produces a 2′,3′-cyclic phosphate group at its cleavage site. The reaction is more reminiscent of that of tRNA endonuclease which cleaves pre-tRNAs to produce 2′,3′-cyclic phosphate termini (Peebles *et al.*, 1983).

Since there is no precedent for an mRNA splicing intermediate with 2′,3′-cyclic phosphate termini, we decided to confirm our observations by using a series of sequential enzymatic modifications to simultaneously analyze the termini of the 3′ and 5′ cleavage fragments. We took advantage of the fact that the fragments produced upon Ire1p(k+t) cleavage of the stem–loop RNAs were small enough so that charge differences due to the presence or absence of terminal phosphate groups altered their electrophoretic mobilities in polyacrylamide gels. As shown in the diagram in Figure 5A, we first used calf intestinal phosphatase (CIP) to remove non-cyclic terminal phosphates. This resulted in a decreased mobility of the 5′ fragment, consistent with the predicted loss of its 5′-triphosphate group, and no mobility change in the 3′ fragment (Figure 5B, lane 2). Next, we treated the fragments produced thus with T4 polynucleotide kinase (T4 PNK) which hydrolyzes 2′,3′-cyclic terminal phosphates. Loss of the additional phosphate group led to a further reduction of the mobility of the 5′ fragment (Figure 5B, lane 3). This provided independent confirmation of the results presented in Figure 4, by demonstrating that a cyclic phosphate was produced at the 3′ end of the 5′ fragment, since a non-cyclic phosphate would have already been removed by CIP in the preceding step. Finally, we used treatment with T4 PNK in the presence of ATP to phosphorylate free 5′-OH groups. This led to an increased mobility of both the 5′ and 3′ fragment (Figure 5B, lane 4), indicating that the Ire1p(k+t)-produced 3′ fragment terminates in a 5′-OH group that can be phosphorylated.

The assignment of the Ire1p(k+t) cleavage site on the 3′ side of the invariant G in the loop was in conflict with an earlier study from our laboratory. In particular, primer extension mapped the Ire1p cleavage sites of *HAC1*$^u$ mRNA to the 5′ side of the invariant G (Sidrauski and Walter, 1997). This raised the possibility that cleavage of the minisubstrates and that of authentic *HAC1*$^u$ mRNA

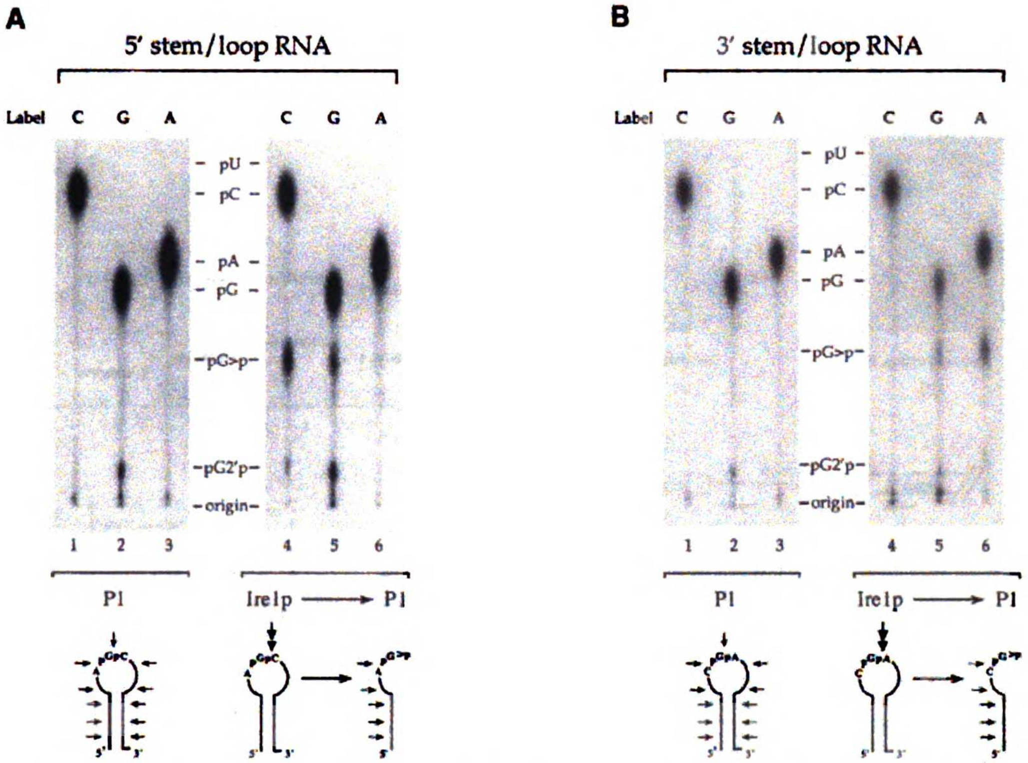


**Fig. 4.** Formation of 3′ terminal guanosine 2′,3′-cyclic phosphate after Ire1p(k+t) cleavage of stem–loop RNA. 5′ stem–loop RNA (**A**) or 3′ stem–loop RNA (**B**) was labeled with [α-$^{32}$P]CTP (lanes 1 and 4), [α-$^{32}$P]GTP (lanes 2 and 5) or [α-$^{32}$P]ATP (lanes 3 and 6). The labeled RNAs were then digested with P1 nuclease (lanes 1–3). Alternatively, they were first digested (40 000 to 60 000 c.p.m. of each stem–loop RNA) with Ire1p(k+t). The 5′ cleavage product was then isolated by preparative gel electrophoresis and digested with P1 nuclease (lanes 4–6). The final digestion products were separated by TLC on PEI cellulose plates and visualized by autoradiography. The position of marker nucleotides on the thin layer plates are indicated, where pG>p refers to a G bearing a 5′ phosphate and a 2′,3′-cyclic phosphate. Digestion of the relevant portions of the stem–loop structures is indicated schematically below each panel. The ratios of labeled pG>p relative to labeled pN were determined and are in good agreement with the expected values. (A) Lanes 4 and 5: pC/pG>p expected = 0.25, actual = 0.26; pG/pG>p expected =0.13, actual = 0.18. (B) Lanes 5 and 6: pG/pG>p expected = 0.17, actual = 0.21; pA/pG>p expected = 0.25, actual = 0.26. Note also that a variant of the wild-type 5′ stem–loop was used in this analysis (hactng-39). In this stem–loop, a GC dinucleotide pair in the stem was changed, leaving the only GC dinucleotide in the loop. The only GA dinucleotide in the wild-type 3′ stem–loop (hactng-10) was in the loop.

might, in fact, be different. The previous studies, however, used oligonucleotide primers distant from the cleavage site, making it possible that the cleavage sites were misassigned. To address this discrepancy and to verify that the cleavage sites in the minisubstrates and intact *HAC1*$^u$ mRNA were, in fact, identical, we repeated primer extension analysis on Ire1p(k+t)-cleaved *HAC1*$^u$ 508 RNA (the same construct used previously) but using primers located closer to the cleavage sites (Figure 6). The analysis shows unambiguously that Ire1p(k+t) cleaves the 5′ and 3′ splice junctions of *HAC1*$^u$ RNA at the 3′ side of the invariant Gs. These are the same positions at which Ire1p(k+t) cleaves the 5′ and 3′ stem–loop RNAs, thereby affirming the use of the minisubstrates to characterize the mechanism of the Ire1p(k+t)-catalyzed cleavage reaction.

### *Additional elements in HAC1$^U$ mRNA contribute to Ire1p(k+t) cleavage efficiency*

Although Ire1p(k+t) accurately cleaves the stem–loop substrates, additional structural and/or sequence elements present in *HAC1*$^u$ RNA are likely to contribute significantly to the efficiency of the reaction. To compare the cleavage efficiencies of the different substrates, we analyzed time courses of *HAC1*$^u$ RNA and stem–loop RNA Ire1p(k+t) digestions performed under identical conditions. Figure 7A shows a time course of *HAC1*$^u$ RNA digestion. We note that the substrate RNA disappears rapidly with a half time of ~2 min (Figure 7A, filled diamonds). Digestion intermediates appear transiently in a manner reminiscent of classical pulse–chase kinetics (Figure 7A, open circles and open squares), and the liberation of 5′ and 3′ exons (crossed circles and crossed squares) occurs rapidly and at comparable rates. In contrast, liberation of the intron (open triangles) follows a short lag time, concordant with the fact that two sequential cleavage events are required for its production. Taken together, these kinetics are consistent with a reaction in which both splice sites are cleaved independently and in random order (Sidrauski and Walter, 1997; Kawahara *et al.*, 1998).

In contrast, the smaller stem–loop RNAs were cleaved at significantly reduced rates. After 8 h of incubation under identical reaction conditions, <25% of either stem–loop RNA was cleaved (Figure 7B and C), whereas cleavage of *HAC1*$^u$ RNA was virtually complete within 16 min. Incubation of both the 5′ and 3′ stem–loops

UCSF LIBRARY

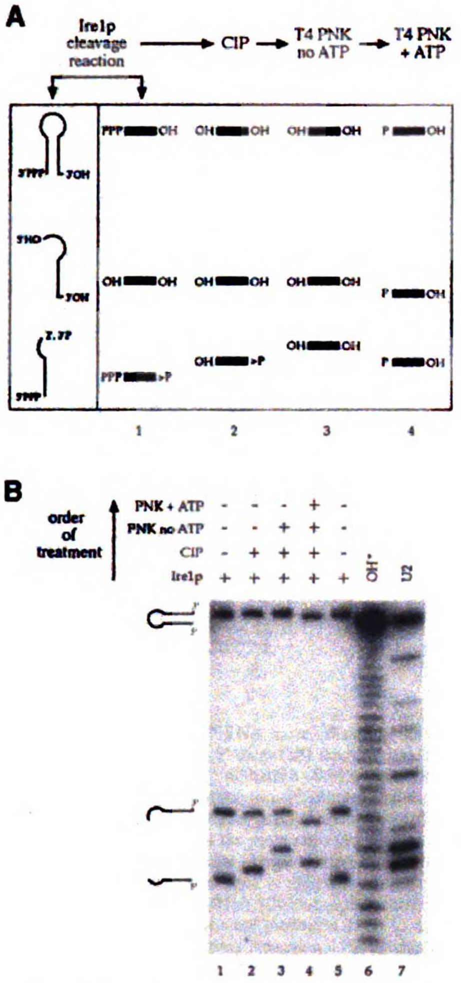


**Fig. 5.** Characterization of termini produced after Ire1p(k+t) cleavage of stem–loop RNA by sequential enzymatic modification. (A) A schematic of the cleavage products generated upon Ire1p(k+t) cleavage and subsequent enzymatic modification of stem–loop RNA is shown. Different gel mobilities of the short fragments result from charge differences due to terminal phosphate groups as indicated. '>P' represents a 2',3'-cyclic phosphate. Lanes 1–4 correspond to lanes 1–4 in (B). (B) Ire1p(k+t) cleavage of internally labeled 3' stem–loop RNA was performed as described in Figure 2. Lanes 1 and 5 show the products of an Ire1p(k+t) cleavage reactions of 3' stem–loop RNA. The products were then treated sequentially with CIP (lane 2), with T4 PNK in the absence of ATP (lane 3), and with T4 PNK in the presence of ATP (lane 4). Markers were produced by alkaline hydrolysis (lane 6) or nuclease U2 digestion (lane 7) of 5' end-labeled wild-type 3' stem–loop RNA.

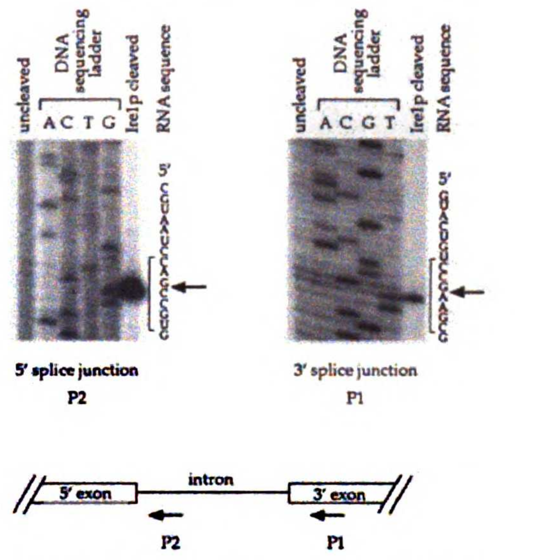


**Fig. 6.** Ire1p(k+t) cleaves $HAC1^u$ RNA at the same positions as it cleaves the 5' and 3' stem–loops. Primer extension analysis was performed on both undigested and Ire1p(k+t)-digested $HAC1^u$ 508 RNA using 5' end-labeled DNA primers. $HAC1^u$ 508 RNA consists of a truncated 5' exon (181 nucleotides), an intact $HAC1^u$ intron (252 nucleotides), and a truncated 3' exon (75 nucleotides), as previously described (Sidrauski and Walter, 1997). Extensions on uncleaved RNA served as controls for natural pausing during the extension reaction. The Ire1p(k+t) cleavage site is indicated by an arrow along side the RNA sequence corresponding to the DNA ladder; the residues in the 7-membered loop are marked with a bracket.

together in the same reaction did not stimulate the cleavage of either stem–loop, nor did heating and quick cooling of the stem–loop substrates prior to their addition to the reaction (data not shown). We therefore conclude that structural elements of $HAC1^u$ RNA, in addition to the stem–loop structures characterized here, must contribute significantly to the efficiency of cleavage by Ire1p(k+t), possibly by enhancing substrate recognition by Ire1p(k+t).

### *Ire1p(k+t) liberates base paired exons*

The predicted secondary structure of the $HAC1^u$ RNA intron (Figure 1A) suggested that the exon sequences flanking the intron are also base paired and hence may contribute to folding the intron–exon junctions into a more complex RNA domain. A prediction of the proposed structure is that the two exons remain associated after removal of the intron by Ire1p-mediated cleavage. To test this possibility directly, we fractionated Ire1p(k+t)-generated $HAC1^u$ RNA cleavage products on a non-denaturing gel (Figure 8A). Individual bands were excised from the non-denaturing gel and re-electrophoresed under denaturing conditions to determine their composition (Figure 8B). Only two major bands were obtained when Ire1p(k+t)-digested $HAC1^u$ RNA was fractionated under native conditions (Figure 8A, lane 1; labeled *a* and *b*). If the cleavage reaction products were first denatured by boiling, band *a* disappeared and two new, faster migrating bands appeared in its place (Figure 8A, lane 2; labeled *d* and *e*). As apparent from the analysis shown in Figure 8B, band *a* contains both exons which upon boiling become separated into band *d* containing the 5' exon and band *e* containing the 3' exon. Thus, upon Ire1p(k+t) digestion, the 5' and 3' exons remain non-covalently associated and comigrate in a native gel, whereas the intron is released. In the context of the $HAC1^u$ RNA splicing reaction, this association may be advantageous in that it holds the two exons in position for efficient and specific ligation by tRNA ligase.

### *The phosphate linking the two exons of spliced HAC1 mRNA is derived from nucleotide triphosphate*

The data presented so far show that, in many respects, the mechanism of $HAC1^u$ mRNA splicing follows the

UCSF LIBRARY

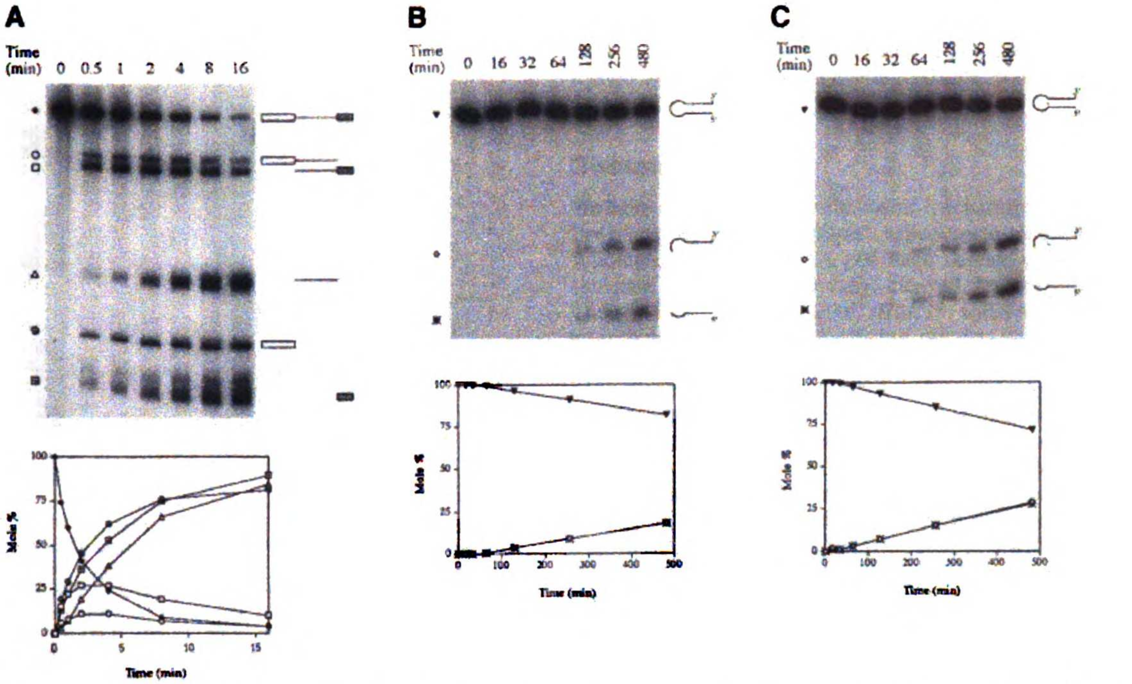


**Fig. 7.** Ire1p(k+t) cleaves *HAC1*$^u$ RNA more efficiently than stem–loop RNAs. (**A**) A time course of *HAC1*$^u$ 600 RNA cleavage is shown. *HAC1*$^u$ 600 RNA consists of a truncated 5′ exon (181 nucleotides), an intact *HAC1*$^u$ intron (252 nucleotides), and a truncated 3′ exon (167 nucleotides). A cocktail of 140 000 c.p.m. *HAC1*$^u$ 600 RNA (0.07 pM), cleavage buffer, and 1.75 μg Ire1p(k+t) in a volume of 140 μl was incubated at 30°C. Samples (20 μl) were removed at the times indicated and the reactions were stopped by addition of 20 volumes phenol/chloroform and 20 volumes of stop buffer. Reactions were ethanol-precipitated, electrophoresed through a 5% denaturing polyacrylamide gel, visualized by autoradiography and quantitated. Percent conversion is the mole ratio (amount of fragment produced)/(amount of fragment expected if 100% cleavage occurred) adjusted for the number of labeled phosphate groups carried by each fragment. Symbols next to the autoradiogram correspond to the symbols used for each fragment in the graph below. The icons to the right represent the fragments produced in the Ire1p(k+t) cleavage reaction. A time course of 5′ stem–loop (**B**) and 3′ stem–loop RNA (**C**) cleavage is shown. Cocktails of 44 000 c.p.m. of stem–loop RNAs (0.22 pM), cleavage buffer and 5.5 μg Ire1p(k+t) in a volume of 440 μl were incubated at 30°C. Samples (40 μl) were removed at the times indicated, rapidly stopped and precipitated as described above, and electrophoresed through a 15% denaturing polyacrylamide gel. The mole ratio of Ire1p(k+t) to RNA substrate is identical in all panels. Visualization and quantitation as well as symbols and icons are as in (A). Note the significantly longer time scales of these experiments. Also note that the amount of Ire1p(k+t) used in the experiments shown in this figure was less than the amount used in the experiments shown in Figures 2 and 5 and thus accounts for the reduced cleavage of the stem–loop substrates at comparable time points.

paradigms established for pre-tRNA splicing (Westaway and Abelson, 1995; Abelson *et al.*, 1998). During pre-tRNA splicing, tRNA ligase phosphorylates the 5′ end of the 3′ exon in a step prior to exon ligation. As a consequence, the phosphate group in the phosphodiester bond linking the 5′ and 3′ exons in spliced tRNAs originates from the γ-phosphate of the GTP or ATP used by tRNA ligase and not from the tRNA itself (Belford *et al.*, 1993). To determine if this is also true for *HAC1*$^u$ RNA splicing, we added purified tRNA ligase to Ire1p(k+t)-cleaved *HAC1*$^u$ RNA, thus reconstituting a complete splicing *in vitro* reaction as previously described (Sidrauski and Walter, 1997). As shown in Figure 9A, ligation occurred very efficiently in our assay system; tRNA ligase addition to Ire1p(k+t)-cleaved *HAC1*$^u$ RNA resulted in a virtually quantitative conversion of the exons (compare lanes 3 and 4) to the ligated product which comigrated with a marker transcript lacking the intron (Figure 8A, lane 1). This reaction was then repeated using unlabeled input RNA and [γ-$^{32}$P]GTP (Figure 9A, lane 5). Under these conditions, any incorporated radioactivity was exclusively derived from the γ-position of the added nucleotide triphosphate. Three prominent, labeled bands were obtained under these conditions (Figure 9A, lane 5), corresponding to the ligated exons, the intron and residual 3′ exon. Thus, as expected, tRNA ligase phosphorylated the two RNA fragments containing free 5′-OH groups (intron and 3′ exon), but not the 5′ exon as it already contains a 5′-triphosphate group. Importantly, these data also suggest that phosphorylated 3′ exon was incorporated into ligated RNA, as previously described for pre-tRNA splicing (Belford *et al.*, 1993).

During pre-tRNA splicing, in the last step prior to exon ligation, tRNA ligase adenylates itself using ATP. Ligase then covalently transfers the AMP to the 5′ phosphate of the 3′ exon which results in the formation of a high energy 5′–5′ phosphoanhydride bond. Upon exon ligation, the AMP is released from the RNA (Greer *et al.*, 1983; Xu *et al.*, 1990). To determine if this is also true for *HAC1*$^u$ RNA splicing, we incubated unlabeled input RNA with Ire1p(k+t), tRNA ligase and [α-$^{32}$P]ATP (Figure 9A, lanes 6 and 7). Under these conditions the 3′ exon and the intron were labeled. In addition, no label was incorporated into the fully spliced RNA. Both these

UCSF LIBRARY

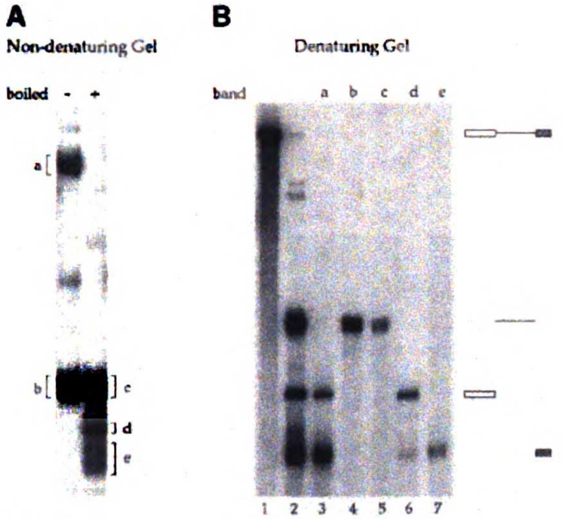


**Fig. 8.** *HAC1*$^{u}$ RNA exons remain base paired following Ire1p(k+t) cleavage. (**A**) An aliquot of *HAC1*$^{u}$ 600 RNA (40 000 c.p.m.) was incubated with 1 µg Ire1p(k+t) in cleavage buffer in 40 µl for 45 min at 30°C. As indicated, reactions were either boiled or not and electrophoresed under non-denaturing conditions through a 5% polyacrylamide gel in TBE and visualized by autoradiography. The bands labeled a, b, c, d and e were excised from the gel. The RNA was eluted from the gel slices overnight. (**B**) Eluted a, b, c, d and e RNAs were resuspended in denaturing loading buffer, boiled and electrophoresed through a 5% denaturing polyacrylamide gel. The marker lanes 1 and 2 carry uncleaved *HAC1*$^{u}$ 600 RNA and Ire1p(k+t) cleaved *HAC1*$^{u}$ RNA, respectively. Icons to the right of the figure indicate the fragments produced in the Ire1p(k+t) cleavage reaction.

observations suggest that *HAC1* exons are ligated by tRNA ligase utilizing the same mechanism with which tRNA exons are joined.

To show that the label in the spliced RNA was indeed incorporated at the splice junction, we took advantage of the fact that during pre-tRNA splicing, tRNA ligase opens up the 2′,3′-cyclic phosphate at the 3′ terminus of the 5′ exon and leaves this phosphate attached at the 2′ position of the splice junction. For our purposes, the presence of the 2′ phosphate would provide a convenient marker for the splice junction as it would render this single phosphodiester bond in the spliced RNA resistant to cleavage by P1 nuclease. To determine whether the labeled γ-phosphate group donated by GTP during the splicing reaction was indeed incorporated at the splice junction, we purified the spliced product from a reaction as shown in Figure 9A, lane 5, and subjected it to digestion with P1 nuclease. TLC of the digestion products revealed that the labeled phosphate was quantitatively recovered in a slow migrating spot (Figure 9B, lane 4) with a mobility identical to that of the marker nucleotide 5′pG(2′p)3′p5′A that was produced in a control reaction by digestion of an *in vitro* spliced pre-tRNA substrate (Figure 9B, lane 3). Consistent with the assigned structure, treatment of either sample with snake venom phosphodiesterase (SVP) (which cleaves phosphodiester bonds even in the presence of a 2′-phosphate) quantitatively converted the labeled dinucleotides to pA. Furthermore, in a related approach, digestion with P1 nuclease liberated labeled pA if the spliced RNA was first treated with CIP to remove the 2′-phosphate at the splice junction (McCraith and Phizicky, 1990; data not shown). Thus by the criteria addressed here, the mechanism by which *HAC1*$^{u}$ RNA exons become ligated by tRNA ligase is indistinguishable from that used during tRNA splicing.

## Discussion

We have characterized the enzymatic mechanism by which Ire1p carries out the unique, spliceosome-independent splicing of *HAC1*$^{u}$ mRNA and have defined a minimal substrate which is accurately cleaved by Ire1p(k+t). Though the mechanism of *HAC1*$^{u}$ mRNA splicing resembles that of pre-tRNA splicing, our data also revealed differences between the two pathways. In particular, we have shown that Ire1p and tRNA endonuclease recognize fundamentally different structural features in their respective substrate RNAs. In some ways, this is not surprising given that these two nucleases lack any recognizable structural similarity. Indeed, the most similar protein to Ire1p is RNase L, a ribonuclease which non-specifically cleaves RNA in cells infected with double-stranded RNA viruses and for which no role in pre-mRNA processing is suspected (Bork and Sander, 1993; Hassel *et al.*, 1993; Dong *et al.*, 1994; Dong and Silverman, 1997; Zhou *et al.*, 1997). Thus, though splicing of *HAC1*$^{u}$ mRNA is unlike spliceosome-mediated pre-mRNA splicing, neither is it completely like pre-tRNA splicing. *HAC1*$^{u}$ mRNA splicing thus defines a unique RNA processing event.

Our experiments have demonstrated that, mechanistically, splicing of *HAC1*$^{u}$ mRNA resembles that of pre-tRNA splicing. In particular, as summarized in Figure 10, the chemistry of endonucleolytic cleavage by Ire1p leaves 2′,3′-cyclic phosphates, the two exons remain associated by base pairing after cleavage has occurred, and exon ligation by tRNA ligase follows the same chemical steps previously characterized for pre-tRNA splicing. This mechanism of processing is unprecedented for a messenger RNA.

By defining a minimal RNA element that is sufficient and required for cleavage by Ire1p(k+t), we have shown that the structural features which Ire1p recognizes in its substrate differ significantly from those which are recognized by tRNA endonuclease. Eukaryotic tRNA endonuclease, for example, primarily recognizes the folded tertiary structure of the pre-tRNA as well as local structures at the intron–exon boundaries and pays little heed to the nucleotide sequences at or near the splice sites (Mattoccia *et al.*, 1988; Reyes and Abelson, 1988; Fabbri *et al.*, 1998). Indeed, changing the length of the stem that separates the cleavage sites from the knee in the tRNA tertiary fold shifts the cleavage sites correspondingly (Reyes and Abelson, 1988). This indicates that eukaryotic tRNA endonuclease can use distance measurements from a substrate binding site to its active site to specify the cleavage sites on its substrate RNA. Recently, a small, doubly bulged stem containing both cleavage sites of pre-tRNA was shown to be a suitable substrate for tRNA endonuclease (Fabbri *et al.*, 1998), indicating that, as for *HAC1*$^{u}$ mRNA cleavage by Ire1p, local structural features in pre-tRNA can be sufficient to confer cleavage specificity. Archaeal tRNA endonuclease recognizes a symmetric, bulged stem structure, also with little regard for the

UCSF LIBRARY

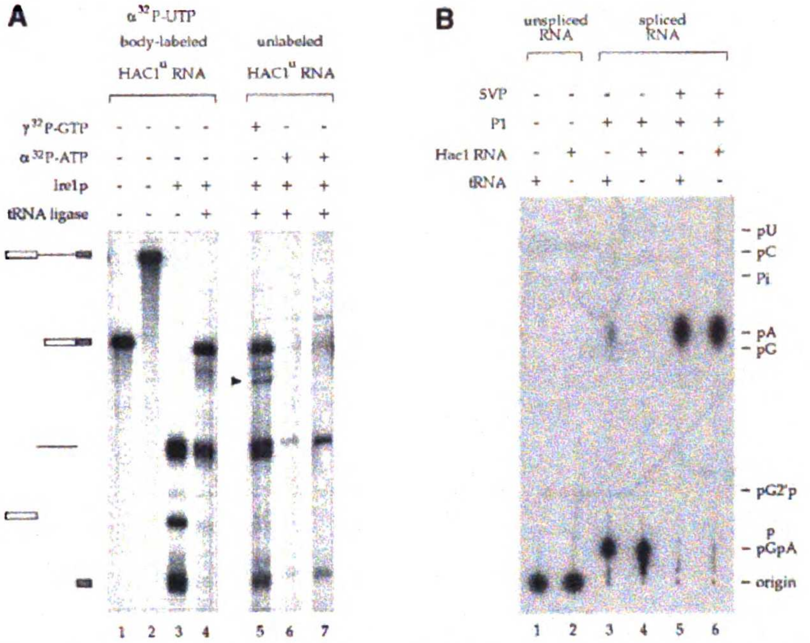


**Fig. 9.** The phosphate group at the splice junction derives from nucleotide triphosphate. **(A)** Unlabeled *HAC1*$^{u}$ 600 RNA (1 pM), cleavage buffer and 0.5 µg Ire1p(k+t) were incubated at 30°C for 45 min. Then 0.3 µg purified tRNA ligase was added along with either 1 mM ATP, 65 µCi [$\gamma$-$^{32}$P]GTP (lane 5) or 200 µCi [$\alpha$-$^{32}$P]ATP, 1 mM GTP (lanes 6 and 7), and the reaction incubated for another 30 min. Lanes 6 and 7 only differ in the length of time they were exposed to film. The remaining lanes carry markers derived from internally labeled *HAC1*$^{u}$ RNA. *HAC1*$^{u}$ 600 RNA was incubated in cleavage buffer with Ire1p(k+t) (lane 3), or Ire1p(k+t), tRNA ligase, ATP and GTP (lane 4). These two lanes provided markers for the Ire1p(k+t) cleavage reaction, and the products of the complete Ire1p(k+t)/tRNA ligase splicing reaction, respectively. Markers for uncleaved *HAC1*$^{u}$ 600 RNA (lane 2) as well as *HAC1*$^{i}$ 348 RNA corresponding to spliced *HAC1*$^{u}$ 600 RNA (lane 1) are shown. Note that the labeled intron in lane 5 splits into a doublet; it is likely that these two forms correspond to a 5′-phosphorylated and a 5′-phosphorylated and also adenylated RNA, both known tRNA ligase reaction products. The nature of the band in lane 5 indicated by a triangle is unknown; one possibility is that it represents a circularized or concatenated intron. **(B)** *HAC1*$^{u}$ 600 RNA spliced in the presence of [$\gamma$-$^{32}$P]GTP as in (A) was gel-purified from a 5% denaturing polyacrylamide gel, and subjected to nuclease P1 digestion (lane 4), followed by digestion with SVP (lane 6). Digestion products were chromatographed on PEI cellulose thin layer plates and visualized by autoradiography. Nuclease P1 digestion of *in vitro* spliced pre-tRNA$^{Phe}$ (lane 3) provided a marker for the 5′pG(2′p)3′,5′pA splice junction dinucleotide (see Materials and methods) and SVP digestion provided a positive control for SVP digestion of spliced *HAC1*$^{u}$ RNA. The position of other nucleotide markers on the thin layer plate is shown.

sequence surrounding the cleavage sites (Thompson and Daniels, 1990; Lykke-Andersen and Garrett, 1994, 1997; Kleman-Leyer *et al.*, 1997). In contrast, we show that cleavage by Ire1p(k+t) requires a well defined 7-nucleotide loop that must be closed by a stem. Furthermore, unlike tRNA endonucleases, cleavage by Ire1p(k+t) is exquisitely sensitive to the particular nucleotide sequence flanking the cleavage site (Figures 2 and 3) (Sidrauski and Walter, 1997; Kawahara *et al.*, 1998). With the exception of loop residue –2, changes at the other six loop positions can drastically reduce cleavage by Ire1p(k+t). The nucleotides at these sensitive positions may provide critical contacts for Ire1p and/or maintain specific structural arrangements within the loop necessary for Ire1p recognition. Thus, our data define a unique 'signature motif' for Ire1p substrates. This structural information will be invaluable for future studies whose aim is identification of other substrates for this unique RNA processing pathway in yeast and other organisms.

Though the primary sequence of the minisubstrate loops is important for Ire1p(k+t) recognition and cleavage, our data suggest that other sequences within *HAC1*$^{u}$ mRNA significantly enhance recognition by Ire1p. In particular, *HAC1*$^{u}$ RNA substrates carrying the entire intron plus flanking exon sequences were cleaved significantly faster than the minisubstrate stem–loop RNAs (Figure 7). Ire1p may have a higher affinity for these substrates relative to the minisubstrates, possibly because it interacts directly with additional regions of *HAC1*$^{u}$ mRNA that fall outside the stem–loop structures. An alternative albeit not mutually exclusive explanation is that, as previously suggested, active Ire1p is a dimer or higher order oligomer that binds *HAC1*$^{u}$ mRNA so that one monomer interacts with the 5′ splice site and the other interacts with the 3′ splice site (Sidrauski and Walter, 1997). Indeed Ire1p oligomerizes upon activation *in vivo* (Shamu and Walter, 1996), and addition of ADP to *in vitro* reactions stimulates Ire1p(k+t) dimerization/oligomerization (Niwa and Walter, unpublished results) and its endoribonuclease activity (Sidrauski and Walter, 1997). For this reason, all reactions described in this paper were performed under optimal ADP-stimulating conditions. A gain in affinity could result from cooperative binding of Ire1p to the two splice sites and thus lead to increased cleavage. Folding of the intron and/or base pairing interactions between the two exons may significantly contribute to cooperativity by positioning the two splice sites in an optimal orientation with respect to each other.

One of the unique consequences of the Ire1p/tRNA ligase-mediated splicing pathway is that the splice junction

UCSF LIBRARY

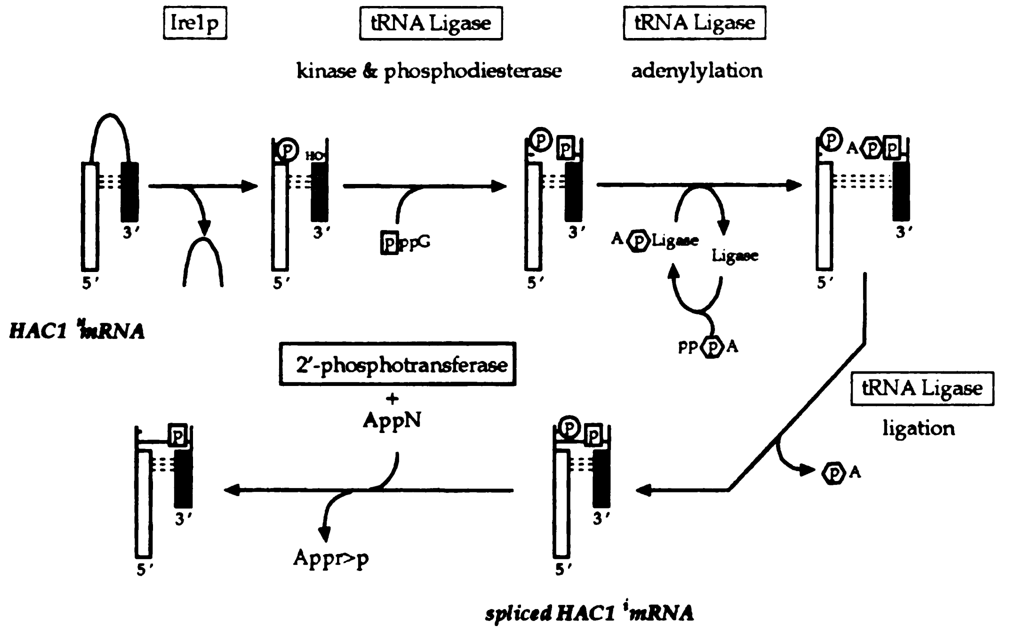


**Fig. 10.** Model of *HAC1*$^u$ mRNA splicing. Ire1p initiates splicing by cleaving *HAC1*$^u$ mRNA 3′ of the conserved G at position –1 in both splice site loops (Figures 1 and 3), releasing the intron and generating a 2′,3′-cyclic phosphate at the 3′ end of the 5′ exon and a free 5′-OH group at the 5′ end of the 3′ exon, respectively. The liberated exons remain base paired, thus holding the appropriate mRNA termini in spatial proximity. tRNA ligase acts upon the paired exons, phosphorylating the 5′ terminus of the 3′ exon with phosphate derived from the γ-position of GTP. This GTP-derived phosphate group ultimately links the two exons together in the spliced *HAC1*$^i$ mRNA. The phosphodiesterase activity of tRNA ligase opens the 2′,3′-cyclic phosphate to the 2′ position. tRNA ligase next adenylates the 5′ terminus of the 3′ exon, forming a high energy A5′pp5′A phosphoanhydride bond and then, using the energy stored in this bond, joins the two exons. As is the case for pre-tRNA splicing, it is likely that tRNA ligase first adenylates itself and then transfers its AMP to the 3′ exon. Upon ligation of the exons, the AMP is released. The splice junction of the newly spliced *HAC1*$^i$ mRNA carries a 2′-phosphate derived from the 5′ splice site. As for pre-tRNA splicing, perhaps NAD-dependent 2′-phosphotransferase transfers this phosphate to NAD to produce ADP-ribose 1″-2″ cyclic phosphate (Appr>p) and *HAC1*$^i$ mRNA free of the splice junction 2′-phosphate.

of the product *HAC1*$^i$ mRNA is initially tagged by a 2′-phosphate group. This tag marks the splice junction of all newly ligated *HAC1*$^i$ mRNA molecules until it is presumably removed. In yeast, the product of an essential gene, a NAD-dependent 2′-phosphotransferase transfers 2′-phosphate groups from spliced tRNAs to NAD, producing ADP-ribose 1″-2″cyclic phosphate (McCraith and Phizicky, 1990; Culver *et al.*, 1993, 1997). The same phosphotransferase may act upon newly ligated *HAC1*$^i$ mRNA to remove its splice junction 2′-phosphate (Spinelli *et al.*, 1997). It is not known whether the unique 2′-phosphate tag or the unique cyclic phosphate-containing metabolite that results from its removal have any physiological function in the cell (Culver *et al.*, 1997). It is conceivable, however, that induction of the UPR, and the resulting splicing of *HAC1*$^u$ mRNA might generate a spike of the unusual ADP-ribose 1″-2″ cyclic phosphate over the level normally produced as a consequence of pre-tRNA splicing. Thus it is tempting to speculate that cells might utilize such a signal, possibly in ways analogous to signals transmitted via other small molecules containing cyclic phosphate groups, and integrate it in as yet unknown ways into the cell's response to protein misfolding in the ER.

## Materials and methods

### Plasmid constructs and recombinant protein expression

The cytoplasmic portion of *S.cerevisiae* Ire1p containing its kinase and C-terminal tail domains, Ire1p(k+t), was expressed in and purified from *Escherichia coli* as previously described (Sidrauski and Walter, 1997). In this study, we used an expression vector which fuses a PreScission Protease (Pharmacia, Uppsala, Sweden) cleavage site between the Ire1p(k+t) and glutathione *S*-transferase (GST) domains of the recombinant polypeptide. In brief, *E.coli* strain DH5α transformed with plasmid pCF210 was grown in liquid culture at 37°C to an OD of 0.5. Expression of recombinant GST–Ire1p(k+t) was induced with isopropyl-β-D-galactopyranoside (IPTG) added to a final concentration of 0.1 mM. Cells were harvested and ruptured with a Microfluidizer (Microfluidics Co., Newton, MA). GST–Ire1p(k+t) was captured by batch binding to glutathione–Sepharose beads (Pharmacia, Uppsala, Sweden), and Ire1p(k+t) was liberated by digestion with PreScission Protease.

The gene encoding tRNA ligase was amplified from the *S.cerevisiae* genome by PCR and the sequence of the amplified gene was verified. The gene was cloned into the expression vector pGEX-6P-2 to create a GST–tRNA ligase fusion with a PreScission Protease cleavage site between GST and tRNA ligase. Expression of the recombinant protein from this plasmid (pSD103) was as for Ire1p(k+t) (see above). Harvested cells were ruptured by sonication and expressed GST–tRNA ligase was bound to glutathione–Sepharose beads. The beads were subsequently washed to remove non-recombinant protein and GST–tRNA ligase was eluted from the beads with 20 mM glutathione, pH 7.5. The eluted fusion protein was dialyzed to remove glutathione and cleaved with PreScission Protease to liberate tRNA ligase. In the final step, PreScission Protease and free GST were removed by incubating the cleaved material with glutathione–Sepharose beads, thus yielding pure tRNA ligase.

### In vitro RNA transcription

*In vitro* transcriptions of *HAC1*$^u$ 600, *HAC1*$^u$ 508, *HAC*$^i$ 348 [see Sidrauski and Walter (1997) for details of the plasmids encoding these RNAs] and pre-tRNA$^{Phe}$ (from plasmid pUC12T7, a gift from C.Greer, University of California, Irvine) were carried out as follows. Reactions (20 μl) containing 1 mM each of ATP, CTP, GTP, 0.1 mM of UTP, 25 μCi of [α-$^{32}$P]UTP (3000 Ci/mM; Amersham, Arlington Heights, IL), 1 μg linearized plasmid DNA, 40 U RNasin (Promega, Madison, WI), and 20 U T7 RNA polymerase (Boehringer Mannheim, Indianapolis, IN) were incubated at 37°C for 1.5 h. To generate unlabeled RNA, transcriptions were carried out in the presence of 1 mM UTP without [α-$^{32}$P]UTP.

Smaller RNAs were transcribed using single-stranded DNA oligonucleotides as templates to which the 18mer 5′TAATACGACTCACTATAG 'T7 promoter oligonucleotide' was annealed to create the double-stranded T7 RNA polymerase promoter (Milligan *et al.*, 1987). The following oligonucleotides were used: hactng-10 (encoding wild-type 3′ stem–loop RNA): 5′TGAGGTCAAACCTGACTGCGCTTCGGACAGTAC-AAGCTTGACCTATAGTGAGTCGTATTA; hactng-32 (encoding 3′ stem–loop RNA with GC stem): 5′GAGCGGCGCGGCGCTTCGGAC-CGCGCCGCTATAGTGAGTCGTATTA; hactng-33 (encoding 3′ stem–loop RNA with GA stem): 5′GATCTTCTCTTCGCTTCGGACCTCTC-CTCTATAGTGAGTCGTATTA; hactng-38 (encoding wild-type 5′ stem–loop RNA): 5′TGAGCCGGTCATCGTAATCACGGCTGGATT-ACGCCAACCGGCTATAGTGAGTCGTATTA, and hactng-39 (encoding 5′ stem–loop used in TLC analysis): 5′TGAGGGGGTCATCGTA-ATCACGGCTGGATTACGACAACCCCCTATAGTGAGTCGTATTA. Oligonucleotides containing the appropriate point mutations indicated in Figures 2B and 3 are modifications of hactng-10.

A solution containing 15 pM T7 promoter oligonucleotide and 0.25 pM template oligonucleotide was heated to 100°C for 3 min and immediately placed on ice. For the nearest neighbor analysis, RNA transcripts were internally labeled by incorporation of any one of three (A, C or G) [α-$^{32}$P]NTPs (3000 Ci/mM; Amersham Corporation, Arlington Heights, IL). In this case, all NTPs in the transcription reactions were at 1 mM, except the labeled NTP of which 300 μCi was added diluted with unlabeled NTP to bring the final concentration to 0.1 mM. To transcribe stem–loop RNAs internally labeled with [α-$^{32}$P]UTP (3000 Ci/mM; Amersham Corporation, Arlington Heights, IL), 1 mM each of ATP, CTP, GTP, 0.1 mM of UTP and 50 μCi [α-$^{32}$P]UTP were added to the reactions instead.

Except where noted, all transcribed RNAs were purified on denaturing polyacrylamide gels, excised from the gels and eluted from gel slices overnight at 4°C with 0.3 M NaOAc, pH 5.2, 10 mM Mg(OAc)$_2$/phenol/chloroform (1/0.5/0.5) followed by ethanol precipitation.

### *Ire1p(k+t) cleavage and splicing reactions*

Ire1p(k+t) cleavage of *HAC1*$^{u}$ 600 RNA, *HAC1*$^{u}$ 508 RNA, and stem-loop RNAs was carried out in cleavage buffer [20 mM HEPES pH 7.6, 50 mM KOAc, 1 mM Mg(OAc)$_2$, 1 mM DTT, 2 mM ADP, 40 U RNasin (Promega, Madison, WI)] at 30°C. The amount of Ire1p(k+t) and RNA in the reactions varied and is given in the figure legends or specific methods for each experiment. For splicing reactions with internally radiolabeled *HAC1*$^{u}$ 600 RNA, the RNA was first incubated with Ire1p(k+t) for 30 min at which point 1 mM ATP, 1 mM GTP and 0.3 μg tRNA ligase were added, and the reaction incubated for an additional 30 min at 30°C. Two different sets of splicing reactions were performed utilizing 1 pM unlabeled RNA each. In one, 65 μCi of [γ-$^{32}$P]GTP (5000 Ci/mM) (Amersham Corporation, Arlington Heights, IL) was added instead of GTP. In the other, 200 μCi [α-$^{32}$P]ATP (3000 Ci/mM) (Amersham Corporation, Arlington Heights, IL) was added instead of ATP. All reactions were terminated with 20 vol of stop buffer (50 mM NaOAc, pH 5.2, 1 mM EDTA, 0.1% SDS), extracted with phenol-chloroform, and ethanol-precipitated. Unless otherwise noted, reactions containing stem–loop RNAs and those containing the larger RNAs were analyzed on 15 and 5% denaturing polyacrylamide gels, respectively, and visualized by autoradiography of the gels.

### *Enzymatic modification of Ire1p(k+t) cleavage products*

A 100 μl Ire1p(k+t) cleavage reaction containing cleavage buffer (see above), 10 000 c.p.m. wild-type 3′ stem–loop RNA, and 5 μg Ire1p(k+t) was incubated at 30°C for 2 h. The reaction was stopped with stop buffer (see above) and phenol-chloroform extracted. Two fifths of the reaction was removed and set aside. The remaining sample was ethanol precipitated and treated with 30 U CIP (New England Biolabs, MA) for 1.5 h at 37°C. The reaction was phenol–chloroform extracted, and one-third of the sample was set aside. The remaining sample was ethanol precipitated and treated with 30 U T4 polynucleotide kinase (New England Biolabs, MA) in the absence of ATP for 1.5 h at 37°C followed by phenol-chloroform extraction. Half of the sample was set aside, and the remaining sample was ethanol precipitated and treated with 30 U T4 polynucleotide kinase plus 2 mM ATP for 1.5 h at 37°C followed by phenol-chloroform extraction and ethanol precipitation. For analysis, all samples were ethanol precipitated, fractionated on a 15% denaturing polyacrylamide gel and visualized by autoradiography.

### *TLC analyses*

Prior to digestion, the 5′ and 3′ stem–loop RNAs or the ligated products of the *HAC1*$^{u}$ 600 and pre-tRNA$^{Phe}$ splicing reactions were gel-purified and eluted overnight in water/phenol/chloroform (1/0.5/0.5) at 4°C. Eluted RNA was precipitated by extraction with N-butanol, washed with ethanol, and dried. P1 nuclease digests of 3′ and 5′ stem–loop RNAs were prepared by incubation of the RNA sample with 0.2 U P1 nuclease (Boehringer Mannheim, Indianapolis, IN) at 37°C for 30 min in a buffer that contained 0.5 μg tRNA and 20 mM NaOAc, pH 5.2. P1 digests of spliced *HAC1*$^{u}$ 600 RNA as well as spliced pre-tRNA$^{Phe}$ were the same but contained only 50 ng tRNA. For further digestion of the P1 nuclease products with SVP, the reaction mix was adjusted to 12.5 mM Tris pH 9.3 and 0.1 μg tRNA, and 0.025 U SVP (Worthington Biochemical Co., Freehold, NJ) were added. The reaction was incubated at 25°C for 40 min.

Spliced pre-tRNA$^{Phe}$ and *HAC1*$^{u}$ share the same splice junction sequence: G-junction-A (Valenzuela *et al.*, 1978; Reyes and Abelson, 1987; Sidrauski and Walter, 1997; Kawahara *et al.*, 1998). Because of this, spliced pre-tRNA$^{Phe}$ was used to generate 5′pG(2′p)3′p5′A marker dinucleotide for the TLC analysis of the *HAC1*$^{u}$ splice junction. To do so, pre-tRNA$^{Phe}$ cleavage reactions were carried out in tRNA endonuclease buffer (Greer *et al.*, 1987) for 30 min at 30°C. Unlabeled pre-tRNA$^{Phe}$ (1 pM) was first incubated with tRNA endonuclease for 30 min as described (Greer *et al.*, 1987), at which point 1 mM ATP, 0.3 μg tRNA ligase and 65 μCi of [γ-$^{32}$P]GTP (5000 Ci/mM) were added and the reaction incubated for an additional 20 min at 30°C. The resulting spliced tRNA$^{Phe}$ was gel purified. To generate labeled 5′pG(2′p)3′p5′A, purified tRNA$^{Phe}$ was digested with P1 nuclease as above. Further digestion with SVP as described above generated labeled pA as expected.

PEI cellulose thin layer plates (EM Science, Gibbstown, NJ) were developed with 1 M LiCl, and samples visualized by autoradiography. The radiolabeled markers 5′pN, 5′pN2′p, 5′pN3′p and pN2′3′(cyclic)p (where N = A, C, G or U) were prepared by phosphorylating N2′p, N3′p, and N2′3′(cyclic)p nucleotides (Sigma) in the presence of [γ-$^{32}$P]ATP and either wild-type T4 polynucleotide kinase (New England Biolabs, MA) or mutant T4 polynucleotide kinase (Boehringer Mannheim, Indianapolis, IN) lacking the 3′ phosphatase activity of the wild-type enzyme (Cameron and Uhlenbeck, 1977; Cameron *et al.*, 1978).

### *Primer extension*

The sequencing primers P1 (TGSP-3; 5′GAAGAAATCATTCAATTCA-AATGAATTC) and P2 (TGSP-1; 5′GCTAGTGTTCTTGTTCACTG) were used to map the 5′ and 3′ Ire1p(k+t) cleavage sites in *HAC1*$^{u}$ 508 RNA, respectively. Reactions contained 1 pM 5′ end-labeled primer, 10 ng Ire1p(k+t)-cleaved or uncleaved *HAC1*$^{u}$ 508 RNA, 20 mM NaCl, 15 mM HEPES pH 7.6. Reactions were heated to 100°C for 3 min and slowly cooled to 40°C. Next, 0.1 mM dNTPs and 3 U AMV reverse transcriptase (Boehringer Mannheim, Indianapolis, IN) were added, and the reactions incubated at 40°C for 30 min. Sequencing ladders were generated in the same manner, except that reactions also contained 0.1 mM of either ddATP, ddCTP, ddGTP or ddTTP. Samples were analyzed on 10% denaturing polyacrylamide gels and visualized by autoradiography.

### *Non-denaturing gel electrophoresis*

Ire1p(k+t) cleavage reactions were electrophoresed at 4°C through a 5% polyacrylamide gel in 1× TBE buffer. Bands were visualized by autoradiography, excised from the gel, and eluted overnight in 0.3 M NaOAc pH 5.2, 10 mM Mg(OAc)$_2$/phenol/chloroform (1/0.5/0.5) at 4°C. Eluted RNA was ethanol precipitated, resuspended in denaturing loading buffer (99% formamide, 1 mM Tris pH 7.8, 0.1 mM EDTA), boiled and loaded onto a 5% denaturing polyacrylamide gel. RNA was visualized by autoradiography.

## Acknowledgements

We thank C.Greer (University of California, Irvine) for plasmid pUC12T7; C.Trotta and J.Abelson (Caltech) for purified tRNA endonuclease; and C.Patil (Walter laboratory, UCSF) for providing us with the secondary structure prediction of the HAC1$^{u}$ mRNA intron used in Figure 1A. We also thank G.Culver for very helpful advice regarding TLC as well as L.Gonzalez, R.Gonzalez and J.Leber for their support throughout these studies. Special thanks go to M.Niwa, T.Powers, P.Peluso and A.Frankel for advice, discussions and comments on the manuscript. This work was supported by a UCSF Biomedical Science Research Career Enhancement Fellowship from the National Institute of

General Medical Science to T.N.G. and by grants from the National Institutes of Health to P.W. P.W. is an Investigator and C.S. is a Postdoctoral Associate of the Howard Hughes Medical Institute.

## References

Abelson,J., Trotta,C.R. and Li,H. (1998) tRNA splicing. *J. Biol. Chem.*, **273**, 12685–12688.

Belford,H.G., Westaway,S.K., Abelson,J. and Greer,C.L. (1993) Multiple nucleotide cofactor use by yeast ligase in tRNA splicing. Evidence for independent ATP- and GTP-binding sites. *J. Biol. Chem.*, **268**, 2444–2450.

Bork,P. and Sander,C. (1993) A hybrid protein kinase-RNase in an interferon-induced pathway? *FEBS Lett.*, **334**, 149–152.

Cameron,V. and Uhlenbeck,O.C. (1977) 3′-Phosphatase activity in T4 polynucleotide kinase. *Biochemistry*, **16**, 5120–5126.

Cameron,V., Soltis,D. and Uhlenbeck,O.C. (1978) Polynucleotide kinase from a T4 mutant which lacks the 3′ phosphatase activity. *Nucleic Acids Res.*, **5**, 825–833.

Chapman,R.E. and Walter,P. (1997) Translational attenuation mediated by an mRNA intron. *Curr. Biol.*, **7**, 850–859.

Chapman,R., Sidrauski,C. and Walter,P. (1998) Intracellular signaling from the endoplasmic reticulum to the nucleus. *Annu. Rev. Cell Dev. Biol*, **14**, 459–485.

Cox,J.S. and Walter,P. (1996) A novel mechanism for regulating activity of a trancription factor that controls the unfolded protein response. *Cell*, **87**, 391–404.

Cox,J.S., Shamu,C.E. and Walter,P. (1993) Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. *Cell*, **73**, 1197–1206.

Culver,G.M., McCraith,S.M., Zillmann,M., Kierzek,R., Michaud,N., LaReau,R.D., Turner,D.H. and Phizicky,E.M. (1993) An NAD derivative produced during transfer RNA splicing: ADP-ribose 1″-2″ cyclic phosphate. *Science*, **261**, 206–208.

Culver,G.M., McCraith,S.M., Consaul,S.A., Stanford,D.R. and Phizicky,E.M. (1997) A 2′-phosphotransferase implicated in tRNA splicing is essential in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **272**, 13203–13210.

Di Nicola Negri,E., Fabbri,S., Bufardeci,E., Baldi,M.I., Gandini Attardi,D., Mattoccia,E. and Tocchini-Valentini,G.P. (1997) The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. *Cell*, **89**, 859–866.

Dong,B. and Silverman,R.H. (1997) A bipartite model of 2-5A-dependent RNase L. *J. Biol. Chem.*, **272**, 22236–22242.

Dong,B., Xu,L., Zhou,A., Hassel,B.A., Lee,X., Torrence,P.F. and Silverman,R.H. (1994) Intrinsic molecular activities of the interferon-induced 2–5A-dependent RNase. *J. Biol. Chem.*, **269**, 14153–14158.

Fabbri,S., Fruscoloni,P., Bufardeci,E., Di Nicola Negri,E., Baldi,M.I., Attardi,D.G., Mattoccia,E. and Tocchini-Valentini,G.P. (1998) Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. *Science*, **280**, 284–286.

Gething,M.-J. and Sambrook,J. (1992) Protein folding in the cell. *Nature*, **355**, 33–45.

Greer,C.L., Peebles,C.L., Gegenheimer,P. and Abelson,J. (1983) Mechanism of action of a yeast RNA ligase in tRNA splicing. *Cell*, **32**, 537–546.

Greer,C.L., Soll,D. and Willis,I. (1987) Substrate recognition and identification of splice sites by the tRNA- splicing endonuclease and ligase from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **7**, 76–84.

Hammond,C. and Helenius,A. (1995) Quality control in the secretory pathway. *Curr. Opin. Cell Biol.*, **7**, 523–529.

Hassel,B.A., Zhou,A., Sotomayor,C., Maran,A. and Silverman,R.H. (1993) A dominant negative mutant of 2–5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon. *EMBO J.*, **12**, 3297–3304.

Kawahara,T., Yanagi,H., Yura,T. and Mori,K. (1997) Endoplasmic reticulum stress-induced mRNA splicing permits synthesis of transcription factor Hac1p/Ern4p that activates the unfolded protein response. *Mol. Biol. Cell*, **8**, 1845–1862.

Kawahara,T., Yanagi,H., Yura,T. and Mori,K. (1998) Unconventional splicing of HAC1/ERN4 mRNA required for the unfolded protein response. Sequence-specific and non-sequential cleavage of the splice sites. *J. Biol. Chem.*, **273**, 1802–1807.

Kleman-Leyer,K., Armbruster,D.W. and Daniels,C.J. (1997) Properties of *H.volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell*, **89**, 839–847.

Knapp,G., Ogden,R.C., Peebles,C.L. and Abelson,J. (1979) Splicing of yeast tRNA precursors: structure of the reaction intermediates. *Cell*, **18**, 37–45.

Kohno,K., Normington,K., Sambrook,J., Gething,M.J. and Mori,K. (1993) The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum. *Mol. Cell. Biol.*, **13**, 877–890.

Kuznetsov,G. and Nigam,S.K. (1998) Folding of secretory and membrane proteins. *N. Engl. J. Med.*, **339**, 1688–1695.

Lykke-Andersen,J. and Garrett,R.A. (1994) Structural characteristics of the stable RNA introns of archaeal hyperthermophiles and their splicing junctions. *J. Mol. Biol.*, **243**, 846–855.

Lykke-Andersen,J. and Garrett,R.A. (1997) RNA–protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO J.*, **16**, 6290–6300.

Mattoccia,E., Baldi,I.M., Gandini-Attardi,D., Ciafre,S. and Tocchini-Valentini,G.P. (1988) Site selection by the tRNA splicing endonuclease of *Xenopus laevis*. *Cell*, **55**, 731–738.

McCraith,S.M. and Phizicky,E.M. (1990) A highly specific phosphatase from *Saccharomyces cerevisiae* implicated in tRNA splicing. *Mol. Cell. Biol.*, **10**, 1049–1055.

Miao,F. and Abelson,J. (1993) Yeast tRNA-splicing endonuclease cleaves precursor tRNA in a random pathway. *J. Biol. Chem.*, **268**, 672–677.

Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Mori,K., Sant,A., Kohno,K., Normington,K., Gething,M.J. and Sambrook,J.F. (1992) A 22 bp *cis*-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. *EMBO J.*, **11**, 2583–2593.

Mori,K., Ma,W., Gething,M.J. and Sambrook,J. (1993) A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus. *Cell*, **74**, 743–756.

Mori,K., Kawahara,T., Yoshida,H., Yanagi,H. and Yura,T. (1996) Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. *Genes Cells*, **1**, 803–817.

Nikawa,J., Akiyoshi,M., Hirata,S. and Fukuda,T. (1996) *Saccharomyces cerevisiae* IRE2/HAC1 is involved in IRE1-mediated KAR2 expression. *Nucleic Acids Res.*, **24**, 4222–4226.

Peebles,C.L., Gegenheimer,P. and Abelson,J. (1983) Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. *Cell*, **32**, 525–536.

Phizicky,E.M., Consaul,S.A., Nehrke,K.W. and Abelson,J. (1992) Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. *J. Biol. Chem.*, **267**, 4577–4582.

Reyes,V.M. and Abelson,J. (1987) A synthetic substrate for tRNA splicing. *Anal. Biochem.*, **166**, 90–106.

Reyes,V.M. and Abelson,J. (1988) Substrate recognition and splice site determination in yeast tRNA splicing. *Cell*, **55**, 719–730.

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

Shamu,C.E., Cox,J.S. and Walter,P. (1994) The unfolded-protein-response pathway in yeast. *Trends Cell Biol.*, **4**, 56–60.

Sidrauski,C. and Walter,P. (1997) The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. *Cell*, **90**, 1–20.

Sidrauski,C., Cox,J.S. and Walter,P. (1996) tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. *Cell*, **87**, 405–413.

Sidrauski,C., Chapman,R. and Walter,P. (1998) The unfolded protein response: an intracellular signalling pathway with many surprising features. *Trends Cell Biol.*, **8**, 245–249.

Spinelli,S.L., Consaul,S.A. and Phizicky,E.M. (1997) A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2′-phosphate and an undermodified base at the splice junction. *RNA*, **3**, 1388–1400.

Thompson,L.D. and Daniels,C.J. (1990) Recognition of exon–intron boundaries by the *Halobacterium volcanii* tRNA intron endonuclease. *J. Biol. Chem.*, **265**, 18104–18111.

Tirasophon,W., Welihinda,A.A. and Kaufman,R.J. (1998) A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells. *Genes Dev.*, **12**, 1812–1824.

Trotta,C.R., Miao,F., Arn,E.A., Stevens,S.W., Ho,C.K., Rauhut,R. and Abelson,J.N. (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*, **89**, 849–858.

Valenzuela,P., Venegas,A., Weinberg,F., Bishop,R. and Rutter,W.J. (1978) Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA. *Proc. Natl Acad. Sci. USA*, **75**, 190–194.

Wang,X.Z., Harding,H.P., Zhang,Y., Jolicoeur,E.M., Kuroda,M. and Ron,D. (1998) Cloning of mammalian Ire1 reveals diversity in the ER stress responses. *EMBO J.*, **17**, 5708–5717.

Welihinda,A.A. and Kaufman,R.J. (1996) The unfolded protein response pathway in *Saccharomyces cerevisiae*. Oligomerization and *trans*-autophosphorylation of Ire1p (Ern1p) are required for kinase activation. *J. Biol. Chem.*, **271**, 18181–18187.

Welihinda,A.A., Tirasophon,W., Green,S.R. and Kaufman,R.J. (1998) Protein serine/threonine phosphatase Ptc2p negatively regulates the unfolded-protein response by dephosphorylating Ire1p kinase. *Mol. Cell. Biol.*, **18**, 1967–1977.

Westaway,S.K. and Abelson,J. (1995) Splicing of tRNA Precursors. In Söll,D. and RajBhandary,U. (eds), *tRNA: Structure, Biosynthesis and Function*. ASM Press, Washington, DC, pp. 79–92.

Xu,Q., Teplow,D., Lee,T.D. and Abelson,J. (1990) Domain structure in yeast tRNA ligase. *Biochemistry*, **29**, 6132–6138.

Zhou,A. *et al.* (1997) Interferon action and apoptosis are defective in mice devoid of 2′,5′-oligoadenylate-dependent RNase L. *EMBO J.*, **16**, 6355–6363.

*Received March 2, 1999; revised and accepted April 15, 1999*

# Chapter 3

# Preliminary characterization of the rlg1-100 tRNA ligase

## *INTRODUCTION*

Introns are a fact of life: organisms in the three domains of life harbor them in their genes. Perhaps the most familiar and well-known introns are those interrupting the coding sequences of eukaryotic mRNAs. However, tRNA introns are more widely distributed across the tree of life; they occur in subsets of tRNA genes found in archaebacteria, eubacteria, and eukaroytes (Abelson et al., 1998; Phizicky and Greer, 1993). [For an excellent review on many aspects of tRNA processing and function, see Hopper and Phizicky (Hopper and Phizicky, 2003).]

The manner in which tRNA introns are excised varies. Found in tRNA genes of chloroplasts, mitochondria, and bacteria, the self-splicing group I and II introns catalyze their own removal by two sequential transesterification reactions; the 5' splice site is cleaved followed by cleavage and exon joining at the 3' splice site. The origin of the nucleophilic attacking group at the 5' splice site differs for Group I and II introns. In Group I it is a 2'OH from a nucleotide within the intron; in Group II, it is a 3'OH from a guanosine cofactor. In both cases, the second cleavage at the 3' splice site uses the 3'OH produced as a consequence of the first cleavage at the 5'splice site. The Group I splicing mechanism is analogous to that utilized in the spliceosomal splicing of mRNAs. As a consequence, it has been proposed that spliceosomal splicing evolved from self-splicing RNAs like the Group I introns of tRNA genes (Eckstein and Lilley, 1997; Phizicky and Greer, 1993).

The splicing of eukaroytic nuclear pre-tRNAs proceeds by a third, very different catalytic mechanism catalyzed by two proteins, tRNA endonuclease and tRNA ligase, and is best understood in yeast (Abelson et al., 1998; Westaway and Abelson, 1995). tRNA endonuclease initiates splicing by cleaving the 5' and 3' splice sites, producing 5'OH and 2',3' cyclic phosphate RNA termini (Knapp et al., 1979; Peebles et al., 1983). Unlike Group I and II intron splicing, the initial cleavage event can occur at either the 5' or 3' splice site (Greer et al., 1987; Miao and Abelson, 1993). Orthologues of the yeast tRNA endonuclease (Trotta et al., 1997) have been identified in archaea, vertebrates, and plants (Akama et al., 2000; Fabbri et al., 1998; Kleman-Leyer et al., 1997; Lykke-Andersen and Garrett, 1997). Chemically, cleavage by tRNA endonuclease resembles the self-cleavage reaction catalyzed by viroids and virusoids, leading to speculation that tRNA endonuclease cleaved introns evolved from self-cleaving tRNA introns (Eckstein and Lilley, 1997; Phizicky and Greer, 1993). In the support of this are reports of self-cleaving introns found in a human and plant pre-tRNATyr (van Tol et al., 1989; Weber et al., 1996). Here, the splice sites are cleaved to produce 5'OH and 2',3'-cyclic phosphates. But unlike the RNA catalyzed Group I and Group II intron splicing, self-ligation does not occur.

Ligation of the exons liberated by tRNA endonuclease is catalyzed by the multifunctional enzyme tRNA ligase (Belford et al., 1993; Greer et al., 1983). Ligation begins with opening of the terminal 2',3'-cyclic phosphate at the end of the 5' exon. Next, the 3' exon is phosphorylated at its 5' end by transfer of a gamma phosphate from GTP or ATP. This phosphate will ultimately link the two exons together at the splice junction. The 5'-phosphate is activated by the transfer of AMP from ligase to form a high energy 5'-5' phosphoanhydride bond. Ligation occurs with the concomitant release of the AMP activating group, and leaves behind a 2' phosphate at the splice junction. The 2'-phosphate is subsequently removed by the NAD-dependent 2'-phosphotransferase, Tpt1p

(McCraith and Phizicky, 1990; McCraith and Phizicky, 1991). Tpt1p transfers the splice junction 2'-phosphophate to NAD to produce the small molecule ADP- ribose 1"-2" cyclic phosphate (Culver et al., 1993). The 2'-phosphate must be removed from spliced tRNAs in order for subsequent base modifications to take place to produce fully functional tRNAs for use in translation by the cell (Culver et al., 1997; Spinelli et al., 1997).

Though the ligation reaction is best understood for the yeast enzyme encoded by the *RLG1* gene in *Saccharomyces cerevisiae*, equivalent biochemical activities have been discovered in mammals and plants (Gegenheimer et al., 1983; Schwartz et al., 1983; Zillmann et al., 1991). Curiously, a second distinctive ligation activity has also been described in cellular extracts from mammals, frogs, and archea (Filipowicz and Shatkin, 1983; Laski et al., 1983; Zofallova et al., 2000). Here, the phosphate which links the two exons together comes from the pre-tRNA itself instead of from GTP or ATP and the final spliced product lacks the 2'-phosphate that marks the splice junction in the yeast-like reaction. The specific proteins responsible for these various ligation activities remain to be identified and their genes cloned.

Adding to this complicated mix is the fact that the NAD-dependent 2-phosphotransferase has been identified by activity or genomic sequence in plants, vertebrates, and archaea, suggesting that the yeast-like splicing mechanism is actively utilized in these organisms (Spinelli et al., 1998; Yukawa et al., 2001; Zillman et al., 1992). It remains to be seen which, if either, tRNA ligation pathway predominants in these organisms.

The yeast tRNA ligase protein is a multidomain protein with three distinct enzymatic activities (Apostol et al., 1991; Xu et al., 1990). The adenylylate synthetase resides in the N-terminal half, with the kinase in the middle quarter, and the cyclic phosphodiesterase in the final C-terminal quarter. Binding to tRNA appears to require multiple non-contiguous sites on the protein and thus has not been pinned down to any single location (Apostol et al., 1991). The massive genomic sequencing projects of the past decade have enabled the identification of regions of homology between tRNA ligase and other functionally related proteins (see appendix A). The adenylylate synthetase domain of tRNA ligase is a member of the covalent nucleotidyl transferase (CNT) superfamily (Aravind and Koonin, 1999; Shuman, 1996). Characterized by the presence of five to six conserved active-site motifs and a common protein fold, this protein superfamily includes ATP- and NAD-dependent DNA ligases, mRNA capping enzymes, and RNA ligases. The tRNA ligase kinase domain shares limited regions of homology with the ATP binding motifs found in T4 polynucleotide kinase and adenylate kinase (Apostol et al., 1991; Koonin and Gorbalenya, 1990). Finally, the 2',3'-cyclic phosphodiesterase domain of tRNA ligase has been identified as a member of the large 2H phosphodiesterase superfamily (Mazumder et al., 2002). This protein family is named for two conserved active site histidine residues each found in the motif Hh(S/T)h (where h is a hydrophobic residue). The histidine and serine/threonine residues of this motif are essential for phosphodiesterase function (Hofmann et al., 2000; Nasr and Filipowicz, 2000). These proteins are predicted to adopt a common fold typified by the structure of the *Arabidopsis* ADP-ribose 1",2" cyclic phosphodiesterase (Hofmann et al., 2002; Hofmann et al., 2000).

tRNA ligase is an essential gene in yeast (Phizicky et al., 1992). Since all genes for tRNAphe, tRNAser, tRNAtrp, and tRNAtyr carry introns, yeast must be able to splice pre-tRNAs in order to live (Westaway and Abelson, 1995). Unexpectedly, ligase was also shown to be required for cell survival when unfolded proteins accumulate in the endoplasmic reticulum (ER) (Sidrauski et al., 1996).

Accumulation of unfolded proteins in the lumen of the ER leads to activation of a cell survival signal transduction pathway, the unfolded protein response (UPR). Activation of the UPR results in the transcriptional up-regulation of genes encoding ER resident proteins that aid protein folding as well as those which mediate ER protein degradation, and aspects of the secretory pathway (Travers et al., 2000) (Ng et al., 2000) (Cox et al., 1997). This response is mediated to a large degree by the bZip transcription factor Hac1p (Cox et al., 1993; Travers et al., 2000). Despite the fact that *HAC1* mRNA levels do not fluctuate more than 2-fold during the UPR (Travers et al., 2000), Hac1p is only produced when the UPR is activated (Chapman and Walter, 1997; Cox and Walter, 1996). Interactions between a 252 nucleotide intron and the *HAC1* mRNA 5' UTR inhibit translation of the message (Ruegsegger et al., 2001). This translational block is relieved and translation resumes upon splicing out of the intron by the combined actions of the endoribonuclease Ire1p and tRNA ligase (Sidrauski et al., 1996; Sidrauski and Walter, 1997).

*HAC1* mRNA splicing and pre-tRNA splicing are mechanistically equivalent. Cleavage of both RNA substrates by their respective endoribonucleases generates 5'-hydroxyls and 2',3-cyclic phosphates; ligation proceeds via the same multistep process catalyzed by tRNA ligase (Gonzalez et al., 1999). Yet despite this mechanistic similarity there exist significant differences between the two splicing pathways. First, RNA cleavage in the two splicing reactions is catalyzed by two endoribonucleases that lack any discernable homology to one another. Nevertheless, the same enzyme is used to ligate the exons that they liberate. Second, while *HAC1* mRNA splicing is regulated and only occurs when the UPR is activated, pre-tRNA splicing is probably constitutive. Third, in yeast, pre-tRNA splicing likely takes place in the nucleus (Clark and Abelson, 1987; Peebles et al., 1983) whereas *HAC1* mRNA splicing occurs in the cytoplasm (Ruegsegger et al., 2001). How tRNA ligase divides its time between constituitive pre-tRNA splicing in the nucleus and regulated *HAC1* mRNA splicing in the cytosol remains an open question.

The role of tRNA ligase in the splicing of *HAC1* mRNA was uncovered when an UPR-defective allele of tRNA ligase, *rlg1-100*, was discovered in a genetic screen designed to identify essential components of the UPR signaling pathway. Cells are viable and splicing of pre-tRNAs is unaffected in strains carrying *rlg1-100* as their sole source of tRNA ligase, yet these strains do not survive on UPR-inducing media (Sidrauski et al., 1996). When the UPR is turned on in *rlg1-100* yeast, Ire1p initiates splicing by cleaving *HAC1* mRNA, but the resulting exon fragments are not ligated together as they are in wild type yeast. Instead, they are rapidly degraded and no Hac1p is made (Sidrauski et al., 1996). The most straightforward interpretation of this result was that the *rlg1-100* ligase, though functional for pre-tRNA splicing, was somehow specifically defective for *HAC1* mRNA splicing. The *rlg1-100* mutation is a histidine to tyrosine change at amino acid 148 in the adenylylate synthetase domain of ligase. Residue 148 is just N-terminal to the active site motif III found in covalent nucleotidyl transferase (CNT) proteins, of which tRNA ligase is a member (Appendix A, Figures A-1 and A-2). Though the mutation is adjacent to the CNT motif, the fact that pre-tRNA splicing is unaffected in *rlg1-100* yeast, strongly suggests that this mutation does not adversely affect the catalytic function of the adenylylate synthetase domain. Thus it is unlikely that *rlg1-100* ligase is defective for *HAC1* mRNA splicing due to significant defects in adenylylate synthetase activity. This is supported by our *in vitro* studies demonstrating that tRNA ligase uses the same catalytic functions for splicing pre-tRNAs and as it does for splicing *HAC1* mRNA (Gonzalez et al., 1999).

Why is the *rlgl1-100* allele of tRNA ligase defective for *HAC1* mRNA splicing and not for pre-tRNA splicing? An understanding of this should help address the question of how tRNA ligase divides it time between these two different splicing pathways *in vivo*. To begin to address these questions, this chapter describes our initial *in vitro* and *in vivo* characterization of the *rlg1-100* tRNA ligase. We end by proposing a new and easily testable model that may explain how tRNA ligase divides its duties between the pre-tRNA splicing and *HAC1* mRNA splicing pathways.

## MATERIALS AND METHODS

### Plasmids and yeast strains

The plasmids used in this study and information about their construction are given in Table III-1. The yeast strains used are listed in Table III-2.

Pop-in/pop-out allele replacement was employed to replace wild type *RLG1* with mutant *rlg1-100* at the wild type gene locus (Guthrie and Fink, 1991). To make strains

TGy-1 and TGy-3, plasmid pTG-1 was linearized within *rlg1-100* coding region and transformed into strains CSY228 and CSY227 respectively. The *rlg1-100* mutation destroys a NlaIII site. Digestion of a short PCR fragment (generated using genomic DNA as template) with NlaIII along with lack of strain growth on UPR-inducing media was used to verify replacement of the *RLG1* gene with rlg1-100.

To tag ligase with the 13Myc epitope, we used a PCR-based transformation method. We amplified a DNA cassette encoding the S.pombe *HIS5* marker along with the 13Myc epitope using oligo-nucleotide primers, each of which bore sequences homologous to the target ligase gene (Longtine et al., 1998). PCR products were transformed into diploid yeast, and the yeast plated onto minimal media lacking histidine. Accurate tagging of each gene was verified by PCR using genomic DNA as the template for amplification. Viable haploid ligase tagged strains were obtained upon sporulation of the diploid strains, indicating that the myc tagged version of ligase was functional in yeast. tRNA ligase is essential for cell viability (Phizicky et al., 1992).

### *Yeast growth media*

Standard recipes for YPD, and synthetic minimal media were used (Guthrie and Fink, 1991). Tunicamycin was added to plates within hours of plating cells by top spreading a solution of tunicamycin dissolved in DMSO. The final concentration of tunicamycin in the media was 0.25 μg/ml.

### *In vitro splicing assays*

Ire1p(k+t) was purified and splicing assays performed as previously described (see Chapter 2 of this thesis) (Gonzalez et al., 1999; Gonzalez and Walter, 2001). GST-ligase fusions were expressed and purified using the same method employed for Ire1p(k+t), except that the GST domain was not proteolytically removed, and remained.

### *Western blot analysis*

Yeast whole cell protein extracts were performed as previously described (Ruegsegger et al., 2001). Western blot analysis of myc-tagged ligase was performed using the monoclonal antibody HA.11 (Santa Cruz Biotechnology, Inc).

### *Isolation of Total RNA and Northern Blot Analysis*

Total yeast RNA was isolated and analyzed by Northern blot as previously described (Ruegsegger et al., 2001). The *HAC1* probe used in all the experiments was generated against the *HAC1* mRNA 5' exon. The *SCR1* probe was generated against the full-length *SCR1* RNA.

### *Immunofluorence, subcellular fractionation*

Immunofluorescence was performed essentially according to Redding et al (Redding et al., 1991). For subcellular fractionation, 50 OD600 of yeast were harvested by centrifugation and the cell pellet frozen in liquid nitrogen. The cell pellet was thawed and resuspended in 0.6 ml 50mM Hepes-KOH pH6.8, 150mM KOAc, 2mM MgOAc + Complete Protease Inhibitors Cocktail (Roche). Next, 0.6ml of 0.5 μm glass beads was added and the sample vortexed at 4°C for 45 seconds, followed by 45 seconds on ice. This vortexing and icing was repeated for a total of 6 rounds. The lysate was recovered from the glass beads and centrifuged at 1000 X *g*. The supernatant was collected (= total cell extract) and centrifuged for 10 min at 21,000 X *g*. The supernatant from this centrifugation carries small membranes and cytosol (soluble fraction) and the pellet contains microsomes and vacuoles (verified by western blot for proteins specific to each subcellular location).

***Calculations for the in vivo concentration of HAC1 mRNA and tRNA ligase***

When total RNA is extracted from yeast grown to mid-log phase, *HAC1* mRNA is present at 1 fM per 2 µg of total RNA (U. Ruesegger, personal communication). Using this value, one can calculate the amount of *HAC1* mRNA in an individual cell. In a typical experiment, we extract 100 ug of total RNA from 10 ml of cells grown to a density of 1.5 X $10^7$ cells/ml. Therefore, in 1 ml of cells we should have 10 ug of total RNA and 5 fM of total *HAC1* mRNA. Using Avagadro's number, the calculated number of *HAC1* mRNA molecules per cell is 200 (or 2.9 X $10^{-22}$ moles). As a comparison, it is estimated that there are 40 to 60 actin mRNA molecules per yeast cell (Kang et al., 2000; Velculescu et al., 1997).

Haploid yeast occupy a volume of 70 um (Guthrie and Fink, 2002) which is equivalent to 70 femtoliters or 7 X $10^{-14}$ liters. Dividing the total number of moles of *HAC1* mRNA per cell (2.9 X $10^{-22}$) by the volume of a single cell (7 X $10^{-14}$ liters) produces a value of 4.1 nM/L for the concentration of *HAC1* mRNA in a single cell. It is estimated that there are 500 tRNA ligase molecules per yeast cell (Clark and Abelson, 1987). Thus, the abundance of ligase per cell is 2.5 times that of *HAC1* mRNA and is 10 nM/L.

## *RESULTS*

### *The rlg1-100 allele of tRNA ligase is a partial UPR-null*

The startling discovery that ligase plays a central role in the UPR was made when a UPR-defect allele of tRNA ligase was pulled out of a genetic screen designed to identify components of the pathway (Sidrauski et al., 1996). It quickly became evident that tRNA ligase was required not only for pre-tRNA splicing in yeast, but also for the splicing of the mRNA encoding the UPR transcription factor, Hac1p (Sidrauski and Walter, 1997). When the UPR is induced in a wild type strain, within 30 minutes, up to 80% of *HAC1* mRNA is spliced. In a *rlg1-100* strain, the majority of the full length *HAC1* mRNA disappears; presumably the cleaved but unspliced *HAC1* mRNA fragments are degraded by the cellular mRNA degradation machinery that recognizes mRNAs lacking 5'-caps and 3' polyA tails as aberrant and destroys them. However, not all of the *HAC1* mRNA is degraded in *rlg1-100* cells, and a little bit of spliced *HAC1* mRNA can been seen on a Northern blot (Figure III-1, lane 4) (Sidrauski et al., 1996). Thus it appears that *rlg1-100* strains carry a tRNA ligase that is severely but not totally defective for *HAC1* mRNA splicing.

To investigate this further, we asked whether over-expression of the mutant ligase could compensate for the inability of *rlg1-100* yeast to grow on UPR-inducing media. Clark and Abelson had previously shown that when they transformed yeast with a high copy 2μ vector carrying the *RLG1* gene under the transcriptional control of its native promoter, steady state levels of wild type tRNA ligase increased by 35-fold (Clark and Abelson, 1987). We took a similar approach, and transformed *rlg1-100* yeast with a high copy 2μ plasmid or a low copy CEN/ARS plasmid carrying the *rlg1-100* gene with its native promoter. When these strains were plated onto UPR-inducing media, the 2μ and CEN/ARS *rlg1-100* plasmids both rescued growth of the *rlg1-100* yeast (Figure III-2, rows 2, 4, 5). In parallel, we also tested *rlg1-100* yeast transformed with 2μ and CEN/ARS plasmids carrying the *HAC1* gene or the *IRE1* gene, both with their native promoters. These plasmids also enabled *rlg1-100* yeast to grow under UPR-inducing conditions, but the effect was not as great as that seen with the *rlg1-100* plasmids (Figure III-2, rows 6, 7, 8).

These results supported the idea that *rlg1-100* ligase is partially active *in vivo*. We speculated that overexpression of the partially active ligase might lead to an increase in levels of spliced *HAC1* mRNA. Northern blot analysis of total RNA purified from the plasmid-bearing *rlg1-100* strains demonstrated that this was so (Figure III-3, lanes 4, 6, 10). Overexpression of *HAC1* mRNA or Ire1p might be expected to likewise increase levels of spliced *HAC1* mRNA in UPR-induced cells. However, we did not verify this. Nonetheless, these results suggest that the inability to grow on UPR-inducing media is due to the inability to produce enough spliced *HAC1* mRNA to in turn produce enough Hac1p to enable the cells to mount a successful UPR response so that they may live.

### ***Steady state levels of wild type and rlg1-100 ligase are comparable in vivo***

Over-expression of the *rlg1-100* ligase rescued growth on UPR-inducing media. This suggested that in the absense of *rlg1-100* overexpression, the steady state level of *rlg1-100* ligase was reduced relative to that found in wild type yeast. To determine is if this was so, we used Western blot analysis to gauge the relative abundance of the wild type and mutant ligase. We tagged each protein with thirteen tandem myc epitopes and verified that that both proteins were functional in yeast (see materials and methods). Epitope-tagged versions of each gene were integrated at the *RLG1* genomic locus so that each strain carried a single copy of a tagged ligase. When visualized by Western blot, the steady state levels of mutant and wild type ligase were comparable (Figure III-4A) and did not change even after 30 minutes of UPR induction (Figure III-4B). Thus differing steady state levels of wild type or mutant tRNA ligase could not explain the differences seen in *HAC1* mRNA splicing and growth during the UPR.

### ***rlg1-100 ligase splices HAC1 mRNA in vitro***

Over-expression of the mutant ligase could increase *HAC1* mRNA splicing by compensating for reduced splicing activity of the mutant relative to the wild type ligase. If this were true, then when compared in side-by-side *in vitro* assays, the mutant ligase should display reduced splicing activity relative to the wild type. To perform the comparison, we GST-tagged, expressed, and purified recombinant ligase. When compared over an 80-fold concentration range, the wild type and *rlg1-100* proteins spliced *HAC1* mRNA equivalently (Figure III-5). Thus, it initially appeared that the two ligases were equivalent when it comes to splicing *HAC1* mRNA *in vitro*. However, closer inspection of the experimental set-up revealed this conclusion to be suspect.

Ligase is estimated to bind pre-tRNAs and cleaved tRNAs with a Kd of 1 nM and 0.1 nM respectively (Apostol and Greer, 1991). In our *in vitro* splicing reactions, the concentration of ligase ranged from 17 μM to 0.2 μM while the concentration of *HAC1* mRNA remained constant at 0.87 nM. Let us assume that wild type ligase binds *HAC1* mRNA as strongly as it binds tRNA (~0.5 nM) and that *rlg1-100* ligase binds *HAC1* mRNA less well. In order to discern a difference between the two ligases in our *in vitro* splicing assay, the binding of *rlg1-100* would have had to be more than 10-fold weaker than that of wild type ligase. We calculate that at a Kd of 0.5 nM, a ligase concentration of 200 nM (the lowest we tested; see Figure III-5, lanes 7 and 12), and a *HAC1* mRNA concentration of 0.87 nM, 98% of the *HAC1* mRNA would bound by ligase. If the Kd were 5 nM (10-fold weaker) or 50 nM (100-fold weaker) under these same conditions, then we calculate that 98% and 60% respectively of *HAC1* would be bound by ligase.

We estimate the *in vivo* concentrations of *HAC1* mRNA and tRNA ligase to be 4 nM and 10 nM respectively (see Materials and Methods for calculations). Again, if we assume that the Kd for binding *HAC1* mRNA is similar to that for binding tRNA (~0.5 nM), then even a 10-fold increase in Kd could be enough to significantly reduce binding of *HAC1* mRNA by ligase *in vivo*. At a Kd of 0.5 nM, we calculate that 92% of *HAC1* mRNA would be bound by ligase; at a Kd of 5 nM, we calculate 60% of *HAC1* mRNA would be bound; and at a Kd of 50 nM, we calculate 16% of *HAC1* mRNA would be bound. Therefore, a 10-fold reduction in binding might be enough to reduce *in vivo* *HAC1* mRNA splicing as seen in *rlg1-100* strains. However, a 10-fold reduction in binding would likely go unnoticed in our *in vitro* splicing assays when performed under the conditions we used in this study (Figure III-5).

It would be best to repeat our *in vitro* splicing assay in a manner where the concentration of ligase is reduced significantly and maintained while the concentration of *HAC1* mRNA is manipulated instead. In addition, it would be very useful to determine directly the binding affinity of wild type and *rlg1-100* ligase for *HAC1* mRNA.

At the very least, this *in vitro* splicing assay demonstrates that the catalytic functions of the *rlg1-100* ligase are not severely compromised. This result supports *in vivo* data demonstrating that *rlg1-100* yeast are not defective for pre-tRNA splicing, a process requiring all the same catalytic functions as *HAC1* mRNA splicing (Gonzalez et al., 1999; Sidrauski et al., 1996).

### *Subcellular localization of tRNA ligase*

Splicing of pre-tRNAs in yeast is thought to be nuclear. When purified from yeast, tRNA endonuclease behaves as an integral membrane protein (Peebles et al., 1983; Rauhut et al., 1990); Sen2p, one of four protein subunits comprising tRNA endonuclease, carries a putative transmembrane sequence likely responsible for tethering the tRNA endonuclease complex to membranes (Trotta et al., 1997). However, it is unclear if the membrane that tRNA endonuclease associates with is nuclear or not. Yeast tRNA ligase fractionates as a soluable protein (Greer et al., 1983; Phizicky et al., 1986) and localizes to the nucleoplasm and nuclear pores (Clark and Abelson, 1987). When overexpressed on a 2µ plasmid, ligase can also be visualized in the cytoplasm (Clark and Abelson, 1987).

In contrast, *HAC1* mRNA splicing takes place in the cytoplasm (Ruegsegger et al., 2001), most likely at the ER membrane where Ire1p is expected to be located. Therefore, some tRNA ligase must be present in the cytoplasm for *HAC1* mRNA splicing to take place. One explanation for the *rlg1-100 HAC1* mRNA splicing defect is that the pool of ligase required for *HAC1* mRNA splicing does not get properly localized to the cytoplasm.

To determine if the subcellular localization of the mutant ligase differed from that of wild type, we used three different approaches. First, we C-terminally tagged tRNA ligase with three tandem green fluorescent proteins (GFP), an approach successfully used by O'Shea and colleagues to visualize the Pho4 transcription factor in yeast (Kaffman et al., 1998). We verified that both wild type and mutant GFP tagged proteins were functional in yeast by testing first for the ability to complement a ligase deletion (ligase is an essential gene), and second, for the ability of wild type GFP-tagged ligase to grow on UPR inducing plates (data not shown). We then used confocal microscopy to inspect yeast carrying these GFP-tagged constructs. Unfortunately, the GFP fluorescence was weak and we could not detect a signal over background (data not shown). It is estimated that there are about 400-500 ligase proteins per yeast cell, making it a low-abundance protein (Clark and Abelson, 1987; Phizicky et al., 1986). If ligase does not cluster significantly at discrete locations in the cell, it is not surprising we could not discern a discrete GFP signal.

In our second approach, we again C-terminally tagged wild type and mutant ligase, but this time we used thirteen tandem myc-epitopes. Again, we verified that both wild type and mutant myc-tagged ligases were functional in yeast using the same criteria as for the GFP-tagged proteins above. These strains were then fixed, probed with anti-myc antibodies, and stained with a secondary antibody conjugated to a fluorescent dye. Again, we were unable to detect a signal over the background in the microscope (data not shown).

In our third attempt at ligase localization, we used a biochemical approach. *HAC1* mRNA splicing likely takes place at the ER membrane where the endonuclease that initiates splicing, Ire1p, is located. In support of this, *HAC1* mRNA is enriched in ER membrane associated polysomes (Diehn et al., 2000). We hypothesized therefore, that *rlg1-100* ligase might not locate properly to the cytoplasm because it might associate less well with membranes compared to wild type ligase. We made cell extracts from yeast carrying the myc-tagged wild type or mutant ligase, separated the extracts into membrane associated (pellet fraction), and soluble fractions, and then analyzed these by Western blot. Both the wild type and the mutant ligases fractionated into the membrane (Figure III-6, lanes 2 and 5) and soluble compartments (Figure III-3, lanes 3 and 6) of the cell to the same degree. Thus it is unlikely that *rlg1-100* is defective for *HAC1* mRNA splicing due to a weakened ability to associate with membranes.

Interestingly, though we separated these proteins under basically the same conditions used in Figure III-4, here we saw a difference in the mobilities of the two ligases (Figure III-6). This could reflect the exciting possibility that one of the proteins undergoes a post-translational modification that may influence its function. Alternatively, and less exciting, is the possibility that one or more of the thirteen tandem myc-epitopes used to tag these proteins recombined out of the genome. We have noted the loss by genomic recombination of myc-epitopes from 13-myc tagged proteins before. It needs to be determined if the *rlg1-100* ligase strain used in this experiment has undergone such a recombination event.

## ***DISCUSSION***

We have investigated mechanisms which might explain the *in vivo* UPR defect seen in *rlg1-100* strains. Based on our data, we now know that the *in vivo* steady state level of *rlg1-100* ligase does not differ significantly from that of wild type ligase. In addition, over-expression of the *rlg1-100* ligase can rescue the ability of cells to splice *HAC1* mRNA and grow under UPR inducing conditions. This suggests that over-expression of the mutant ligase increased *HAC1* mRNA splicing by compensating for reduced splicing activity of the mutant relative to the wild type ligase. Alternatively, *rlg1-100* could be just as active as wild type ligase for splicing, but in the *in vivo* situation, insufficient amounts of the mutant ligase localize to the cytoplasm where *HAC1* mRNA splicing takes place. We have not been able to determine if either of these hypotheses is correct.

In the course of reviewing our data and the published literature, we have developed a model for the *in vivo* splicing of *HAC1* mRNA by Ire1p and tRNA ligase consistent with what is currently known about the process. Our model provides an explanation for how tRNA ligase localizes to the site of *HAC1* mRNA splicing in the cytoplasm and results in a number of easily tested predictions for future investigations of *HAC1* mRNA splicing in both wild type and *rgl1-100* yeast.

In our model (Figure III-7), tRNA ligase binds to newly transcribed *HAC1* mRNA in the nucleus and travels with the mRNA out into the cytoplasm, where ribosomes load onto the mRNA as well. The ligase-mRNA-ribosome complex then makes its way to the ER membrane where UPR activated Ire1p initiates splicing by cleaving *HAC1* mRNA. The mRNA bound tRNA ligase joins the resulting RNA fragments together. Once splicing is achieved, ligase is released from the mRNA and is free to recycle back to the nucleus. Back in the nucleus, it can bind to another *HAC1* mRNA in preparation for the next round of *HAC1* mRNA splicing.

This model is pleasing on a number of levels. It incorporates a mechanism by which ligase is localized to the site of *HAC1* mRNA splicing in the cytoplasm. It also makes the easily tested prediction that ligase should be found bound to polyribosome-associated, unspliced *HAC1* mRNA in sucrose gradients.

In the nucleus, there appear to be two pools of tRNA ligase. One pool localizes to the nuclear pore, where pre-tRNA splicing is hypothesized to happen. The second pool is not associated with any visible membranes and is located in the nucleoplasm within 200 nm of the nuclear envelop (Clark and Abelson, 1987). It is tempting to speculate that the nucleoplasmic pool contains the tRNA ligase involved in *HAC1* mRNA splicing. It is estimated that in any given cell, there are 400-500 tRNA ligase proteins and 100-150 tRNA endonuclease proteins (Clark and Abelson, 1987; Phizicky et al., 1986; Rauhut et al., 1990). This ratio of ligase to tRNA endonuclease also suggests that there could be a pool of tRNA ligase that is not engaged with pre-tRNA splicing and that is readily available for *HAC1* splicing. Interestingly, when ligase is overexpressed in yeast, it can be visualized in the cytosol as well (Clark and Abelson, 1987).

How might our model be used to explain the UPR defect of *rlg1-100*? Two scenarios come to mind. In the first, we envision *rlg1-100* ligase binding *HAC1* mRNA more weakly than wild type ligase does. If so, this would reduce the amount of *HAC1* mRNA leaving the nucleus bound by tRNA ligase. Upon induction of the UPR, Ire1p would cleave these mRNAs, but because they lack bound ligase, splicing would not be completed. The unjoined *HAC1* mRNA fragments would then be destroyed by the cell's mRNA degradation machinery. This is what we witness happening *in vivo* (Sidrauski et al., 1996).

The second scenario involves recycling of ligase back to the nucleus following *HAC1* mRNA splicing. Ligase has a molecular weight of 97kD, well above the 65kD cut off for passive diffusion of molecules through the nuclear pore. Thus, to travel from the cytoplasm to the nucleus following splicing (or even its translation on ribosomes), ligase must be actively transported. Perhaps the mutation in *rlg1-100* prevents interaction with an importin or another protein with which ligase might piggy-back its way into the nucleus. As an outcome of this, we would predict that the nuclear pools of ligase would eventually be depleted, and that tRNA splicing would be adversely affected. As *rlg1-100* strains splice their pre-tRNAs just as well as wild type strains, this scenario appears less plausible to us than the first (Sidrauski et al., 1996). However, this model could still hold up if there exist two separate pools of tRNA ligase, one for *HAC1* mRNA splicing and the other for pre-tRNA splicing, and if insignificant exchange takes place between the two pools.

What are the experimental predictions of these two scenarios? First, if wild type ligase were found to be associated with unspliced *HAC1* mRNA in polysomes, we would expect to see much less of the mutant to be. Second, if we were able to visualize ligase in yeast (perhaps using the method successfully used by Abelson and colleagues), we would expect to see a change in the nuclear pool of ligase in *rlg1-100* strains. If the first scenario were true, we would expect the nuclear pools of ligase to be greater in mutant yeast than those seen in wild type yeast. If the second scenario were true, we would expect to see just the opposite. Lastly, the first scenario predicts that *rlg1-100* ligase binds *HAC1* mRNA more weakly than the wild type ligase does. In the second scenario, a change in affinity for *HAC1* mRNA is not necessary to explain the UPR defect. Thus we would expect that wild type and mutant ligase would bind *HAC1* mRNA to the same degree.

As mentioned in the introduction to this chapter, there appear to be two distinct tRNA ligase activites in mammalian cells. It is unclear which is utilized for pre-tRNA splicing. It is also unclear which is utilized for splicing of the recently identified *HAC1* mammalian homologue, XBP1. Like *HAC1* mRNA splicing in yeast, XBP1 mRNA splicing is initiated by IRE1 cleavage. Perhaps we might be able to identify the mammalian ligase utilized for XBP1 mRNA splicing by identifying those proteins that are bound to the unspliced XBP1 mRNA.

## *Table III-1 Chapter 3 plasmid list*

*Plasmids*

| **Plasmid** | **Yeast marker** | **CEN, 2μ or integrating** | **Cloned gene** | **Comments** |
|---|---|---|---|---|
| pTG-1 | URA3 | integrating | *rlg1-100* | subcloned *rlg1-100* from pCF158 (C. Sidrauski, Walter lab) into pRS306 |
| pTG-4 | TRP | CEN | RLG1 | subcloned *RLG1* from pCF157 (C. Sidrauski, Walter lab) into pRS314 |
| pTG-5 | URA | CEN | rlg1-100 | subcloned *rlg1-100* from pCF158 (C. Sidrauski, Walter lab) into pRS316 |
| pTG-11 | URA | 2μ | rlg1-100 | subcloned *rlg1-100* from pCF158 (C. Sidrauski, Walter lab) into pRS426 |
| pCS110 | TRP | CEN | IRE1 | C. Shamu, Walter lab |
| pCS122 | TRP | 2μ | IRE1 | C. Shamu, Walter lab |
| pJC327 | LEU | CEN | HA-HAC1 | J. Cox, Walter lab |
| pFA6a-13Myc -His3MX6 | HIS5 | ---- | 13Myc tag | PCR template for 13Myc epitope |
| pCF210 | --- | --- | GST-Ire1p(k+t) | C. Sidrauski, Walter lab;<br>E.coli protein expression plasmid; pharmacia precision protease site between GST and Ire1p |
| pSD102 | --- | --- | GST-rlg1-100 | S.Nock, Walter lab<br>E.coli protein expression plasmid; pharmacia precision protease site between GST and rgl1-100p |
| pSD103 | --- | --- | GST-RLG1 | S.Nock, Walter lab<br>E.coli protein expression plasmid; pharmacia precision protease site between GST and Rlg1p |

## *Table III-2 Chapter 3 yeast strains*

| STRAIN | GENOTYPE |
|---|---|
| W303-1A | *MATa; leu2-3,-112; ura3-1,-112; his3-11,-15; trp1-1; ade2-1; can1-100* |
| CSY227 | same as W303-1A, except *his3-11,-15::HIS3-UPRE-lacZ*. CSY227 also known as PWY374 |
| CSY228 | same as CSY228, except *MAT∝*. CSY228 also known as PWY373 |
| TGy-1 | same as CSY228, except *RLG1* replaced by *rlg1-100* using pTG-1 (see methods) |
| TGy-3 | same as CSY227, except *RLG1* replaced by *rlg1-100* using pTG-1 (see methods) |
| TGy-57 | same as W303-1A except *RLG1::13Myc* where *RLG1* C-terminally tagged with thirteen Myc epitopes linked to S.pombe *HIS5* gene. See methods. |
| TGy-58 | same as TGy-57 except *MAT∝*. |
| TGy-130 | same as TGy-1 except *his3-11,-15*, mating type unknown, *rlg1-100::13Myc*. See notes regarding Myc for TGy-57. |

***Figure III-1 Northern blot analysis of HAC1 mRNA splicing in RLG1 and rlg1-100 strains***

Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. Unspliced, spliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are indentified along the left hand side of the blot.

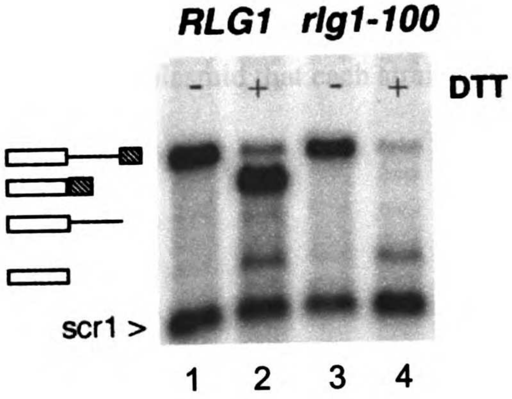
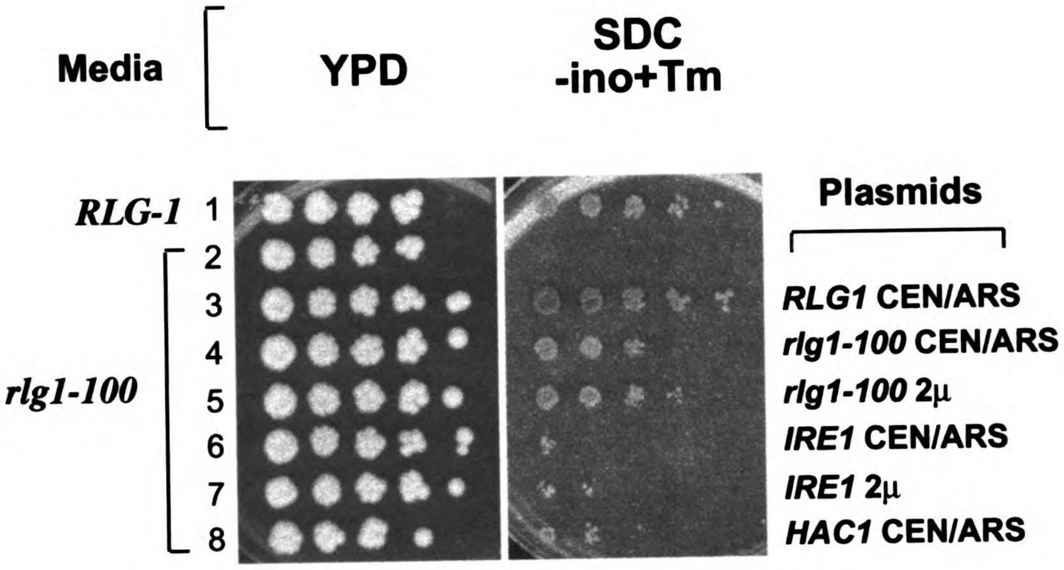

RLG1
rlg1-100
-
+
-
+
DTT
scr1 >
1
2
3
4

***Figure III-2    Growth of rlg1-100 strains overexpressing rlg1-100, RLG1, IRE1, or HAC1***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. SDC-ino+Tm indicates synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 μg/ml. The genotype of each strain is indicated on the left and the plasmid that each strain carried is indicated on the right.


Media
YPD
SDC
-ino+Tm
RLG-1 1
2
3
4
rlg1-100
5
6
7
8
Plasmids
RLG1 CEN/ARS
rlg1-100 CEN/ARS
rlg1-100 2μ
IRE1 CEN/ARS
IRE1 2μ
HAC1 CEN/ARS

UCSF LIBRARY

***Figure III-3 Northern blot analysis of rlg1-100 strains overexpressing rlg1-100***

Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. SCR1 RNA was used as a loading control. Unspliced and spliced *HAC1* mRNA species are indentified along the right hand side of the blot.

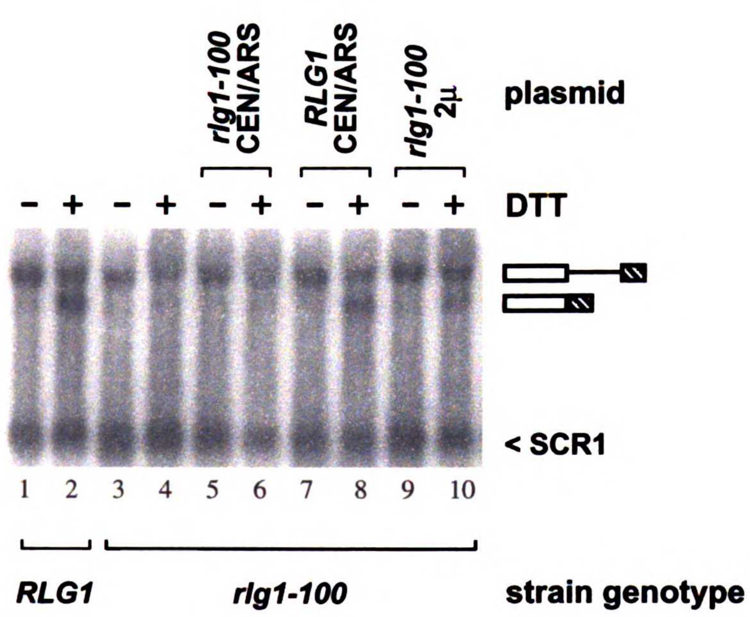
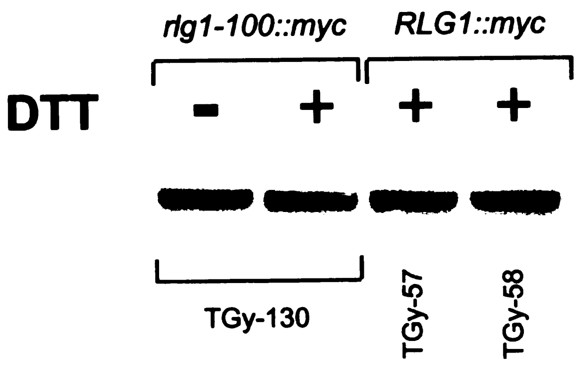

rlg1-100 CEN/ARS
RLG1 CEN/ARS
rlg1-100 2µ
plasmid
- + - + - + - + - +
DTT
< SCR1
1 2 3 4 5 6 7 8 9 10
RLG1
rlg1-100
strain genotype

***Figure III-4 Western blot analysis of tRNA ligase protein in rlg1-100 and RLG1 strains***

Total protein was extracted and analyzed by western blot to visualize myc-tagged tRNA ligase as described in the materials and methods section. The UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Cultures were grown at 30°C in YPD. The strains used for each experiment are indicated (TGy) along with the identity of the myc-tagged ligase carried by each strain. Panel A: *RLG1* ligase and *rlg1-100* ligase levels are equivalent. Panel B: *RLG1* ligase and *rlg1-100* ligase levels do not change upon UPR induction.

**A**


*rlg1-100::myc*
*RLG1::myc*
DTT
-
+
+
+
TGy-130
TGy-57
TGy-58

**B**

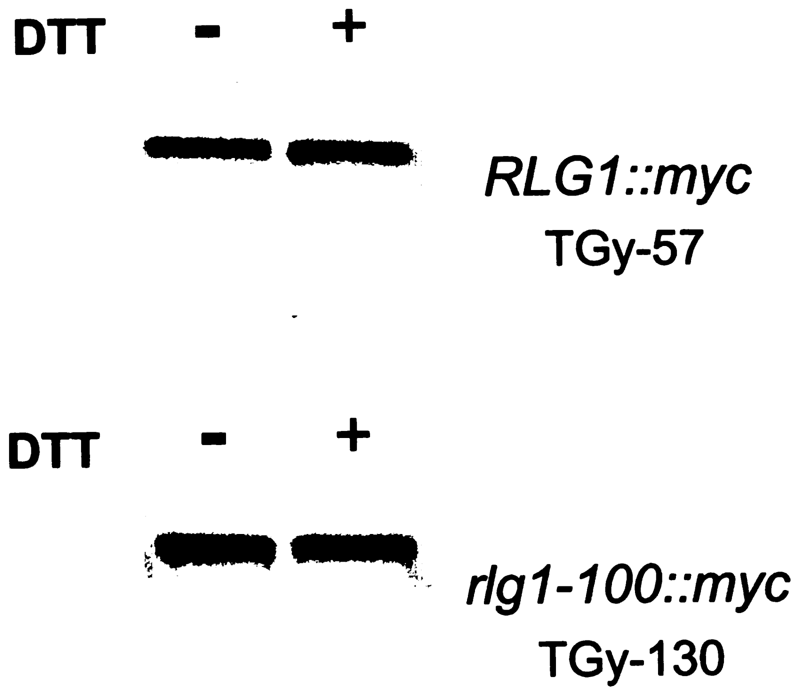
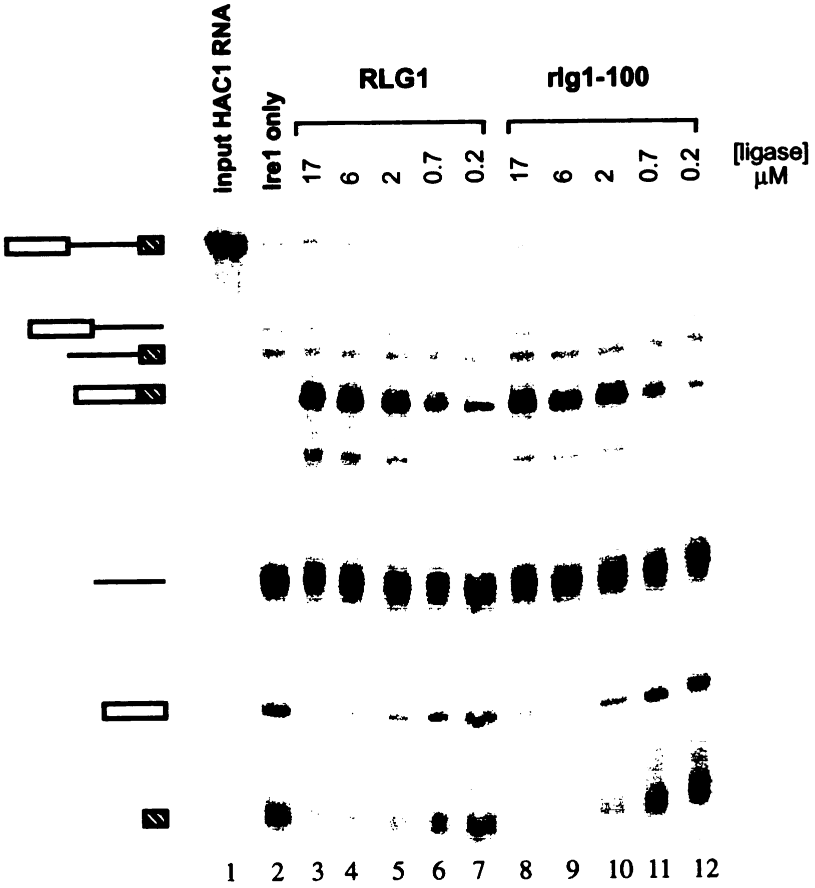

DTT
-
+
*RLG1::myc*
TGy-57
DTT
-
+
*rlg1-100::myc*
TGy-130

***Figure III-5 In vitro splicing assay comparing RLG1 and rlg1-100 ligase activities***

The final concentration of ligase protein used in each reaction is given above each lane. The icons on the left represent the *HAC1* mRNA species produced upon Ire1p cleavage and tRNA ligase ligation. From top to bottom these are: full length *HAC1*, 5'exon + intron *HAC1*, intron + 3'exon *HAC1*, spliced *HAC1*, intron *HAC1*, 5'exon *HAC1*, and 3' exon *HAC1*. The assay was performed as described in the Materials and Methods section.



input HAC1 RNA
Ire1 only
RLG1
rlg1-100
17
6
2
0.7
0.2
17
6
2
0.7
0.2
[ligase]
µM
1
2
3
4
5
6
7
8
9
10
11
12

***Figure III-6 Western blot analysis of ligase in the membrane and soluble cellular fractions of RLG1 and rlg1-100 strains***

Subcellular fraction and Western blot analysis was carried out as described in the Materials and Methods. Two isogenic strains, one carrying a myc-tagged wild type ligase (*RLG1*), the other carrying a myc-tagged *rlg1-100* ligase were analyzed. T, P, and S refer respectively to total cell extract, pelleted fraction, and soluble fraction. We thank Jason Brickner for collaborating with us and for providing us with this figure.

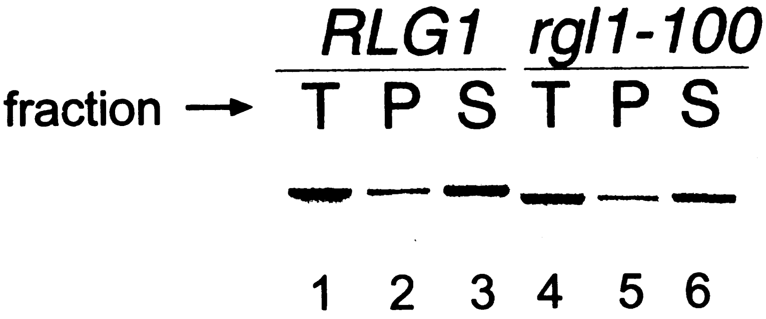
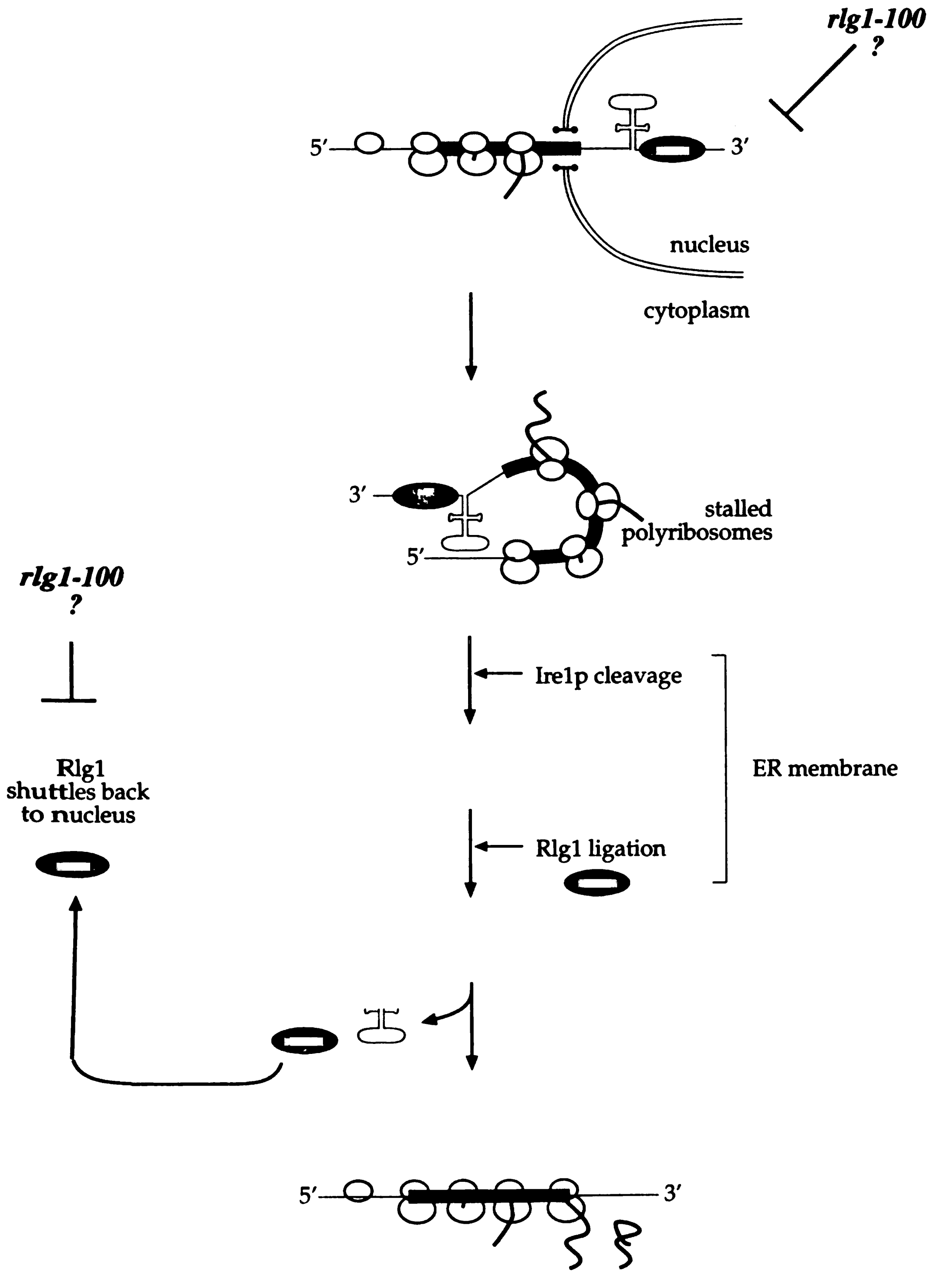

RLG1
rgl1-100
fraction →
T P S T P S
1 2 3 4 5 6

***Figure III-7 A model for HAC1 mRNA splicing in yeast***

tRNA ligase (green oval) binds to newly transcribed *HAC1* mRNA in the nucleus and travels with the mRNA out into the cytoplasm, where ribosomes load onto the mRNA. The ligase-mRNA-ribosome complex then makes its way to the ER membrane where UPR activated Ire1p initiates splicing by cleaving *HAC1* mRNA. The mRNA bound tRNA ligase joins the resulting RNA fragments together. Once splicing is achieved, ligase is released from the mRNA and is free to recycle back to the nucleus. Back in the nucleus, it can bind to another *HAC1* mRNA in preparation for the next round of *HAC1* mRNA splicing. We speculate that *rlg1-100* tRNA ligase might be defective for binding to *HAC1* mRNA in the nucleus or for recycling back to the nucleus following *HAC1* splicing in the cytoplasm.


rlg1-100
?
5′
3′
nucleus
cytoplasm
3′
5′
stalled polyribosomes
rlg1-100
?
Ire1p cleavage
ER membrane
Rlg1 shuttles back to nucleus
Rlg1 ligation
5′
3′

## *REFERENCES*

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. J Biol Chem *273*, 12685-12688.

Akama, K., Junker, V., and Beier, H. (2000). Identification of two catalytic subunits of tRNA splicing endonuclease from Arabidopsis thaliana. Gene *257*, 177-185.

Apostol, B. L., and Greer, C. L. (1991). Preferential binding of yeast tRNA ligase to pre-tRNA substrates. Nucleic Acids Res *19*, 1853-1860.

Apostol, B. L., Westaway, S. K., Abelson, J., and Greer, C. L. (1991). Deletion analysis of a multifunctional yeast tRNA ligase polypeptide. Identification of essential and dispensable functional domains. J Biol Chem *266*, 7445-7455.

Aravind, L., and Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol *287*, 1023-1040.

Belford, H. G., Westaway, S. K., Abelson, J., and Greer, C. L. (1993). Multiple nucleotide cofactor use by yeast ligase in tRNA splicing. Evidence for independent ATP- and GTP-binding sites. J Biol Chem *268*, 2444-2450.

Chapman, R. E., and Walter, P. (1997). Translational attenuation mediated by an mRNA intron. Curr Biol *7*, 850-859.

Clark, M. W., and Abelson, J. (1987). The subnuclear localization of tRNA ligase in yeast. J Cell Biol *105*, 1515-1526.

Cox, J. S., Chapman, R. E., and Walter, P. (1997). The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Mol Biol Cell *8*, 1805-1814.

Cox, J. S., Shamu, C. E., and Walter, P. (1993). Transcriptional Induction of Genes Encoding Endoplasmic Reticulum Resident Proteins Requires a Transmembrane Protein Kinase. Cell *73*, 1197-1206.

Cox, J. S., and Walter, P. (1996). A novel mechanism for regulating activity of a trancription factor that controls the unfolded protein response. Cell *87*, 391-404.

Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., and Phizicky, E. M. (1997). A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. J Biol Chem *272*, 13203-13210.

Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., LaReau, R. D., Turner, D. H., and Phizicky, E. M. (1993). An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"- 2" cyclic phosphate. Science *261*, 206-208.

Diehn, M., Eisen, M. B., Botstein, D., and Brown, P. O. (2000). Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nat Genet *25*, 58-62.

Eckstein, F., and Lilley, D. M. J. (1997). Catalytic RNA, 1st edn (Berlin ; New York, Springer).

Fabbri, S., Fruscoloni, P., Bufardeci, E., Di Nicola Negri, E., Baldi, M. I., Attardi, D. G., Mattoccia, E., and Tocchini-Valentini, G. P. (1998). Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. Science *280*, 284-286.

Filipowicz, W., and Shatkin, A. J. (1983). Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. Cell *32*, 547-557.

Gegenheimer, P., Gabius, H. J., Peebles, C. L., and Abelson, J. (1983). An RNA ligase from wheat germ which participates in transfer RNA splicing in vitro. J Biol Chem *258*, 8365-8373.

Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. Embo J *18*, 3119-3132.

Gonzalez, T. N., and Walter, P. (2001). Ire1p: a kinase and site-specific endoribonuclease. Methods Mol Biol *160*, 25-36.

Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983). Mechanism of action of a yeast RNA ligase in tRNA splicing. Cell *32*, 537-546.

Greer, C. L., Soll, D., and Willis, I. (1987). Substrate recognition and identification of splice sites by the tRNA- splicing endonuclease and ligase from Saccharomyces cerevisiae. Mol Cell Biol *7*, 76-84.

Guthrie, C., and Fink, G. R. (1991). Guide to yeast genetics and molecular biology (San Diego, Academic Press).

Guthrie, C., and Fink, G. R. (2002). Guide to yeast genetics and molecular and cell biology (San Diego, Calif., Academic Press).

Hofmann, A., Grella, M., Botos, I., Filipowicz, W., and Wlodawer, A. (2002). Crystal structures of the semireduced and inhibitor-bound forms of cyclic nucleotide phosphodiesterase from Arabidopsis thaliana. J Biol Chem *277*, 1419-1425.

Hofmann, A., Zdanov, A., Genschik, P., Ruvinov, S., Filipowicz, W., and Wlodawer, A. (2000). Structure and mechanism of activity of the cyclic phosphodiesterase of Appr>p, a product of the tRNA splicing reaction. Embo J *19*, 6207-6217.

Hopper, A. K., and Phizicky, E. M. (2003). tRNA transfers to the limelight. Genes Dev *17*, 162-180.

Kaffman, A., Rank, N. M., O'Neill, E. M., Huang, L. S., and O'Shea, E. K. (1998). The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. Nature *396*, 482-486.

Kang, J. J., Watson, R. M., Fisher, M. E., Higuchi, R., Gelfand, D. H., and Holland, M. J. (2000). Transcript quantitation in total yeast cellular RNA using kinetic PCR. Nucleic Acids Res *28*, e2.

Kleman-Leyer, K., Armbruster, D. W., and Daniels, C. J. (1997). Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Cell *89*, 839-847.

Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979). Splicing of yeast tRNA precursors: structure of the reaction intermediates. Cell *18*, 37-45.

Koonin, E. V., and Gorbalenya, A. E. (1990). Related domains in yeast tRNA ligase, bacteriophage T4 polynucleotide kinase and RNA ligase, and mammalian myelin 2',3'-cyclic nucleotide phosphohydrolase revealed by amino acid sequence comparison. FEBS Lett *268*, 231-234.

Laski, F. A., Fire, A. Z., RajBhandary, U. L., and Sharp, P. A. (1983). Characterization of tRNA precursor splicing in mammalian extracts. J Biol Chem *258*, 11974-11980.

Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Lykke-Andersen, J., and Garrett, R. A. (1997). RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. Embo J *16*, 6290-6300.

Mazumder, R., Iyer, L. M., Vasudevan, S., and Aravind, L. (2002). Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. Nucleic Acids Res *30*, 5229-5243.

McCraith, S. M., and Phizicky, E. M. (1990). A highly specific phosphatase from Saccharomyces cerevisiae implicated in tRNA splicing. Mol Cell Biol *10*, 1049-1055.

McCraith, S. M., and Phizicky, E. M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. J Biol Chem *266*, 11986-11992.

Miao, F., and Abelson, J. (1993). Yeast tRNA-splicing endonuclease cleaves precursor tRNA in a random pathway. J Biol Chem *268*, 672-677.

Nasr, F., and Filipowicz, W. (2000). Characterization of the Saccharomyces cerevisiae cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1",2"-cyclic phosphate. Nucleic Acids Res *28*, 1676-1683.

Ng, D. T., Spear, E. D., and Walter, P. (2000). The unfolded protein response regulates multiple aspects of secretory and membrane protein biogenesis and endoplasmic reticulum quality control. J Cell Biol *150*, 77-88.

Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983). Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. Cell *32*, 525-536.

Phizicky, E. M., Consaul, S. A., Nehrke, K. W., and Abelson, J. (1992). Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. J Biol Chem *267*, 4577-4582.

Phizicky, E. M., and Greer, C. L. (1993). Pre-tRNA splicing: variation on a theme or exception to the rule? Trends Biochem Sci *18*, 31-34.

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986). Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. J Biol Chem *261*, 2978-2986.

Rauhut, R., Green, P. R., and Abelson, J. (1990). Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. J Biol Chem *265*, 18180-18184.

Redding, K., Holcomb, C., and Fuller, R. S. (1991). Immunolocalization of Kex2 protease identifies a putative late Golgi compartment in the yeast Saccharomyces cerevisiae. J Cell Biol *113*, 527-538.

Ruegsegger, U., Leber, J. H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103-114.

Schwartz, R. C., Greer, C. L., Gegenheimer, P., and Abelson, J. (1983). Enzymatic mechanism of an RNA ligase from wheat germ. J Biol Chem *258*, 8374-8383.

Shuman, S. (1996). Closing the gap on DNA ligase. Structure *4*, 653-656.

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response [see comments]. Cell *87*, 405-413.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell *90*, 1-20.

Spinelli, S. L., Consaul, S. A., and Phizicky, E. M. (1997). A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. Rna *3*, 1388-1400.

Spinelli, S. L., Malik, H. S., Consaul, S. A., and Phizicky, E. M. (1998). A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. Proc Natl Acad Sci U S A *95*, 14136-14141.

Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249-258.

Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. Cell *89*, 849-858.

van Tol, H., Gross, H. J., and Beier, H. (1989). Non-enzymatic excision of pre-tRNA introns? Embo J *8*, 293-300.

Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K. W. (1997). Characterization of the yeast transcriptome. Cell *88*, 243-251.

Weber, U., Beier, H., and Gross, H. J. (1996). Another heritage from the RNA world: self-excision of intron sequence from nuclear pre-tRNAs. Nucleic Acids Res *24*, 2212-2219.

Westaway, S. K., and Abelson, J. (1995). Splicing of tRNA Precursors. In tRNA : structure, biosynthesis, and function, D. Söll, and U. RajBhandary, eds. (Washington, D.C., ASM Press), pp. 79-92.

Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990). Domain structure in yeast tRNA ligase. Biochemistry *29*, 6132-6138.

Yukawa, Y., Fan, H., Akama, K., Beier, H., Gross, H. J., and Sugiura, M. (2001). A tobacco nuclear extract supporting transcription, processing, splicing and modification of plant intron-containing tRNA precursors. Plant J *28*, 583-594.

Zillman, M., Gorovsky, M. A., and Phizicky, E. M. (1992). HeLa cells contain a 2'-phosphate-specific phosphotransferase similar to a yeast enzyme implicated in tRNA splicing. J Biol Chem *267*, 10289-10294.

Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1991). Conserved mechanism of tRNA splicing in eukaryotes. Mol Cell Biol *11*, 5410-5416.

Zofallova, L., Guo, Y., and Gupta, R. (2000). Junction phosphate is derived from the precursor in the tRNA spliced by the archaeon Haloferax volcanii cell extract. Rna *6*, 1019-1030.

## Chapter 4

# *In search of rlg1-100 suppressors:*

# *A genetic screen*

## *INTRODUCTION*

The unfolded protein response (UPR) is an intracellular signal transduction pathway that regulates the protein folding capacity of the endoplasmic reticulum (ER) (Kaufman, 2002; Patil and Walter, 2001; Sidrauski et al., 1998). The accumulation of unfolded and misfolded proteins in the ER leads to the transcriptional upregulation of genes encoding ER chaperones such as Bip and protein disulfide isomerase (PDI). In yeast, ER chaperone gene expression is activated by the bZip transcription factor Hac1p. Regulation of Hac1p activity determines when the UPR is on or off.

Hac1p expression is regulated at the translational level. Though steady-state *HAC1* mRNA levels are generally equivalent whether or not the UPR is turned on (Travers et al., 2000), Hac1p is only expressed during UPR induction. *HAC1* mRNA translation is blocked by interactions between the *HAC1* 5'UTR and a small intron located at the 3' end of the Hac1p coding region (Ruegsegger et al., 2001). The block is relieved and translation resumes upon removal of the intron by the combined actions of the endoribonuclease Ire1p and tRNA ligase (Sidrauski and Walter, 1997).

Mechanistically, *HAC1* mRNA splicing most resembles the splicing of pre-tRNAs. Cleavage of both RNA substrates by their respective endoribonucleases generates 5'-hydroxyls and 2',3' cyclic phosphates and ligation proceeds via the same multi-step, tRNA ligase catalyzed process (see chapter 2) (Gonzalez et al., 1999). However, this is where the similarities end. Ire1p and tRNA endoribonuclease lack any discernable homology and not surprisingly, recognize unrelated structural aspects of their individual RNA substrates. Moreover, whereas pre-tRNA splicing likely occurs at the nuclear pore

(Clark and Abelson, 1987), *HAC1* mRNA splicing is cytoplasmic, and likely takes place at the ER membrane (Diehn et al., 2000; Ruegsegger et al., 2001).

That tRNA ligase plays a role during the UPR was first recognized when a UPR defective allele, *rlg1-100*, was isolated in a genetic screen designed to identify components in the yeast UPR pathway (Sidrauski et al., 1996). A histidine to tyrosine change at amino acid 148 in tRNA ligase was identified and verified as the source of the UPR defect. *In vivo*, this mutation severely reduces the ability of tRNA ligase to splice *HAC1* mRNA while it has no detectable effect on pre-tRNA splicing. How this mutation reduces the ability of *rlg1-100* to ligate *HAC1* mRNA remains unresolved.

In Chapter 3 of this thesis, we describe biochemical and molecular approaches we took to uncover the basis for the *rlg1-100* ligase UPR defect. In this chapter, we describe genetic approaches taken to attack this same problem. We undertook a genetic screen designed to identify suppressor mutations that restore the ability of *rlg1-100* strains to grow under UPR inducing conditions. As happens with genetic screens, we were led down unanticipated avenues of inquiry. In our case, whereas we were unable to resolve why the *rlg1-100* tRNA ligase is deficient for *in vivo HAC1* mRNA splicing, we now know that mutations in mRNA turnover pathways can lead to *HAC1* splicing-independent Hac1p production and the ability of *rlg1-100* strains to grow under UPR inducing conditions.

## MATERIALS AND METHODS

### EMS mutagenesis

Strain TGy-1 was mutagenized with ethyl methanesulfonate (EMS) to 20% survival. Cells were plated on YPD to a density of ~300 colonies per plate and allowed to grow at 30°C for 2-3 days before replica plating to synthetic minimal dextrose (SD) plates lacking inositol and supplemented with tunicamycin (SD-ino+Tm) to a final concentration of 0.25 ug/ml. Individual colonies that grew on the tunicamycin plates were struck out onto fresh SD-ino+Tm plates to retest their phenotypes. Only those mutants that grew upon retesting were picked for further analysis.

### Plasmids and yeast strains

The plasmids used in this study and information about their construction are given in Table IV-1. The yeast strains used are listed in Table IV-2.

Pop-in/pop-out allele replacement was employed to replace wild type genes with recombinant versions at the wild type gene locus (Guthrie and Fink, 1991). To make strains TGy-1 and TGy-3, pTG-1 was linearized within *rlg1-100* and transformed into strains CSY228 and CSY227 respectively. The *rlg1-100* mutation destroys a NlaIII site. Along with lack of strain growth on UPR-inducing media, digestion of a short PCR fragment (generated using genomic DNA as template) with NlaIII was used to verify replacement of the *RLG1* gene with *rlg1-100*. To make strain TGy-113, pTG-28 was linearized within *HA-HAC1* and transformed into CSY228. To make TGy-118, pTG-29 was linearized within *HA-HAC1[G885C]* and transformed into CSY228. To make TGy-121, pTG-30 was linearized within *HA-HAC1[G1137C]* and transformed into CSY228.

Strains TGy-113, TGy-118, and TGy-121 were utilized to introduce wild type or mutant *HA-HAC1* genes into other strains by mating. The presence of the HA epitope, the 5' [G885C] splice site mutation, or the 3' [G1137C] splice site mutation was verified by restriction digestion of PCR fragments covering the sequence of interest and generated using the strain-in-question's genomic DNA as a template. This analysis was especially useful when we could not verify the presence of these sequences using a UPR growth phenotype or by western blot analysis. Addition of the HA-epitope to Hac1p creates an NdeI site, the 5' [G885C] splice site mutation destroys a CviJI site, and the 3' [G1137C] splice site mutation destroys a Hpy188I site.

To easily follow *rlg1-100* during crosses and strain construction, we made a strain in which *rlg1-100* was very tightly linked to the marker *URA3*. This strain, TGy-10.1, was made by linearizing pTG-10 within the N-terminally truncated Rlg1p coding region and integrating it at the *rlg1-100* locus in a *his3-11,-15* version of strain TGy-3. Plasmid pTG-10 does not carry *RLG1* sequences that overlapped with the *rlg1-100* allele locus.

To delete each *SKI* gene, we used a PCR-based transformation method. We amplified a DNA cassette encoding the *TPR1* marker using oligonucleotide primers, each of which bore sequences homologous to the target *SKI* gene (Longtine et al., 1998). PCR products were transformed into diploid yeast, and the yeast plated onto minimal media lacking tryptophan. Accurate deletion of each gene was verified by PCR using genomic DNA as a template. Haploid deletion strains were obtained upon sporulation of the diploid strains.

Standard recipes for YPD, and synthetic minimal (SDC) media were used (Guthrie and Fink, 1991). Tunicamycin was added to plates within hours of plating cells by top spreading 100 µl of 75 µg/ml solution of tunicamycin dissolved in DMSO to give a final concentration of 0.25 µg/ml tunicamycin in the media on the plates.

### *Northern and Western Blot Analysis*

Total yeast RNA was isolated and analyzed by Northern blot using the method of Ruegsseger et al (Ruegsegger et al., 2001). The *HAC1* probe used in all the experiments was generated against the *HAC1* 5' exon sequences. The *SCR1* probe was genereated against the full-length *SCR1* RNA.

Yeast whole cell protein extracts were performed as previously described (Ruegsegger et al., 2001). Western blot analysis of HA-tagged Hac1p was performed using monoclonal antibody HA.11 (Santa Cruz Biotechnology, Inc).

## ***RESULTS***

### *Three classes of mutants identified as suppressors of rlg1-100 (SOR)*

Yeast strain TGy-1 was treated for 60 minutes with EMS, resulting in 80% cell death. To screen for growth under UPR inducing conditions, cells were plated onto YPD, grown for 2 days at 30°C and then replica plated to SDC lacking inositol and supplemented with tunicamycin to a final concentration of 0.25ug/ml tunicamycin (this media is referred to as SDC-ino+Tm). Though it has been reported that *rlg1-100* strains do not grow on plates lacking inositol (Sidrauski et al., 1996), we found it necessary to add tunicamycin to the plates to completely prevent *rlg1-100* strain growth. Tunicamycin inhibits the N-linked protein glycosylation that normally takes place in the ER, thus contributing to protein misfolding in the ER. A total of 112 of the colonies that grew robustly were picked and struck out onto SDC-ino+Tm to verify their phenotype. Of these, 70 retested for growth on SDC-ino+Tm (Figures IV-1, -2, and -3) and were further characterized by Northern blot analysis (Figures IV-5, -6, and -7) to visualize *HAC1* mRNA splicing. Mutants were organized into three groups based on their *HAC1* splicing phenotype: C1-Sor$^+$, C2-Sor$^+$, and C3-Sor$^+$ where C stands for class and Sor stands for suppressor of *rlg1-100*. C1-Sor$^+$ strains reestablished *HAC1* mRNA splicing to near wild type levels (Figures IV-4 and –5); C2-Sor$^+$ strains accumulated significant amounts of cleaved *HAC1* mRNA fragments (Figures IV-4 and –6); and C3-Sor$^+$ strains did not demonstrate significant changes in *HAC1* splicing nor in the accumulation of cleaved *HAC1* mRNA fragments (Figures IV-4 and –7).

By far, the vast majority (50/70) of Sor$^+$ mutants fell into the C3-Sor$^+$ class. These strains might represent a class of suppressor mutants which increase products normally produced downstream of *HAC1* mRNA splicing and which aid the cell in coping with the increases in misfolded proteins in the ER (Ng et al., 2000; Travers et al., 2000). This might include increases in the ER associated degradation machinery (ERAD), or ER chaperones such as Kar2p/Bip. Alternatively, any mutation leading to a decrease in the rate at which proteins are cotranslationally translocated into the ER might be expected to have a similar effect by reducing the folding load on the ER. Since *HAC1* mRNA splicing was not greatly altered in these C3-Sor$^+$ strains compared to the *rlg-100* parental strain, we did not pursue them. Instead, we focused our attention on the C1-Sor$^+$ and C2-Sor$^+$ strains.

### *C1-Sor⁺ strains*

C1-Sor⁺ strains grew well on SDC-ino+Tm (Figure IV-1) and spliced *HAC1* mRNA more robustly than the parental *rlg1-100* strain (Figures IV-4, -5, and -8). Whereas treatment of the *rgl1-100* strain TGy-1 with DTT led to < 10% *HAC1* mRNA splicing, treatment of C1-Sor⁺ strains with DTT lead to between 26% and 55% *HAC1* mRNA splicing and likely accounted for their ability to grow under conditions requiring a functional UPR. As a comparison, a wild type ligase (*RLG1*) strain splices 70% of *HAC1* mRNA (Figures IV-4, -8). Thus our C1-Sor⁺ strains were excellent candidates for the identification of suppressors of the *HAC1* mRNA splicing defect in *rlg1-100* yeast.

To assess whether the C1-Sor⁺ mutants were dominant or recessive and if they were due to a single mutation, each C1-Sor⁺ strain was mated to a parental *rlg1-100* strain. The phenotype of the diploid was determined, and the segregation of the suppressor phenotype analyzed by tetrad dissection of the spores generated from the diploid (Table IV-3). By this analysis, all ten C1-Sor⁺ strains were dominant. Of the seven that underwent tetrad analysis, all showed a 2:2 segregation of the suppressor phenotype (as accessed by the ability to grow on SDC-ino+Tm), demonstrating that the suppressor mutation was encoded in a single gene.

To determine whether the mutations responsible for the C1-Sor⁺ phenotype were intragenic (that is, mutations within the *rlg1-100* gene itself) or extragenic, we performed linkage analysis. C1-Sor⁺ strains were mated to an *rlg1-100* strain carrying *URA3* integrated just 3' of the *rlg1-100* gene. These diploids were then sporulated and tetrad analysis performed. If the suppressor phenotype was due to a mutation in *rlg1-100*, then we expected that the suppressor phenotype would fail to co-segregate (i.e. exhibit linkage) with the *URA3* gene. Conversely, if the suppressor mutation was not caused by an *rlg1-100* mutation, then we expected to see random segregation of the Sor$^+$ and Ura$^+$ phenotypes. For the seven C1-Sor$^+$ strains tested in this manner, the suppressor phenotype never co-segregated with the *URA3* marker, indicating that these strains carried suppressing mutations within the *rlg1-100* gene itself (Table IV-3). We obtained partial DNA sequences for the *rlg1-100* genes in these strains and found that none of these suppressors had reverted *rlg1-100* back to *RLG1* (Table IV-3). Only in the case of strain Sor-72.1 was a mutation identified in the *rlg1-100* gene. In Sor-72.1, glutamate 289 was mutated to a lysine. It is unclear whether this mutation is responsible for the Sor-72.1 UPR phenotype or if it is the only *rlg1-100* mutation in this strain. Currently, the entire *rlg1-100* gene in all ten C1-Sor$^+$ suppressors is being sequenced.

### *C2-Sor$^+$ strains*

Though C2-Sor$^+$ strains grew well on SDC-ino+Tm (Figure IV-2), in general, they did not splice *HAC1* mRNA to a greater extent than the parental *rlg1-100* strain (Figures IV-4, -6, and -8). Instead, upon induction of the UPR, the most striking feature was a marked increase in *HAC1* mRNA fragments corresponding to the 5'exon + intron (5'E+I) and the 5' exon (5'E).

Each C2-Sor$^+$ strain was mated to a parental *rlg1-100* strain carrying *URA3* just 3' of the *rlg1-100* gene. The phenotype of the diploid was determined, and the 2:2 segregation of the suppressor phenotype analyzed by tetrad dissection of the spores generated from the diploid (Table IV-4). By this analysis, seven C2-Sor$^+$ strains were dominant and 2 were recessive. Of the five that underwent tetrad analysis, each showed a 2:2 segregation of the Sor$^+$:Sor$^-$ phenotype (as tested by the ability to grow on SDC-ino+Tm), demonstrating that these suppressor mutations were each located in a single gene. To determine whether the mutations responsible for the C2-Sor$^+$ phenotypes were intragenic or extragenic, we performed linkage analysis with *rlg1-100*. Diploids were sporulated and tetrad analysis performed as outlined above for the C1-Sor$^+$ strains. For the five C2-Sor$^+$ strains tested in this manner, the suppressor phenotype randomly segregated with the *URA3* marker, indicating that these C2-Sor$^+$ strains carried extragenic suppressors of the *rlg1-100* phenotype (Table IV-4). Since the recessive nature of the C2-Sor$^+$ mutations in the strains Sor-43.1 and Sor-65.1 would make them amenable to gene cloning by library transformation and phenotypic complementation, they were picked for further analysis. Because results obtained in this study with the Sor-43.1 and Sor-65.1 strains were essentially the same, in the sections to follow, for the most part, only data for Sor-43.1 will be shown and discussed.

### *Model for C2-Sor$^+$ suppression of the rlg1-100 UPR growth defect*

The accumulation of *HAC1* mRNA (5'E+I) and 5'E fragments in our C2-Sor$^+$ strains was striking. If the ability of C2-Sor$^+$ yeast to grow on UPR-inducing media was a direct consequence of the accumulation of these two mRNA fragments, then we postulated that perhaps translation of one or both of these fragments was taking place to produce some form of Hac1p.

mRNAs lacking polyA tails are rapidly degraded. However, mutations which block mRNA degradation pathways stabilize such mRNAs (Wilusz et al., 2001). In addition, polyA tails are not absolutely required for mRNA translation (Searfoss and Wickner, 2000). Thus we hypothesized that the suppressor mutations carried in our C2-Sor$^+$ mutants caused the stabilization of mRNAs lacking polyA tails and that the *HAC1* 5'exon (5'E) fragment that accumulated in our C2-Sor$^+$ mutants was being translated to produce Hac1p.

Translation of the unspliced form of *HAC1* mRNA is inhibited by basepairing interactions between the 5'UTR and the intron of the mRNA (Ruegsegger et al., 2001). These basepairing interactions are disrupted and the translational block relieved upon splicing out to of the intron when the UPR is turned on (Figure IV-9). We suspected then, that in our C2-Sor$^+$mutants, cleavage at the 5' splice site was likewise relieving the block to translation; moreover, since mRNA degradation seemed reduced in our strains, the 5'exon accumulated to the extent that it was translated to produce Hac1p. Concomitantly, translation of the 5'E+I fragment would not be expected to take place because interactions between the 5'UTR and intron remain intact, inhibiting translation (Figure IV-10).

If this were so, then relative to the Hac1p produced from spliced *HAC1* mRNA, Hac1$^i$p (where i stands for induced), the protein produced from translation of the 5'exon alone would be C-terminally truncated by 18 amino acids (Figure IV-11). This could obviously have deleterious affects on Hac1p function. However, there was evidence from the literature that such a C-terminally truncated Hac1p (Hac1pΔtail) is functional.

Hac1pΔtail induces expression of a UPR reporter gene just as well as Hac1$^i$p (Cox and Walter, 1996).

***Sor-43.1 suppressor phenotype requires cleavage at the HAC1 mRNA 5' splice site***

Based on our model, we made some testable predictions. First, upon induction of the UPR, the Sor-43.1 strain should produce a truncated version of Hac1p, Hac1pΔtail. Second, preventing cleavage at the *HAC1* 5' splice site but not the 3' splice site in the Sor-43.1 strain should prevent both production of Hac1pΔtail as well as growth on media requiring a functional UPR for yeast growth and survival.

To carry this out, we constructed Sor-43.1 strains in which the endogenous *HAC1* gene was replaced with HA-epitope tagged *HAC1* versions carrying otherwise wild type sequences, or mutations preventing Ire1p-mediated cleavage at either the 5' or 3' splice site (Gonzalez et al., 1999; Sidrauski and Walter, 1997). As controls, these same *HA-HAC1* genes were also introduced into isogenic *RLG1* and *rlg1-100* strains. We used restriction digest analysis (see Materials and Methods) and Northern blot analysis of *HAC1* mRNA splicing to verify that each strain carried the appropriate *HA-HAC1* gene. Upon UPR induction, Sor-43.1 yeast carrying the wild type *HA-HAC1* produced both the 5'E+I and 5'E fragments (Figure IV-12, strain 116). However, when cleavage was prevented by a mutation at the 5' or 3' splice site, only the 5'E+I fragment and 5'E fragment were respectively detected (Figure IV-12, strains 120 and 123). In the *RLG1* and *rlg1-100* strains, when *HAC1* mRNA splicing was inhibited by mutations in the splice junctions, the unligated fragments did not accumulate appreciably (Figure IV-12, strains 118, 121, 119, and 122).

Total cell extracts from UPR induced and uninduced strains were prepared and examined by Western blot analysis for the presence of Hac1p. As predicted, upon UPR induction, Sor-43.1 yeast produced Hac1pΔtail-S (S for Sor$^+$) that migrated faster than Hac1$^i$p and close to the position of a Hac1pΔtail marker protein (Figure IV-13, strains 116, 113, and 96). The Sor-43.1 Hac1pΔtail-S consistently ran faster than the Hac1pΔtail marker protein, and we did not resolve the cause for this discrepancy. Consistent with our model, the Sor-43.1 Hac1pΔtail-S protein was not produced in the Sor-43.1 strain carrying the 5' splice site mutant *HA-HAC1* gene (Figure IV-13, strain 120) but was produced in the strain carrying the 3' splice site mutant *HA-HAC1* gene (Figure IV-13, strain 123). Interestingly, a small amount of the Hac1pΔtail-S protein was made in the RLG1 strain carrying the *HA-HAC1* 3' splice site mutant (Figure IV-13, strain 121), indicating that the 5'E *HAC1* mRNA fragment is competent for translation even in a non-C2-Sor$^+$ strain. Also, a small amount of Hac1$^i$p was produced in the *rlg1-100* strain (Figure IV-13, strain 115). This is consistent with the ability of *rlg1-100* strains to grow weakly on media lacking inositol but not on the same media supplemented with tunicamycin. Apparently, *rlg1-100* ligase can splice just enough *HAC1* mRNA to produce the small amount of Hac1$^i$p needed to support weak growth on media lacking inositol but not enough for growth under the more demanding UPR conditions presented by the addition of tunicamycin to the media.

To assess strain growth under UPR inducing conditions, we plated serial dilutions of the Sor-43.1, *rlg1-100*, and *RLG1* strains carrying wild type or splice site mutant *HA-HAC1* genes on SDC-ino+Tm media (Figure IV-14). Each strain's growth phenotype correlated well with its ability to make some form of Hac1p. While the sor2-43.1 strains carrying either wild type or 3' splice site mutated *HA-HAC1* grew (Figure IV-14, strains 116 and 123), the strain with the 5' splice site mutant did not (Figure IV-14, strain 120). Interestingly, the 3' splice site mutant's growth was not quite as robust as that for the wild type. This might reflect residual interactions between the 5'E fragment and the Intron+3'Exon fragment facilitated not only by the intron and 5'UTR basepairing interactions (FigIV-10), but also by basepairing between the two exons (see Chapter 2, Figures 1 and 8). Following 5' splice site cleavage, this combination might help the intron maintain interactions with the 5'UTR to the extent that some translational inhibition takes place, reducing the amount of Hac1pΔtail-S made, and limiting the ability of the yeast to grow under UPR inducing conditions.

### *Sor-43.1 and Sor-65.1 are defective for non-stop mRNA degradation*

In our model for Sor-43.1 mediated suppression of the *rlg1-100* UPR defect, we asserted that the Sor-43.1 mutation was likely affecting cellular mRNA degradation. If this were true, we suspected that introduction of the Sor-43.1 mutation into a wild type ligase (*RLG1*) strain might result in accumulation of *HAC1* mRNA cleavage fragments similar to those seen in the original Sor-43.1 strain (which carries *rlg1-100*). We mated a Sor-43.1, *rlg1-100* strain to a *RLG1* strain, sporulated the resultant diploid, and by tetrad analysis, picked five spores that we predicted should carry both the Sor-43.1 mutation and *RLG1*. When induced for the UPR, four of the five haploid strains accumulated *HAC1* 5'E and 5'E+I mRNA fragments (Figure IV-15), providing more support for our mRNA degradation theory. Interestingly, this also highlighted the fact that the joining by wild type tRNA ligase of the two exons generated by Ire1p cleavage of *HAC1* mRNA is not 100% efficient. Quite a bit of the Ire1p-generated *HAC1* mRNA fragments must normally be degraded when the UPR is turned on.

As we performed our C2-Sor$^+$ experiments, a novel mechanism for the targeted destruction of anomalous yeast mRNAs was described in the literature. During nonstop mediated mRNA decay, 5'capped, 3' polyadenylated mRNAs lacking stop codons are specifically targeted for degradation (Frischmeyer et al., 2002; van Hoof et al., 2002; Vasudevan et al., 2002). The proteins required for nonstop decay overlapped with those required for antiviral protection in budding yeast. The yeast RNA LA virus produces RNA transcripts for translation that lack both a 3' polyA tail and a 5' cap. Yeast prevent viral proliferation by recognizing these mRNAs as abnormal (they lack a 5'cap and a polyA tail) and destroying them (Benard et al., 1999; Wickner, 1993; Widner and Wickner, 1993). Nonstop mediated decay and the link to proteins involved in the destruction of mRNAs lacking a polyA tail made us think of the accumulation in our C2-Sor$^+$ strains of the *HAC1* mRNA 5'E fragment which lacks both an in-frame stop codon and a 3' polyA tail. We therefore set out to determine if the Sor-43.1 and Sor-65.1 yeast were defective for nonstop mRNA decay.

Roy Parker, Ambro van Hoff and colleagues have developed a plasmid-based assay useful in the identification of yeast strains deficient for nonstop mRNA decay (van Hoof et al., 2002). They designed a *HIS3* gene in which all in-frame stop codons all the way to the end of the polyA tail are removed. When this plasmid is introduced into *his3* deletion strains, only those strains that are also defective for nonstop mRNA decay can complement the *his3* deletion. When this nonstop *his3* plasmid was introduced into our Sor-43.1 and Sor-65.1 strains, both grew on media lacking histidine; the parental *rlg1-*

*100* strain did not (Figure IV-16). With this piece of information, it appeared that a defect in the nonstop mediated mRNA degradation pathway was responsible for the C2-Sor$^+$ suppressor phenotype. The next step was to identify which component in the nonstop mRNA decay pathway was defective in our mutants.

***Genetic interactions between Sor-43.1, rlg1-100, and SKI2, SKI3, and SKI8***

The exosome is a multiprotein complex whose 3'-5' riboexonuclease activity is enlisted in the 3' end formation of numerous non-messenger RNAs, the processing of ribosomal RNAs, the destruction of left over RNA processing intermediates, and the degradation of mRNAs (Butler, 2002). The exosome exists in two forms, one cytoplasmic, the other nuclear. Each form consists of a common core packed full of 3'-5' exonucleases: six putative and three confirmed out of ten total core proteins. The nuclear and cytoplasmic exosomes differ as a consequence of the constellation of associated protein factors specific to each. Whereas the nuclear exosome associates with Rrp6p and the putative helicase Mtr4p, the cytoplasmic exosome associates with Ski7p and the *SKI* helicase complex. Deletion of *SKI7*, or any component of the *SKI* helicase complex (*SKI2, SKI3, SKI8*) cripples the nonstop mRNA decay pathway in yeast (Frischmeyer et al., 2002; van Hoof et al., 2002). Roy Parker's group has proposed a model describing how the exosome, Ski7p, and the *SKI* complex coordinate their attack on mRNAs (Figure IV-17) (Maquat, 2002). Given their central roles in nonstop mRNA decay, we directed our efforts to determine if a mutant form of Ski2p, Ski3p, Ski7p, or Ski8p was responsible for the Sor-43.1 phenotype.

Like a haploid *rlg1-100*, Sor-43.1 strain, a diploid homozygous for both these alleles grew on –ino+Tm media. However, an otherwise isogenic wild type/Sor-43.1 diploid did not. Consequently, we asked what was the phenotype of diploids made by mating an *rlg1-100*, Sor-43.1 strain to isogenic *rlg1-100* strains deleted for *SKI2*, *SKI3*, or *SKI8*. If Sor-43.1 yeast harbor a mutant form of one of these genes, then the diploid should grow on –ino+Tm media. In this case, deleting *SKI2*, *SKI3*, *or SKI8* allowed for diploid growth on the –ion+Tm media (Figure IV-18). Evidently, deletion of these *SKI* genes resulted in non-allelic non-complementation (also known as unlinked non-complementation). This genetic phenomenon occurs in diploids in which recessive mutations in two different genes unmask the phenotype of one of the parental haploid strains despite the presence of two wild type copies of the two recessive genes. Consequently, we could not conclude that the Sor-43.1 mutation was in any one of the three *SKI* genes tested. Nevertheless, since deletion of the *SKI* genes did uncover the Sor-43.1 phenotype in the doubly heterozygous diploid, it suggests that the Sor-43.1 gene product and the *SKI2*, *SKI3,* and *SKI8* gene products function *in vivo* to bring about nonstop mRNA decay. Non-allelic non-complementation genetic interactions often identify proteins that interact *in vivo* (Phizicky and Fields, 1995).

Given the nonstop decay connection, we also tested the growth phenotype of haploid *rlg1-100* strains deleted for *SKI2*, *SKI3,* or *SKI8*. We surmised that loss of any one of these components of the *SKI* complex would lead to accumulation of the *HAC1* 5'exon upon UPR induction and that this would be reflected in the ability, unlike that of the parental *rlg1-100* strain, of these strains to grow on –ino+Tm plates. Deletion of *SKI2*, *SKI3,* or *SKI8* in an *rlg1-100* strain did not recapitulate the Sor-43.1 growth phenotype. Whereas Sor-43.1 grew on –ino+Tm plates after 2 days, the strains deleted for *SKI2*, *SKI3*, or *SKI8* did not (Figure IV-19). However, weak growth of these strains was evident after 6-7 days (data not shown).

The haploid and diploid data demonstrate that though *SKI2*, *SKI3*, and *SKI8* likely functionally overlap with the Sor-43.1 gene product, whatever the recessive mutation in Sor–43.1 strains is, it leds to a much stronger UPR$^+$ phenotype than deleting any of the three *SKI* genes in a *rlg1-100* strain. Though we did not pursue Northern analysis of the three deletion *SKI* strains, given their haploid phenotypes, it is probable that they do not accumulate the *HAC1* 5'E mRNA fragment to the degree that Sor-43.1 yeast do and that this accounts for their weak growth on UPR-inducing media. This also suggests that were we to examine other nonstop decay substrates, we would expect to see the most accumulation of mRNA in Sor-43.1 compared to the *SKI* mutants.

### *Other approaches taken to identify the Sor-43.1 mutation gene*

We tried a number of other approaches to identify the gene responsible for the C2-Sor$^+$ phenotype of Sor-43.1 yeast. These included the cloning of genes identified in the literature as having cytoplasmic exosome phenotypes when mutated and various library screening attempts. Unfortunately, none of them yielded the elusive Sor-43.1 gene. Nevertheless, we deem it useful to briefly summarize these approaches and make note of ways in which they might be tried successfully in the future.

In addition to Ski7p and the genes encoding the three subunits of the *SKI* complex, Ski2p, Ski3p, and Ski8p, alleles of three genes encoding core exosome proteins, Rrp6p, Ski6p, and Ski4p, are known which cripple the cytoplasmic exosome but lack discernable affects on the nuclear exosome (Butler, 2002). Whereas Rrp4p and Ski6p are known 3'-5' riboexonucleases, Ski4p's exosomal function is unclear. We used PCR to amplify the coding regions for the seven genes, *SKI2, SKI3, SKI4, SKI6, SKI7, SKI8,* and *RRP4*, from yeast genomic DNA and then subcloned each PCR product into a CEN/ARS yeast plasmid vector. All genes were under the control of their native promoters. We then transformed these plasmids into a Sor-43.1, *rlg1-100* strain and observed how they affected growth on –ino+Tm media. Since the Sor-43.1 allele was recessive, we expected that a plasmid carrying the wild type Sor-43.1 gene would inhibit the ability of the Sor2-43.1 strain to grow on –ino+Tm media. None of the plasmids we tested clearly reduced growth to that of a non-suppressed, parental *rlg1-100 strain* (data not shown). However, since we did not verify that each cloned gene was functional in our hands following PCR amplification, the inability of the cloned genes to complement the Sor-43.1 mutation could have simply been due to inactivating mutations accumulated during PCR amplification of the genes. Thus, these experiments were inconclusive. Should this approach be taken again, the activity of each subcloned gene should be verified.

In addition to gene candidate approaches, we also screened two yeast CEN/ARS libraries in search of the Sor-43.1 gene. Again, we expected that a plasmid bearing the wild type Sor-43.1 gene would inhibit the ability of the Sor-43.1 strain to grow on –ino+Tm media. This approach yielded plasmids that did not positively retest for Sor-43.1 complementation.

When screening these yeast CEN/ARS libraries, because we were looking for lack of colony growth in a sea of growing colonies, we were at a disadvantage. It would be much easier to identify positively complementing library plasmids if, instead of inhibiting yeast growth, they promoted it. To this end, we attempted to design a nonstop allele of the yeast *URA3* gene. *URA3* has the advantage that its presence in yeast can be selected for or against by plating yeast onto media lacking uracil or supplemented with the chemical 5-fluoro-orotic acid (5-FOA). Plating cells with wild type *URA3* onto 5-FOA media kills them because 5-FOA is metabolized into 5-fluoro-uracil, a nucleotide analogue toxic to yeast. Wild type yeast harboring a nonstop *ura3* gene should grow on 5-FOA containing media but not on media lacking uracil. Conversely, yeast with defects in nonstop mRNA decay bearing this nonstop *ura3* gene should not grow on 5-FOA media but instead grow on media lacking uracil. Our strategy then, was to make and introduce a nonstop *ura3* gene into our Sor-43.1 strain and then screen yeast libraries for plasmids that promoted colony growth on 5'FOA media.

To make the nonstop *ura3* gene, we took advantage of the nonstop *his3* gene given to us by Ambro van Hoof in the Parker lab (van Hoof et al., 2002). This nonstop *his3* gene lacked in-frame stop codons all the way through the 3'UTR and poly(A) tail of its mRNA. When we examined the *URA3* gene sequence, removing all in frame stop codons through the 3'UTR and poly(A) tail would have required extensive site directed mutagenesis. Instead, we opted to make an inframe fusion between the *URA3* coding region and the nonstop *his3* 3'UTR. The fusion occurred where the wild type *URA3* and *HIS3* stop codons would have been. We then integrated the plasmid with this nonstop *ura3* gene into the genome of Sor-43.1 strains and isogenic *rlg1-100* strains. Our hope was that when plated onto 5-FOA media, the Sor-43.1 strain would die and that the *rlg1-100* strain would grow. If so, we would be set to screen yeast libraries in search of plasmids that promoted growth of the Sor-43.1 strain on 5-FOA media. Indeed, the Sor-

43.1 strain carrying the nonstop *ura3* grew robustly on media lacking uracil and failed to grow on 5-FOA media. Unfortunately, the isogenic *rlg1-100* strain grew weakly on uracil lacking media. This growth was enough to kill the *rlg1-100* strain when plated on 5-FOA media. Apparently, even in a strain wild type for nonstop mRNA decay, enough of the nonstop *ura3* mRNA could be translated to make the cells unable to grow on 5-FOA.

The nonstop *ura3* approach could still be a fruitful avenue to pursue. By manipulating the 5-FOA and uracil concentrations in the 5-FOA media (since yeast that can grow on 5-FOA are uracil auxotrophs, 5-FOA media must be supplemented with uracil), conditions might be found which meet our screening needs. In addition, we integrated the nonstop *ura3* carrying plasmid at the *TRP1* genomic locus by linearizing the plasmid within the *TRP1* gene it bore. It is thus possible that the plasmid also integrated at the *ura3* site, complicating the possible phenotypic outcomes. Indeed, when different clones derived from one integration experiment were tested on media lacking uracil or supplemented with 5-FOA, we observed a noticeable difference in the growth of the supposedly isogenic clones. In the future, this might be resolved by integrating the nonstop *ura3* carrying plasmid at the *TRP1* locus in strains in which the entire *URA3* gene has been deleted.

### *Sor-43.1 and Sor-65.1 mutations are encoded by different genes*

To determine if the recessive mutations in the Sor-43.1 and Sor-65.1 strains were in the same gene, we mated the strains to each other and observed the phenotypes of the resulting diploid strain and the tetrads produced by sporulating the diploid. The diploid grew robustly on –ino+Tm media, suggesting that the alleles were in the same gene (data not shown). However, tetrad analysis revealed that different genes encoded Sor-43 and Sor-65. If the two alleles mapped to the same gene, then all the resulting spores should have grown on –ino+Tm. However, in more than half the tetrads analyzed, one spore did not grow, indicating that it carried neither the Sor-43.1 nor the Sor-65.1 mutation. Though encoded by different genes, growth of the diploid on –ino+Tm media indicated that the two gene products likely interact *in vivo* to promote nonstop mRNA decay – another case of non-allelic non-complementation. Non-allelic non-complementation genetic interactions often identify proteins that interact *in vivo* (Phizicky and Fields, 1995).

## DISCUSSION

We have isolated mutants that suppress the UPR growth defect of *rlg1-100* strains, naming them Sor for suppressor of rlg1-100. These mutants fall into three categories: C1-Sor$^+$ mutants re-establish *HAC1* mRNA splicing to near wild type levels; C2-Sor$^+$ mutants accumulate *HAC1* mRNA 5'exon and 5'E+I fragments; and C3-Sor$^+$ mutants display little change in HAC1 mRNA splicing relative to the parental *rlg1-100* strain.

### C1-Sor$^+$ strains

Seven of the ten C1-Sor$^+$ mutants mapped to the *rgl1-100* locus by linkage analysis. Interestingly, of those that were sequenced across the *rlg1-100* allelic region (6/10), none had reverted to wild type ligase at the site of the *rlg1-100* mutation. It remains to be determined what the intragenic C1-Sor$^+$ mutations are and how they change a tRNA ligase that does not splice *HAC1* mRNA *in vivo* into one that does. Given the results and model we propose in Chapter 3 of this thesis, some possibilities come to mind. First, any mutation that increases the steady state levels of the *rlg1-100* ligase would increase *HAC1* mRNA splicing. Such mutations could affect the transcription or stability of the ligase mRNA, the translation of the ligase mRNA, or the stability of the ligase protein itself. Second, if tRNA ligase is defective for *HAC1* mRNA binding, an intragenic suppressor mutation could restore the ability of ligase to bind *HAC1* mRNA. Third, if tRNA ligase is defective in its ability to recycle back to the nucleus following *HAC1* mRNA splicing in the cytoplasm, perhaps the intragenic suppressor mutation ameliorates this. For instance, if the *rlg1-100* ligase is defective for binding to a nuclear import factor, such as an importin, the intragenic suppressor mutation might restore this interaction. If the polysome experiments suggested in Chapter 3 indeed support diminished *HAC1* mRNA binding or diminished nuclear recycling by the *rgl1-100* ligase, it would be interesting to see how tRNA ligase from these intragenic C1-Sor$^+$ suppressors behave in these assays as well.

### *Recessive C2-Sor$^+$ strains*

Of the eight C2-Sor$^+$ strains we isolated, we chose to more fully characterize the two recessive mutants, Sor-43.1 and Sor-65.1. We demonstrate that translation of the *HAC1* 5' exon fragment that accumulates in these strains is responsible for their ability to grow under UPR inducing conditions. Because cleavage at the 5' splice site effectively removes the translation inhibition imposed by the *HAC1* intron, the 5'exon fragment is translated to produce a C-terminally truncated Hac1p (Hac1pΔtail-S). We also demonstrate that in these C2-Sor$^+$ strains, cleavage at the 5' splice site is required for both the production of Hac1pΔtail-S protein and growth on UPR inducing media. Furthermore, our experiments demonstrate that our suppressor strains carry mutations that reduce mRNA degradation and not mutations that solely affect the UPR.

This study is not the first demonstrating the ability of a Hac1p N-terminally truncated at the 5' splice site to function during the UPR. Cox and Walter (Cox and Walter, 1996) made a *HAC1* gene construct in which they engineered a stop codon after the last codon located immediately before the 5' splice site; actin transcriptional terminator and 3'UTR sequences were added immediately following the new stop codon. Using this construct, they showed that Hac1pΔtail could induce expression of a reporter UPRE-*lacZ* gene as well as wild type Hac1$^{i}$p produced from spliced *HAC1* mRNA. Our results add to this by demonstrating that Hac1pΔtail-S (S for Sor) production can also rescue growth of strains normally unable to grow on UPR-inducing media. When we compared the migration by SDS-PAGE of the Hac1pΔtail-S produced by our Sor-43.1 and Sor-65.1 strains with Hac1pΔtail, the Hac1pΔtail-S protein consistently migrated slightly faster than the Hac1pΔtail version. This could reflect different modes by which translational termination and ribosome release proceed when these two proteins are produced. Whereas the Hac1pΔtail protein translationally terminates at a stop codon, the C2-Sor$^{+}$ Hac1pΔtail-S protein does not. The *HAC1* 5' exon fragment from which the Hac1pΔtail-S is translated lacks a stop codon. This could be interesting to follow up on. Ribosome release and recycling following translation termination is a much better understood phenomenon in bacteria than it is in eukaryotes. Additional characterization of these two Hac1pΔtail proteins might further our understanding of this process in eukaryotes.

Mori and colleagues propose that splicing of *HAC1* mRNA does more than just relieve the translational block imposed by the intron (Mori et al., 2000). If the unspliced *HAC1* mRNA were translated, it would produce a protein, $\text{Hac1}^{u}\text{p}$ (u for uninduced or unspliced) that differs from the protein produced from the spliced *HAC1* mRNA, $\text{Hac1}^{i}\text{p}$ (i for induced). Specifically, splicing replaces the last 10 amino acids of the $\text{Hac1}^{u}\text{p}$ with a new set of C-terminal 18 amino acids in $\text{Hac1}^{i}\text{p}$ (Figure IV-11). When these two proteins were tested for their ability to activate transcription, $\text{Hac1}^{i}\text{p}$ was a much more potent transcriptional activator than $\text{Hac1}^{u}\text{p}$. Thus it was concluded that *HAC1* mRNA splicing is also critical for producing an especially potent UPR transcription factor. However, Mori and colleagues did not examine the transcriptional potency of a Hac1p version lacking either C-terminal tail. Thus our data is not inconsistent with their data. In fact, our data suggests that swapping of the 18 amino acid tail via *HAC1* mRNA splicing does not so much enhance the transcriptional transactivation abilities of Hac1p as it gets rid of a stretch of 10 amino acids that inhibits transactivation. Side-by-side comparisons of the transcriptional activities of each of these Hac1p variants should verify if this is indeed the case.

Unlike normally translated cellular mRNAs, the 5'exon fragment of *HAC1* mRNA that is translated in our C2-$\text{Sor}^{+}$ mutants lacks an in-frame stop codon and likely lacks a 3'poly(A) tail. As we undertook our experiments, the nonstop RNA mediated decay pathway was discovered and described in the literature. During nonstop mediated mRNA decay, 5'capped, 3' polyadenylated mRNAs lacking stop codons are specifically targeted for degradation in a translation dependent manner (Frischmeyer et al., 2002; van Hoof et al., 2002; Vasudevan et al., 2002). The proteins required for nonstop decay overlap with those required for antiviral protection in yeast. The yeast RNA L-A virus produces RNA messages for translation that lack 5'caps and 3'poly(A) tails. Yeast inhibit viral proliferation by recognizing these mRNAs as abnormal (they lack a 5'cap and a poly(A) tail) and destroy them (Benard et al., 1999; Wickner, 1993; Widner and Wickner, 1993). Because the translated *HAC1* 5'exon in our C2-Sor$^+$ strains lacked a stop codon and likely lacked a poly(A) tail, we investigated the possibility that these strains were defective for nonstop mediated mRNA decay. Our experiments using the nonstop *his3* reporter strongly suggest that our mutants are in this pathway. The non-allelic non-complementation phenotype observed in our diploid strains supports this as well. However, results in *rlg1-100* strains deleted for *SKI2, SKI3*, or *SKI8* - key components in the nonstop mRNA degradation pathway – are inconsistent with this. These deletion strains did not recapitulate our C2-Sor$^+$ suppressor phenotype as they only very weakly rescued the ability of an *rlg1-100* strain to grow on UPR-inducing media. Though we did not test an *rlg1-100* strain deleted for *SKI7* for growth on UPR-inducing media, we believe that it is unlikely that our C2-Sor$^+$ strains are defective for Ski7p activity; yeast deleted for *SKI2, SKI3, SKI7*, or *SKI8* stabilize nonstop mRNAs to the same extent (van Hoof et. al. 2002). We have not, however, ruled out the possibility that the C2-Sor+ strains Sor-43.1 and Sor-65.1 are defective for a core component of the cytoplasmic exosome. Perhaps our mutants represent alleles of the core exosomal components *SKI4, SKI6*, or *RRP6* that block nonstop mRNA decay more strongly than loss of *SKI7* or the *SKI* complex. Known alleles in these three genes block cytoplasmic exosome function without affecting nuclear exosome function (Butler, 2002). Alternatively, our mutants might represent alleles in genes whose role during nonstop mRNA decay has until now not been revealed.

Are there other known genes involved in mRNA degradation that when mutated might be expected to produce the C2-Sor$^+$ suppressor phenotype? Yes! Like the nonstop mRNAs studied by Parker and colleagues, the *HAC1* mRNA 5' exon lacks a stop codon; however unlike the nonstop mRNAs, the *HAC1* 5' exon probably also lacks a 3'poly(A) tail. Therefore the *HAC1* 5'exon would be primed and ready as a substrate for the 5' to 3' RNA degradation pathway, the dominant mRNA turnover pathway in yeast (Wilusz et al., 2001). The 5' to 3' degradation of RNA takes place in three stages: First, the proteins Pan2/3 and Ccr4/Caf1 remove the 3'poly(A) tail. Second, the decapping enzymes Dcp1p and Dcp2p remove the 5'cap. Third, the Xrnlp 5' to 3' exonuclease degrades the RNA. In yeast, this pathway is estimated to be 2 to 5 times faster than the 3' to 5' exonucleolytic digestion catalyzed by the exosome (Cao and Parker, 2001; Muhlrad et al., 1995). *XRN1* is also known by the name *SKI1* (Toh et al., 1978; Toh and Wickner, 1980), having been isolated as a gene involved in the response of yeast to the RNA L-A virus. In yeast defective for Xrnlp, decapped and deadenylated mRNAs are polysome associated, suggesting that they are translated (Hsu and Stevens, 1993). Thus it is possible that mutations affecting degradation at the 5' end of mRNA could lead to stabilization and translation of the 5exon fragment of *HAC1* mRNA. Perhaps our C2-Sor$^+$ strains carry mutations in the exoribonuclease Xrnlp, or the decapping enzymes Dcp1p and Dcp2p (Steiger et al. 2003). In addition, mutations in proteins known to stimulate 5' decapping could be involved. The Pat1p/Lsm-1 complex (Bonnert et al. 2000; Bouveret et al. 2000;

Tharun et al. 2000), the putative helicase Dhh1p (Fischer and Weis, 2002) and the RNA binding proteins Edc1p and Edc2p (Schwartz et al. 2003; Steiger et al. 2003) all stimulate mRNA decapping. Alternatively, mutations that increase binding of the translation initiation factor eIF4E might also be involved. Binding of eIF4E not only stimulates translation, it also inhibits 5'decapping and thus mRNA degradation (Schwartz and Parker 1999; Schwartz and Parker 2000; Ramirez et al. 2002).

These possibilities lead to a number of predictions. First, if the C2-Sor$^+$ strains are defective for Xrn1p activity, then we should see an accumulation of mRNAs lacking both 5' caps and 3' poly(A) tails. Second, if instead, the C2-Sor$^+$ strains are defective for 5' decapping, then we should see an accumulation of 5'capped mRNAs lacking 3'poly(A) tails.

If our C2-Sor$^+$ strains are defective for 5' to 3' mRNA decay instead of nonstop decay, how do we explain the results we obtained with the nonstop *his3* constructs? Our results are not necessarily incompatible with defects in the 5' to 3'degradation pathway. The nonstop-mediated degradation pathway recognizes mRNAs lacking stop codons as aberrant, and then starts degrading them at the 3' end. Perhaps mRNAs that start being degraded by the 3'-5' exosome can be finished off by the 5' to 3' mRNA decay pathway, the major mRNA degradation pathway in yeast. There is, however, a major caveat to invoking the 5'-3' mRNA degradation pathway in explaining our results: yeast deleted for the 5'-3' exonuclease *XRN1* and which carry the nonstop *his3* plasmid are His$^-$, suggesting that blocking 5'-3' mRNA degradation does not produce defects in nonstop mRNA decay (Ambro van Hoof, personal communication). We need to verify that if this is case in the yeast strain background used in our experiments.

***Dominant C2-Sor$^+$ strains***

It is intriguing that the majority (7/9) of our C2-Sor$^+$ mutants were dominant. It would be interesting to determine if like the Sor-43.1 and Sor-65.1 strains, these strains produce Hac1pΔtail-S, require its production in order to grow on UPR-inducing media, and if they are defective for nonstop mRNA decay using the nonstop *his3* reporter plasmid. Nonstop mediated decay involves at least three large protein complexes: the *SKI* complex, the cytoplasmic exosome, and the ribosome. Ski7p interacts with the *SKI* and exosome complexes and is thought to bind the ribosomal A-site (Araki et. al 2001; van Hoof et. al. 2002). Given all these protein-protein interactions, our dominant C2-Sor$^+$ strains might represent dominant-negative mutations in genes encoding components of these complexes; these would be expected to retain the ability to interact with other proteins while at the same time altering the function of the complex as a whole.

A dominant-negative allele of the ribosomal protein L3 exits which increases levels of the yeast viral L-A dsRNA (Peltz et. al. 1999). L-A dsRNA abundance is proportional to the abundance of L-A mRNA. Thus, the L3 dominant-negative allele must also increase the levels of the L-A mRNA, an mRNA lacking both a 5'cap and a 3'poly(A) tail. A portion of the L3 ribosomal protein lies near the peptidyl-transferase center and the A-site (Puglisi et. al. 2000). Binding of Ski7p to the A-site would place it near L3. Thus, perturbations in L3 structure could lead to perturbations in Ski7p binding and might reduce nonstop mRNA decay as well as L-A mRNA decay. It is tempting to speculate that some of our C2-Sor$^+$ dominant alleles might likewise involve ribosomal proteins.

## *Table IV-1 Chapter 4 plasmid list*

***Plasmids***

| **Plasmid** | **Yeast marker** | **CEN, 2µ or integrating** | **Cloned gene** | **Comments** |
|---|---|---|---|---|
| pTG-1 | URA3 | integrating | *rlg1-100* | subcloned *rlg1-100* from pCF158 (Sidrauski et al. 1998) into pRS306 |
| pTG-10 | URA3 | integrating | N-terminally truncated *RLG1* | parental plasmd pRS306 carries *RLG1* aa585-827 plus 3'UTR |
| pTG-25 | TRP1 | CEN | *HA-HAC1 Δtail* | subcloned *HA-HAC1Δtail* from pJC 836 (Cox and Walter 1996) into pRS314 |
| pTG-28 | URA3 | integrating | *HA-HAC1* | subcloned *HA-HAC1* from pJC321 into pRS306 |
| pTG-29 | URA3 | integrating | *HA-HAC1* 5'ss G885C | subcloned *HA-HAC1* 5'ss G855C from pCF205 into pRS306 |
| pTG-30 | URA3 | integrating | *HA-HAC1* 3'ss G1137C | subcloned *HA-HAC1* 3'ss G1137C from pCF203 into pRS306 |
| pAV188 | URA3 | CEN | nonstop *his3* | parental plasmid pRS416 a gift from A. von Hoff |
| pRS306 | URA3 | integrating | none | from Sikorski and Hieter (1989) |
| pRS314 | TRP1 | CEN | none | from Sikorski and Hieter (1989) |
| pRS315 | LEU2 | CEN | none | from Sikorski and Hieter (1989) |
| pRS316 | URA3 | CEN | none | from Sikorski and Hieter (1989) |
| pRS416 | URA3 | 2µ | none | from Sikorski and Hieter (1989) |

**Table IV-2 Chapter 4 yeast strain list**

| *Strain* | *Genotype* |
|---|---|
| W303-1A | *MATa; leu2-3,-112; ura3-1,-112; his3-11,-15; trp1-1; ade2-1; can1-100* |
| CSY227 | same as W303-1A, except *his3-11,-15::HIS3-UPRE-lacZ.* CSY227 also known as PWY374 |
| CSY228 | same as CSY227, except *MAT∝*. CSY228 also known as PWY373 |
| SKI2 | *MATa, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0* |
| Δski2 | same as SKI2 except *Δski2::NEO*$^R$ |
| TGy-1 | same as CSY228, except *RLG1* replaced by *rlg1-100* using pTG-1 (see methods) |
| TGy-3 | same as CSY227, except *RLG1* replaced by *rlg1-100* using pTG-1 (see methods) |
| TGy-10.1 | same as TGy-3, except *his3-11,-15, rlg1-100::URA3.*<br>*N*ote: *URA3* inserted 3' of *rlg1-100* for linkage analysis using pTG-10 |
| TGy-16 | diploid of cross CSY227 & CSY228 |
| TGy-41 | same as TGy-3, except *Sor-43. T*his is a spore from mating TGy-10.1& original *Sor-43* strain |
| TGy-43 | same as TGy-10.1, except *MAT∝, Sor-43, his3-11,-15::HIS3-UPRE-lacZ.*<br>This is a spore from mating TGy-10.1 & original *Sor-43* strain |
| TGy-61 | diploid of cross TGy-43 & TGy-41, carries pRS315 also |
| TGy-62 | diploid of cross TGy-3 & TGy-43, carries pRS315 also |
| TGy-88 | diploid W303 except *rlg1-100/rlg1-100,* pRS314, & pRS316 |
| TGy-96 | same as CSY228 except *Δhac1::LEU2,* & pTG-25 |
| TGy-113 | same as CSY228 except genomic *HAC1* replaced with *HA-HAC1* using pTG-28 (see methods) |
| TGy-115 | same as TGy-113 except *rlg1-100::URA3* (see TGy-10.1) |
| TGy-116 | same as TGy-115 except *Sor-43* |
| TGy-117 | same as TGy-115 except *Sor-65* |
| TGy-118 | same as TGy-113, except *HA-HAC1[G885C]* using pTG-29. G885 is the *HAC1* 5' splice site |
| TGy-119 | same as TGy-118, except *MATa, rlg1-100::URA3* (see TGy-10.1) |
| TGy-120 | same as TGy-119, except *MAT∝, sor2-43* |
| TGy-121 | same as TGy-113, except *HA-HAC1[G1137C]* using pTG-30. G1137 is the *HAC1* 3' splice site |
| TGy-122 | same as TGy-121, except *rlg1-100::URA3* (see TGy-10.1) |
| TGy-123 | same as TGy-122, except *Sor-43* |
| TGy-127 | same as W303-1A except *rlg1-100::URA3* (see TGy-10.1), *Δski8::TRP1* |
| TGy-128 | same as W303-1A except *rlg1-100::URA3* (see TGy-10.1), *Δski3::TRP1* |
| TGy-129 | same as W303-1A except *rlg1-100::URA3* (see TGy-10.1), *Δski2::TRP1* |
| TGy-131 to 134 | same as TGy-113 except *Sor-43, &* mating type unknown, TGy-131= clone 9C,<br>TGy-132= clone 10A, TGy-133= clone 12C, TGy-134= clone 12A. |

***Table IV-3 Linkage analysis for C1-Sor⁺ strains***

| Sor strain | Dom or Rec | *URA3* rlg1-100 linkage | Sor⁺ phenotype segregation | |
|---|---|---|---|---|
| **5.1** | dom | linked | 2:2 | * |
| **8.1** | dom | linked | 2:2 | * |
| **59.2** | dom | na | na | |
| **61.1** | dom | linked | 2:2 | * |
| **62.2** | dom | linked | 2:2 | * |
| **63.1** | dom | na | na | |
| **72.1** | dom | linked | 2:2 | * # |
| **96.1** | dom | linked | 2:2 | * |
| **302D1.1** | dom | linked | 2:2 | |
| **303D3.1** | dom | na | na | |

na = not available

* = DNA sequencing verified did not revert to *rlg1-100* at position 148.
Note *RLG1* = H148 and *rlg1-100* = Y148.

# = 72.1 mutation E289K

### *Table IV-4 Linkage analysis for C2-Sor$^+$ strains*

| **Sor strain** | Dom or Rec | *URA3* rlg1-100 linkage | Sor$^+$ phenotype segregation |
|---|---|---|---|
| **11.1** | dom | not linked | 2:2 |
| **16.1** | dom | na | na |
| **36.1** | dom | na | na |
| **41.1** | dom | not linked | 2:2 |
| **43.1** | rec | not linked | 2:2 |
| **65.1** | rec | not linked | 2:2 |
| **602D**19.2 | dom | na | na |
| **603D**5.1 | dom | na | na |
| **902D**1.1 | dom | not linked | 2:2 |

**na** = not available

***Figure IV-1 C1-Sor$^+$ strain growth on UPR inducing media***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. SDC, synthetic dextrose complete media; SDC-ino+Tm, synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 μg/ml. The Sor strain number of each strain is indicated on the left. *RLG1* and *rlg1-100* designate strains carrying wild type or UPR-defective tRNA ligase respectively.

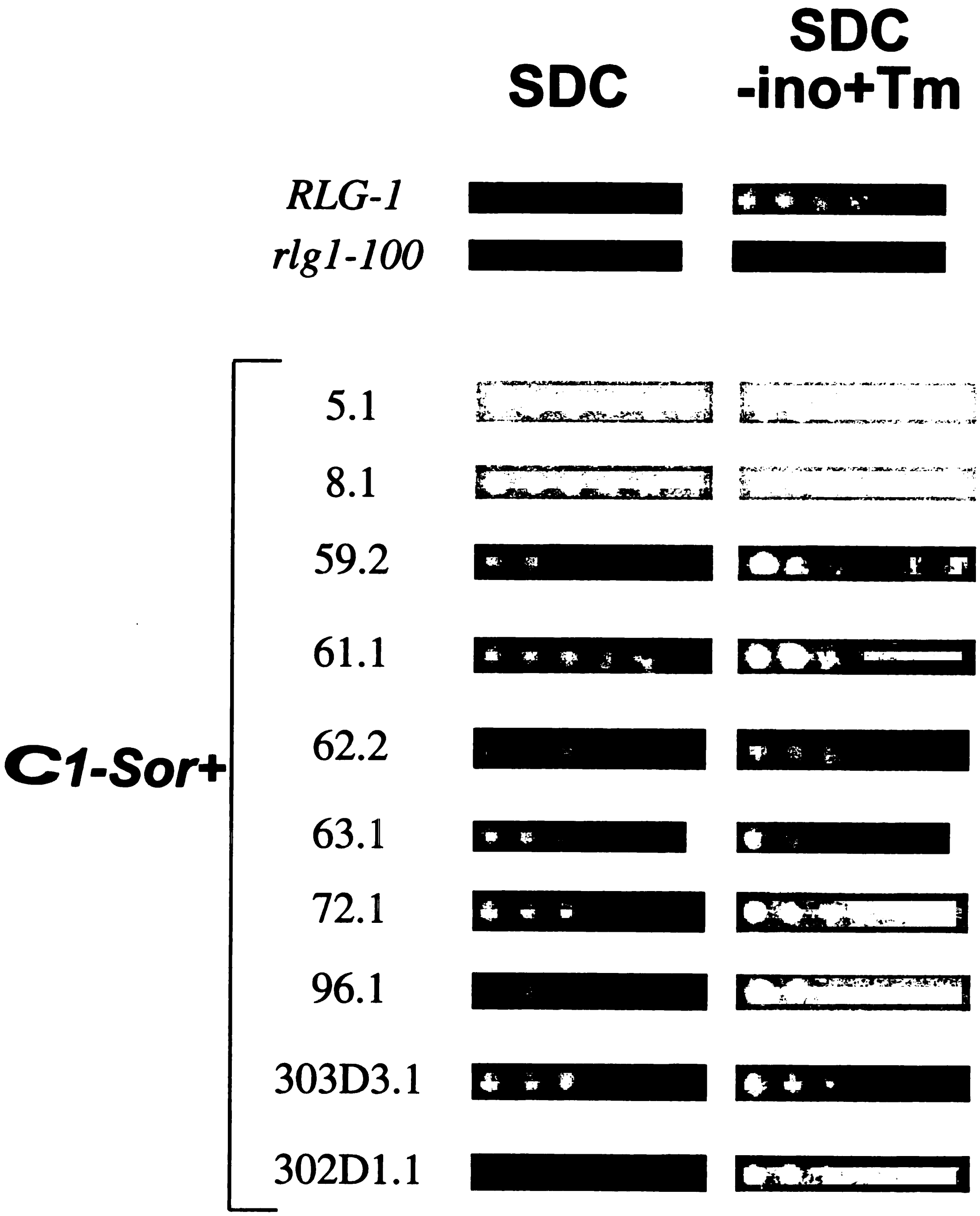

SDC
SDC -ino+Tm
RLG-1
rlg1-100
C1-Sor+
5.1
8.1
59.2
61.1
62.2
63.1
72.1
96.1
303D3.1
302D1.1

***Figure IV-2 C2-Sor⁺ strain growth on UPR inducing media***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. SDC, synthetic dextrose complete media; SDC-ino+Tm, synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 μg/ml. The Sor strain number of each strain is indicated on the left. *RLG1* and *rlg1-100* designate strains carrying wild type or UPR-defective tRNA ligase respectively.

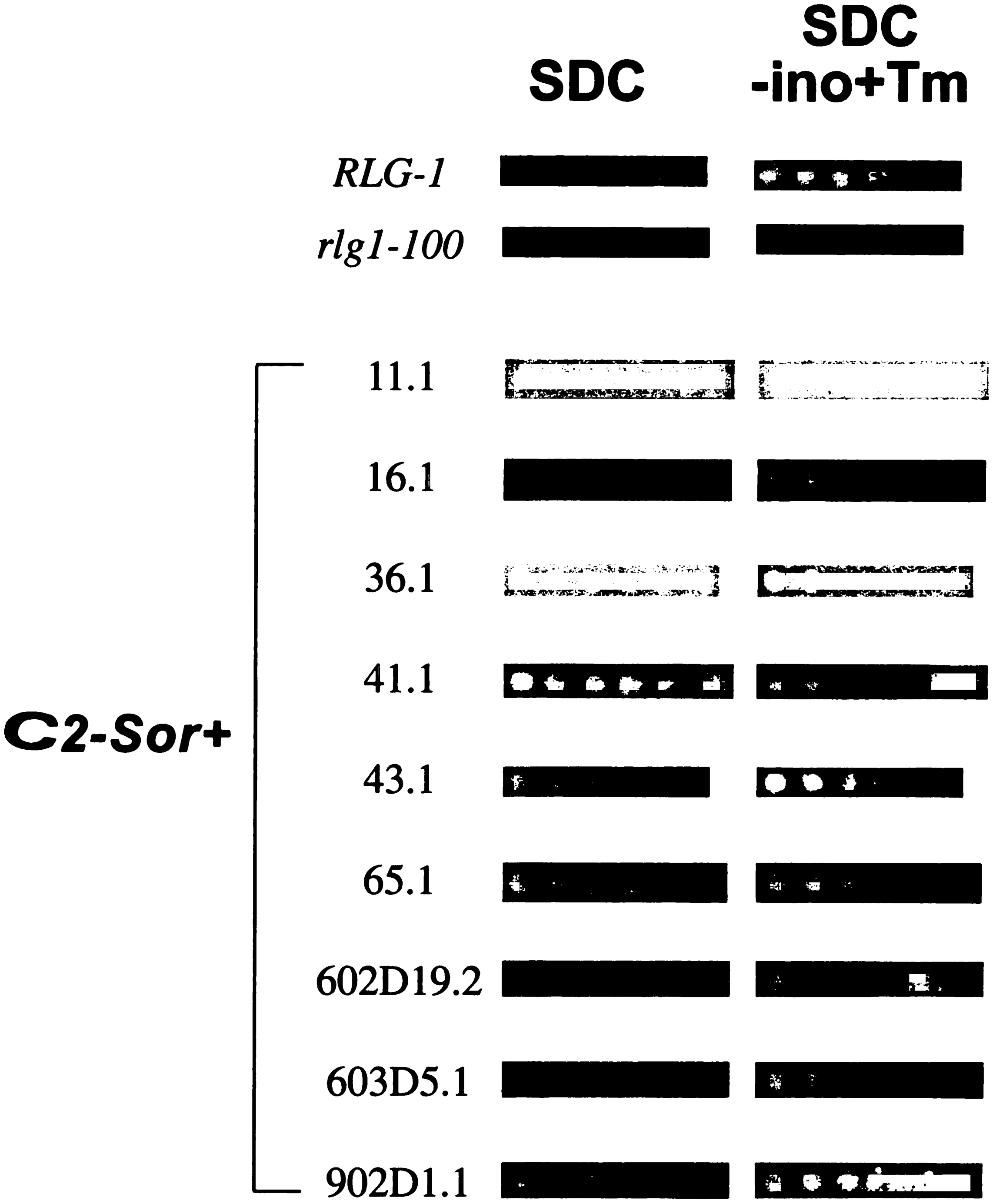

SDC
SDC -ino+Tm
*RLG-1*
*rlg1-100*
11.1
16.1
36.1
41.1
43.1
65.1
602D19.2
603D5.1
902D1.1
*C2-Sor+*

***Figure IV-3 C3-Sor⁺ strain growth on UPR inducing media***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. SDC, synthetic dextrose complete media; SDC-ino+Tm, synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 µg/ml. The Sor strain number of each strain is indicated on the left. *RLG1* and *rlg1-100* designate strains carrying wild type or UPR-defective tRNA ligase respectively. Note, only 11 out of the 50 total C3-Sor⁺ strains are shown.

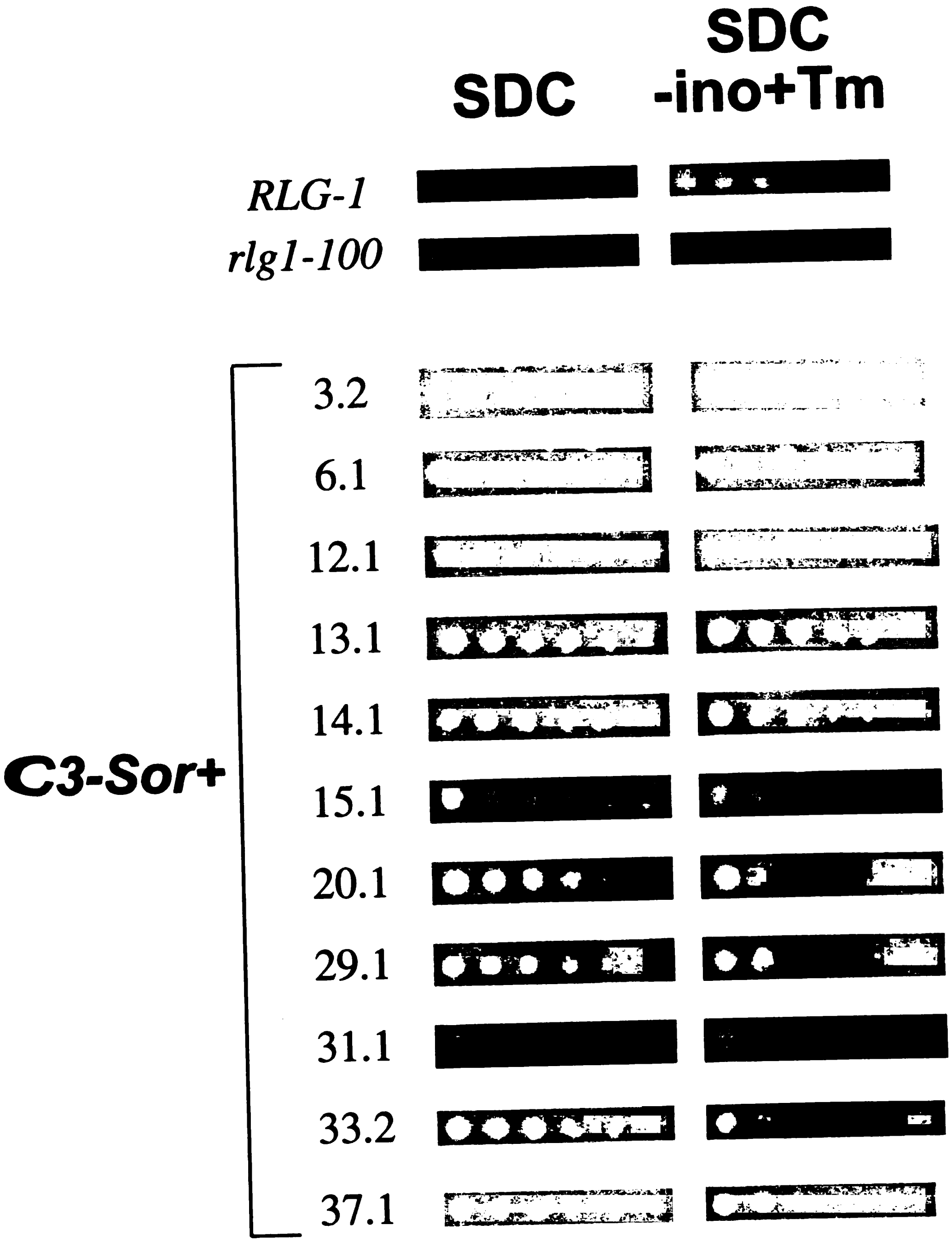

SDC
SDC -ino+Tm
RLG-1
rlg1-100
C3-Sor+
3.2
6.1
12.1
13.1
14.1
15.1
20.1
29.1
31.1
33.2
37.1

***Figure IV-4 Northern blot analysis of HAC1 mRNA splicing in RLG1 and rlg1-100 strains***

Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. Unspliced, spliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are indentified along the left hand side of the blot.

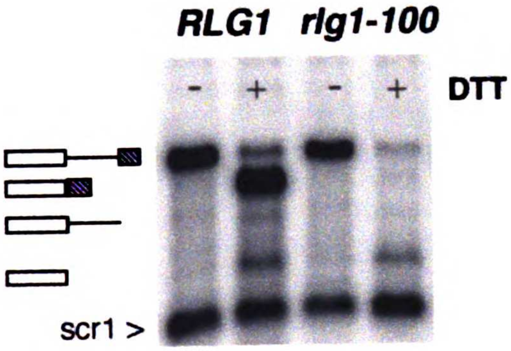

RLG1
rlg1-100
-
+
-
+
DTT
scr1 >

***Figure IV-5 Northern blot analysis of HAC1 mRNA splicing in C1-Sor$^+$ strains.*** Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. Unspliced and spliced *HAC1* mRNA species are indentified along the left hand side of the blot. Sor strain numbers are given above each Northern blot.

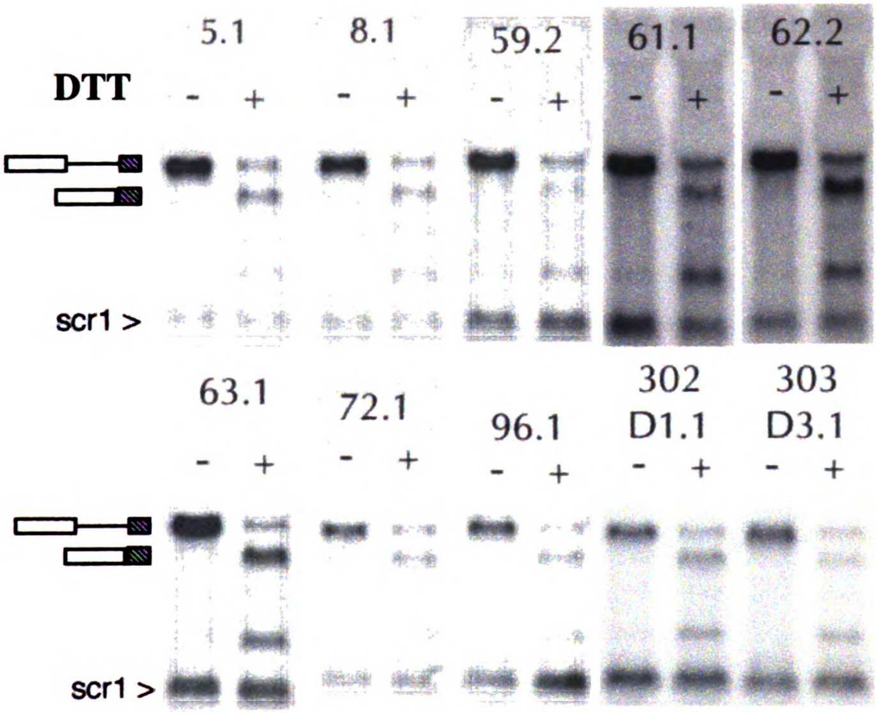

DTT
5.1
8.1
59.2
61.1
62.2
- +
scr1 >
63.1
72.1
96.1
302 D1.1
303 D3.1
- +
scr1 >

***Figure IV-6 Northern blot analysis of HAC1 mRNA splicing in C2-Sor$^+$ strains.*** Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. Unspliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are identified along the left hand side of the blot. Sor strain numbers are given above each Northern blot.

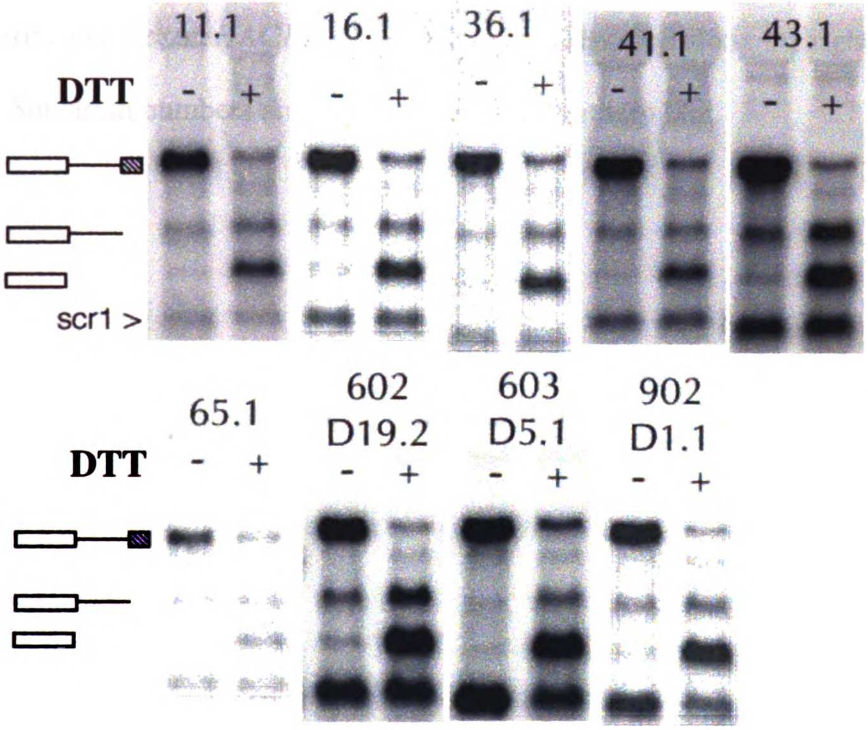

11.1
16.1
36.1
41.1
43.1
DTT
- +
- +
- +
- +
- +
scr1 >
65.1
602 D19.2
603 D5.1
902 D1.1
DTT
- +
- +
- +
- +

***Figure IV-7 Northern blot analysis of HAC1 mRNA splicing in C3-Sor$^+$ strains.*** Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. Unspliced, spliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are identified along the right hand side of the blot. Sor strain numbers are given above each Northern blot.

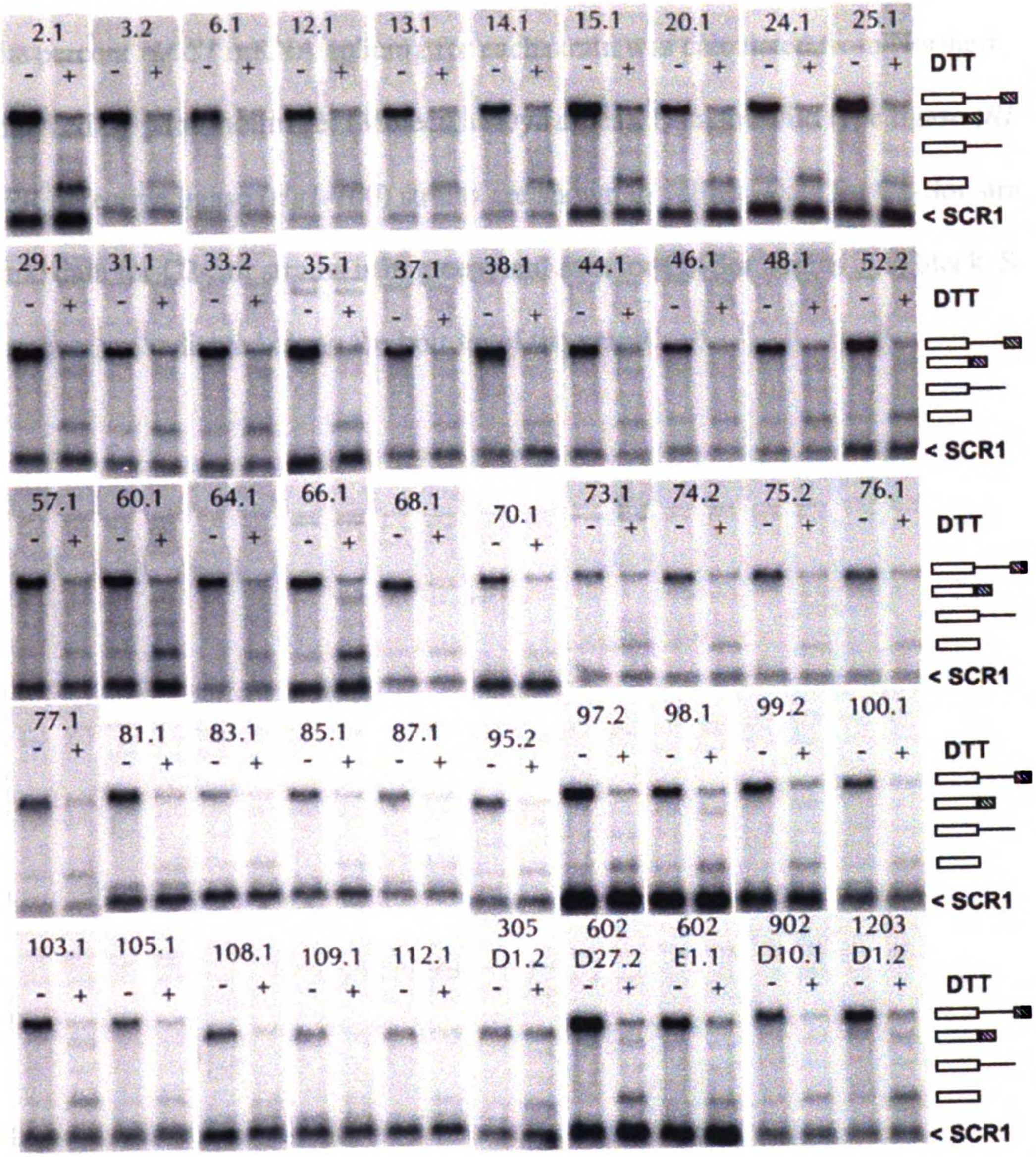

2.1
3.2
6.1
12.1
13.1
14.1
15.1
20.1
24.1
25.1
DTT
< SCR1
29.1
31.1
33.2
35.1
37.1
38.1
44.1
46.1
48.1
52.2
DTT
< SCR1
57.1
60.1
64.1
66.1
68.1
70.1
73.1
74.2
75.2
76.1
DTT
< SCR1
77.1
81.1
83.1
85.1
87.1
95.2
97.2
98.1
99.2
100.1
DTT
< SCR1
103.1
105.1
108.1
109.1
112.1
305 D1.2
602 D27.2
602 E1.1
902 D10.1
1203 D1.2
DTT
< SCR1

***Figure IV-8 Calculated percent HAC1 mRNA splicing for RLG1, rgl1-100, C1-Sor⁺, C2-Sor, and C3-Sor strains***

The percent *HAC1* mRNA splicing for each strain was calculated from Northern blot data analyzed by phosphoimager (Molecular Dynamics). Data for wild type ligase (*RLG1*) and UPR defective ligase (*rlg1-100*) strains are shown in yellow. Data for C1-Sor strains is in pink; data for C2-Sor strains is in green; and data for C3-Sor strains is in black. Sor strain numbers are indicated along the bottom of the graph.

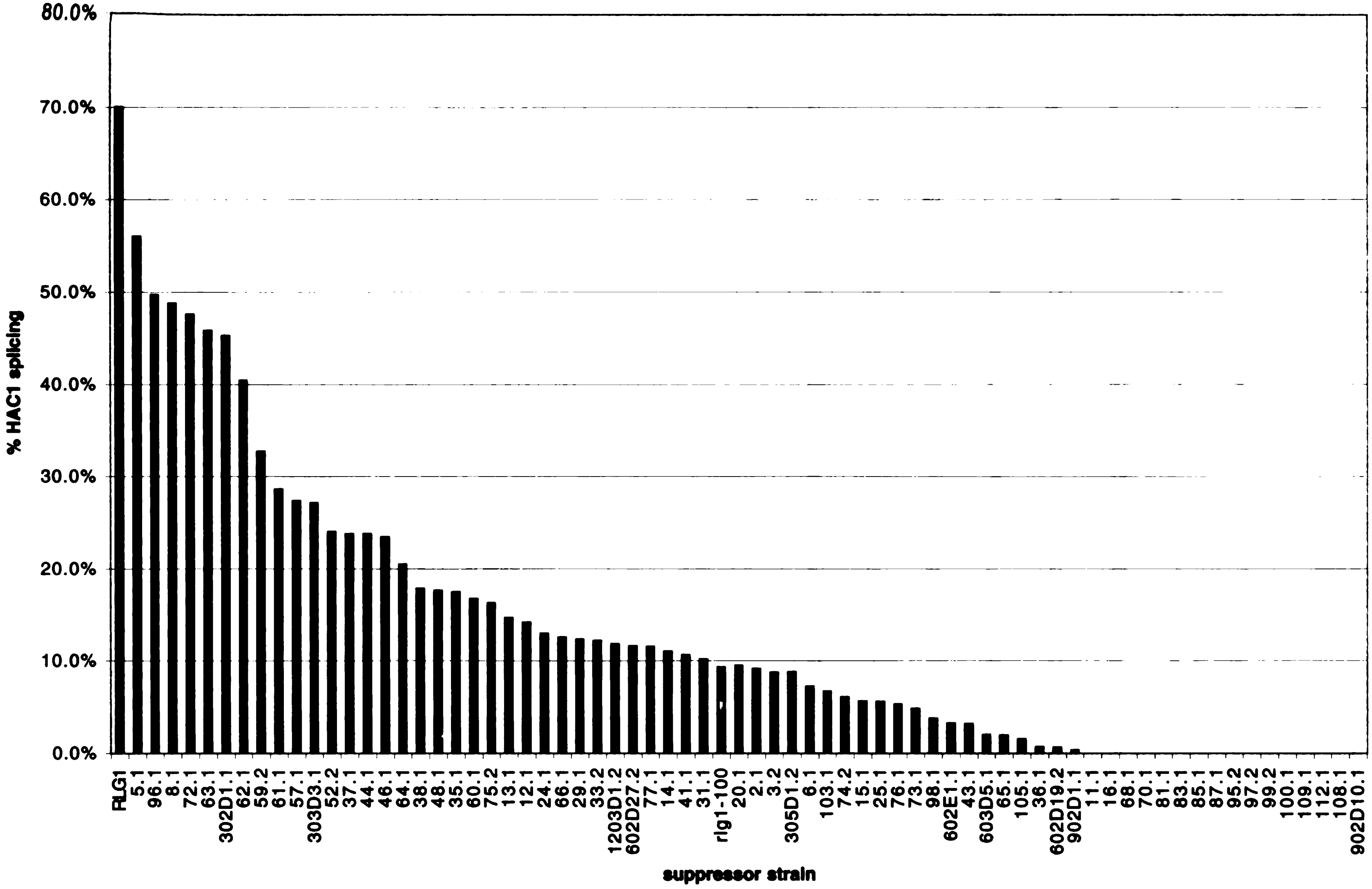

80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
% HAC1 splicing
RLG1
5.1
96.1
8.1
72.1
63.1
302D1.1
62.1
59.2
61.1
57.1
303D3.1
52.2
37.1
44.1
46.1
64.1
38.1
48.1
35.1
60.1
75.2
13.1
12.1
24.1
66.1
29.1
33.2
1203D1.2
602D27.2
77.1
14.1
41.1
31.1
rlg1-100
20.1
2.1
3.2
305D1.2
6.1
103.1
74.2
15.1
25.1
76.1
73.1
98.1
602E1.1
43.1
603D5.1
65.1
105.1
36.1
602D19.2
902D1.1
11.1
16.1
68.1
70.1
81.1
83.1
85.1
87.1
95.2
97.2
99.2
100.1
109.1
112.1
108.1
902D10.1
suppressor strain

***Figure IV-9 Model for HAC1 mRNA splicing in wild type (RLG1) and UPR defective (rlg1-100) ligase strains***

Adapted from (Ruegsegger et al., 2001). The *HAC1* 5'exon is shown in black, the intron in red, and the 3'exon in green. Basepairing interactions between the 5' UTR and intron are shown. Upon induction of the UPR in wild type tRNA ligase strains, *HAC1* mRNA is cleaved by Ire1p and ligated by tRNA ligase to produce spliced *HAC1* mRNA. Splicing disrupts basepairing between the *HAC1* intron and 5'UTR. This relieves the translational block normally imposed by this basepairing interaction; thus the spliced *HAC1* mRNA is translated to produce Hac1p. In *rlg1-100* yeast, because tRNA ligase does not ligate the *HAC1* 5' and 3' exons produced by Ire1p cleavage, the *HAC1* fragments are degraded and not Hac1p is produced.

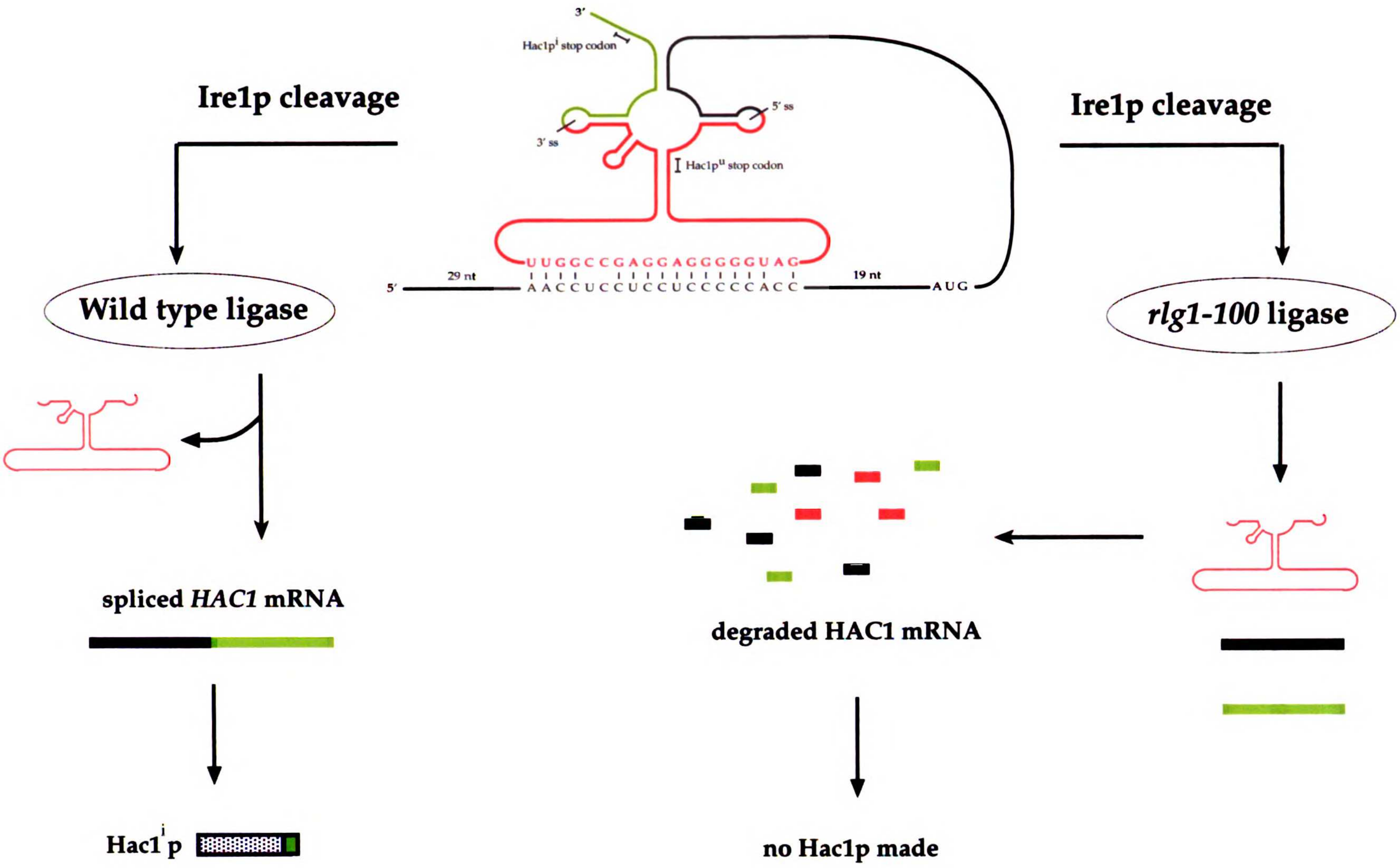

3'
Hac1p$^{i}$ stop codon
Ire1p cleavage
5' ss
3' ss
Hac1p$^{u}$ stop codon
Ire1p cleavage
UUGGCCGAGGAGGGGGUAG
AACCUCCUCCUCCCCCACC
5'
29 nt
19 nt
AUG
Wild type ligase
rlg1-100 ligase
spliced HAC1 mRNA
degraded HAC1 mRNA
Hac1$^{i}$p
no Hac1p made

***Figure IV-10 Model for HAC1 mRNA splicing in C2-Sor$^+$ strains***

Adapted from (Ruegsegger et al., 2001). The *HAC1* 5' exon is shown in black, the intron in red, and the 3'exon in green. Basepairing interactions between the 5' UTR and intron are shown. Ire1p mediated cleavage at the 5' splice site produces the 5'exon *HAC1* mRNA fragment that is free of basepairing interactions with the intron, and therefore is competent for translation. Ire1p mediated cleavage at the 3' splice site produces the (5'exon+intron) *HAC1* mRNA fragment that retains basepairing interactions with the intron and is thus translationally inhibited.

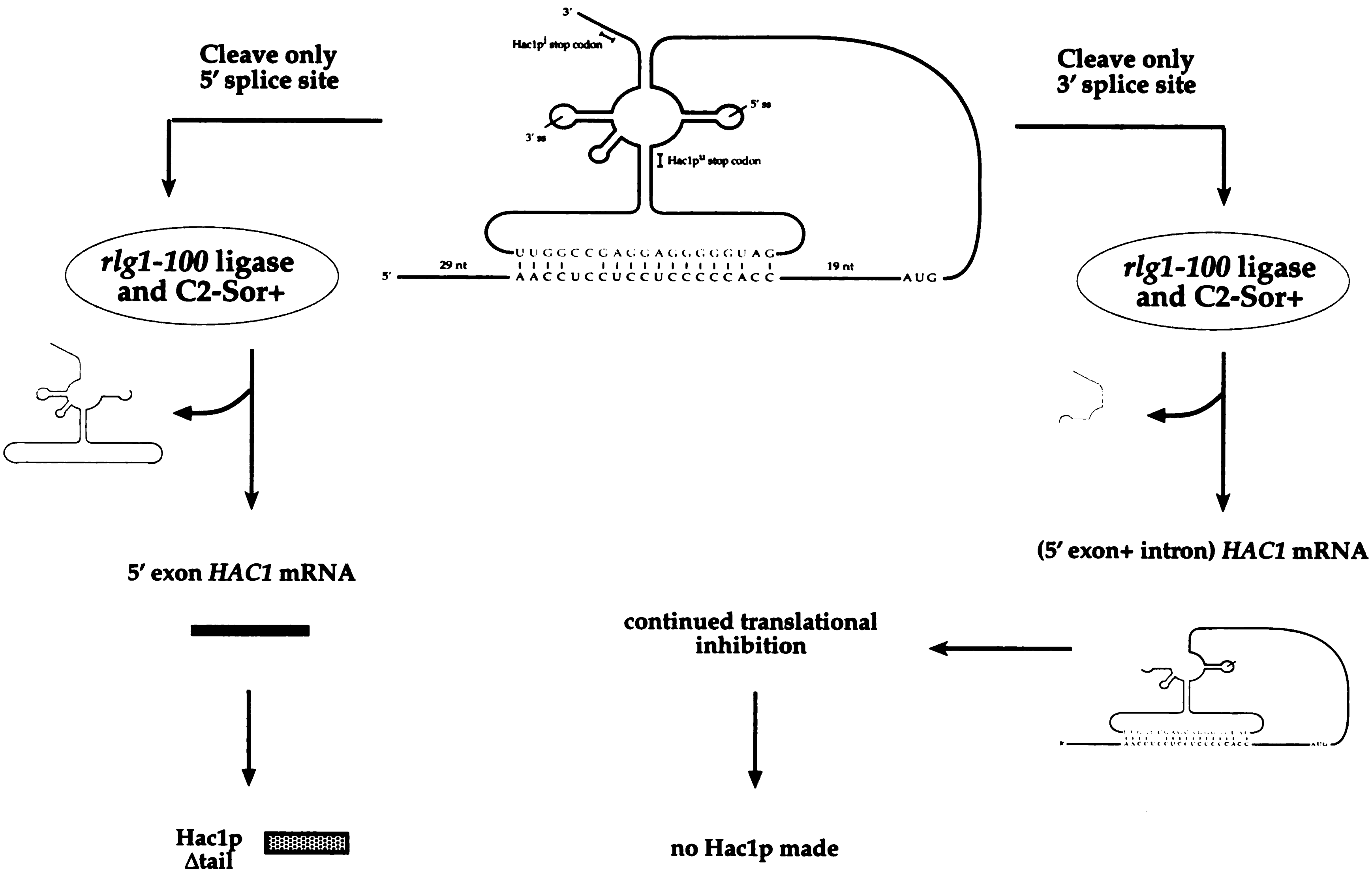
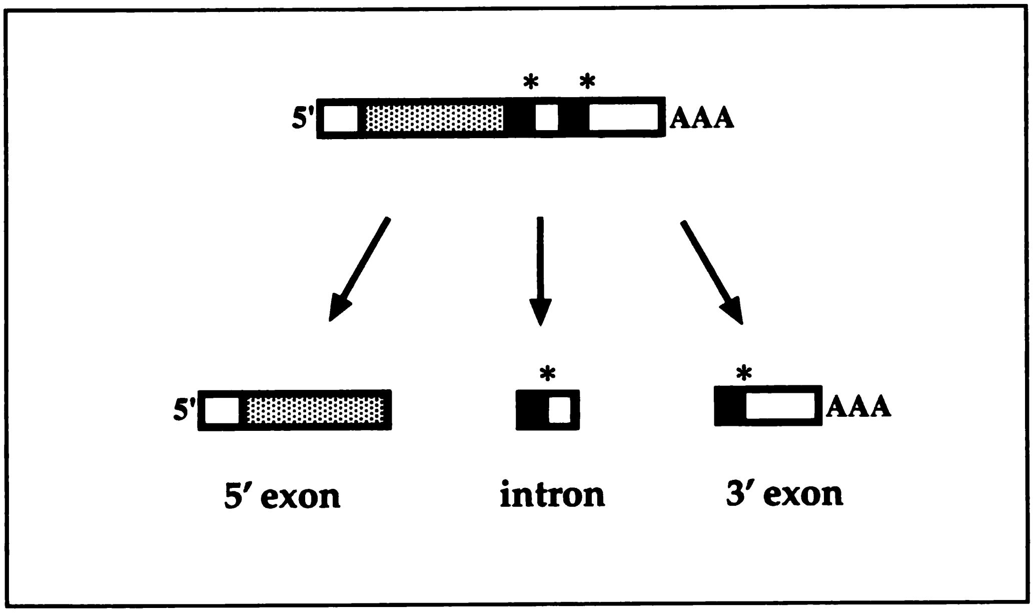

Cleave only
5′ splice site
3′
Hac1p$^{i}$ stop codon
5′ ss
3′ ss
Hac1p$^{u}$ stop codon
5′
29 nt
UUGGCCGAGGAGGGGGUAG
AACCUCCUCCUCCCCCACC
19 nt
AUG
Cleave only
3′ splice site
*rlg1-100* ligase
and C2-Sor+
*rlg1-100* ligase
and C2-Sor+
5′ exon *HAC1* mRNA
(5′ exon+ intron) *HAC1* mRNA
continued translational
inhibition
Hac1p
Δtail
no Hac1p made

***Figure IV-11 Hac1p proteins capable of being produced from HAC1 mRNA***

The top box shows the three fragments that are produced when Ire1p cleaves *HAC1* mRNA. In order from left to right are shown 5'exon, intron, and 3' exon. The bottom box shows the various forms of Hac1p that can be made by translation of the unspliced *HAC1* mRNA to produce Hac1p$^{u}$ (u for uninduced or unspliced), the 5'exon to produce Hac1pΔtail, or the spliced *HAC1* mRNA to produce Hac1p$^{i}$ (i doe induced). Note that final 10 C-terminal amino acids of Hac1p$^{u}$ (indicated in red) are replaced by a different set of 18 amino acids in Hac1p$^{i}$ (indicated in green). Hac1pΔtail lacks either C-terminal tail. Stop codons are highlighted by *.


*
*
5'
AAA
*
5'
*
AAA
5′ exon
intron
3′ exon

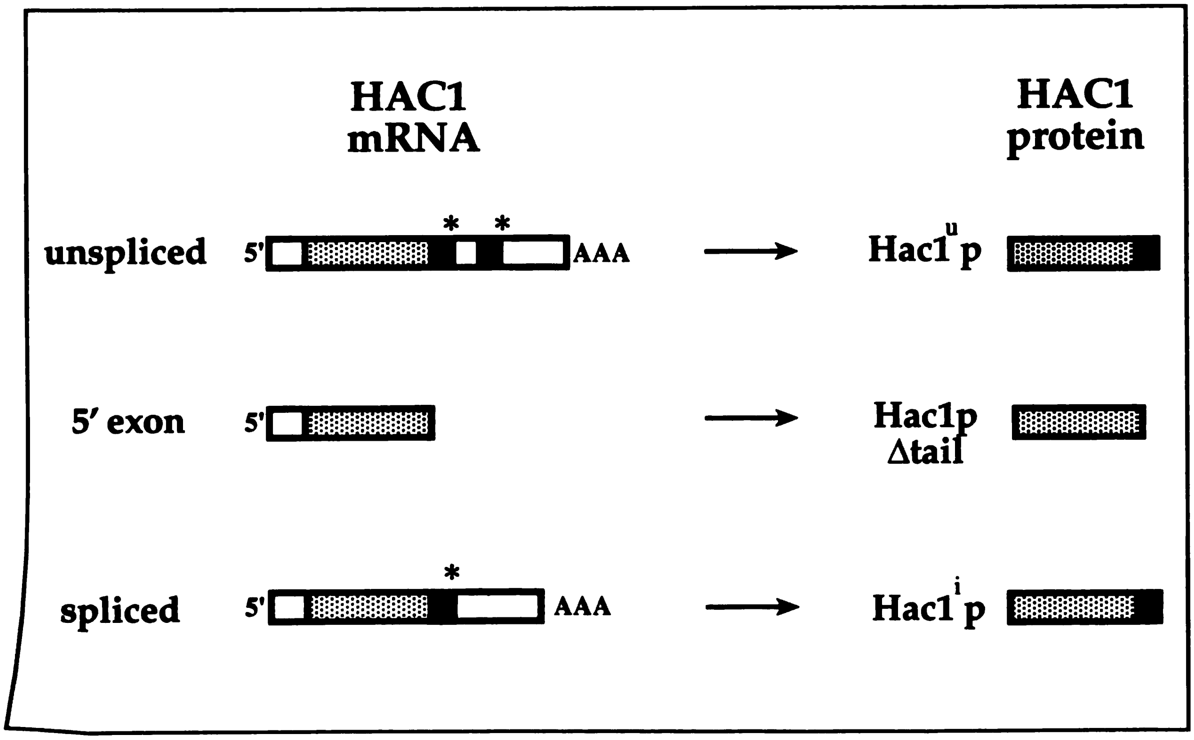

HAC1
mRNA
HAC1
protein
unspliced
*
*
5'
AAA
Hac1$^{u}$p
5′ exon
5'
Hac1p
Δtail
spliced
*
5'
AAA
Hac1$^{i}$p

***Figure IV-12 Northern blot analysis of HAC1 mRNA splicing in C2-Sor⁺ strain***

Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the materials and methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. The genotype of each strain is indicated above the blot. Unspliced, spliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are identified along the right hand side of the blot. The strains (TGy) used in this experiment are indicated at the bottom. Western blot analysis on these same samples is shown in Figure IV-12.

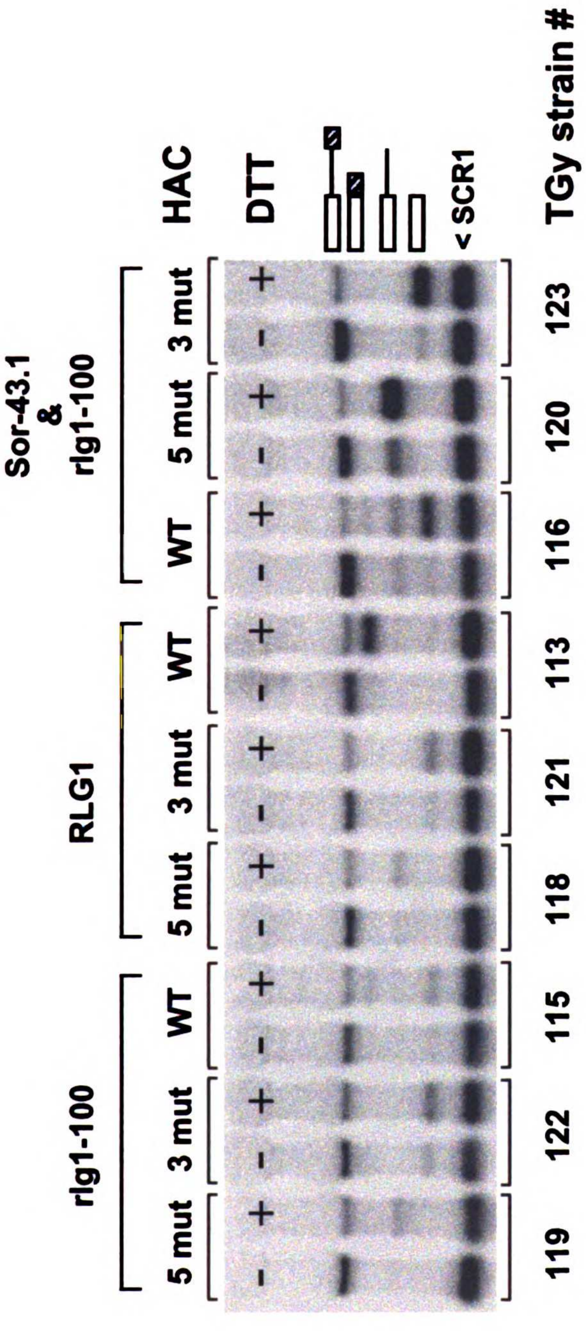
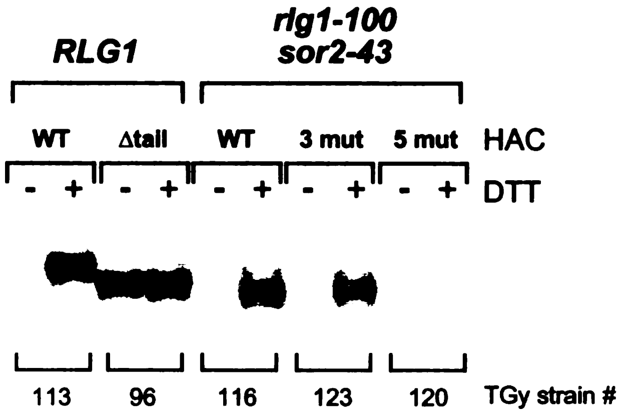

rlg1-100
RLG1
Sor-43.1
&
rlg1-100
HAC
5 mut
3 mut
WT
5 mut
3 mut
WT
WT
5 mut
3 mut
DTT
-
+
< SCR1
TGy strain #
119
122
115
118
121
113
116
120
123

***Figure IV-13 Western blot analysis of Hac1p production in a C2-Sor$^+$ strain***

Total protein was extracted and analyzed by western blot to visualize HA-tagged Hac1p as described in the Materials and Methods section. The UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Cultures were grown at 30°C in YPD to mid-log phase. The strains used are indicated (TGy strain #) along the bottom. The particular HA-tagged Hac1p version carried by each strain as well as its genotype is indicated above each blot. Northern blot analysis on these same samples is shown in Figure IV-12.


RLG1
rlg1-100
sor2-43
WT
Δtail
WT
3 mut
5 mut
HAC
- + - + - + - + - +
DTT
113
96
116
123
120
TGy strain #

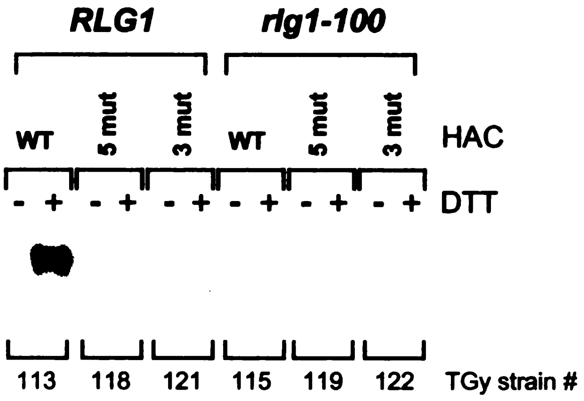

RLG1
rlg1-100
WT
5 mut
3 mut
WT
5 mut
3 mut
HAC
- + - + - + - + - + - +
DTT
113
118
121
115
119
122
TGy strain #

***Figure IV-14 Growth of a C2-Sor$^{+}$ strain carrying wild type, or splice site mutant versions of HAC1***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. SDC-ino+Tm is synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 µg/ml. The Sor strain number of each strain is indicated on the right. The genotype of each strain is indicated on the left as well as the particular *HAC1* version carried by each strain.

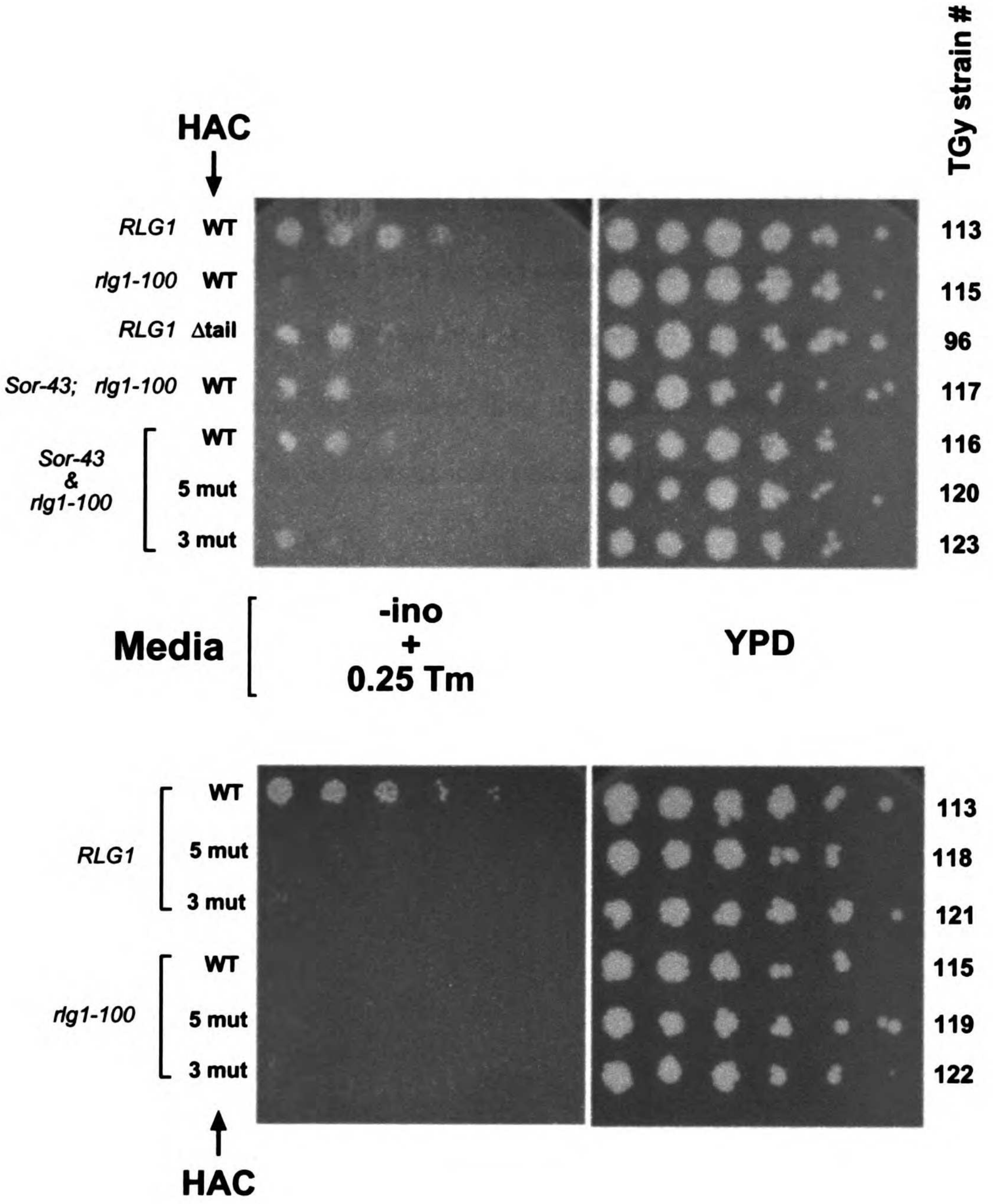

HAC
TGy strain #
RLG1 WT 113
rlg1-100 WT 115
RLG1 Δtail 96
Sor-43; rlg1-100 WT 117
Sor-43 & rlg1-100
WT 116
5 mut 120
3 mut 123
Media
-ino + 0.25 Tm
YPD
RLG1
WT 113
5 mut 118
3 mut 121
rlg1-100
WT 115
5 mut 119
3 mut 122
HAC

***Figure IV-15 Northern blot analysis of HAC1 mRNA splicing in RLG1, C2-Sor$^+$ strains***

Strains were grown at 30°C to mid-log phase and the UPR was induced by addition of DTT to a final concentration of 8 mM for 30 minutes. Total RNA was extracted and analyzed as described in the Materials and Methods. The blot was probed for *SCR1* and *HAC1* RNAs. *SCR1* RNA was used as a loading control. The particular spore from which each strain germinated is indicated above the blot. Unspliced, spliced, 5'exon+intron, and 5'exon *HAC1* mRNA species are identified along the right hand side of the blot. The strains (TGy) used in this experiment are indicated at the bottom.

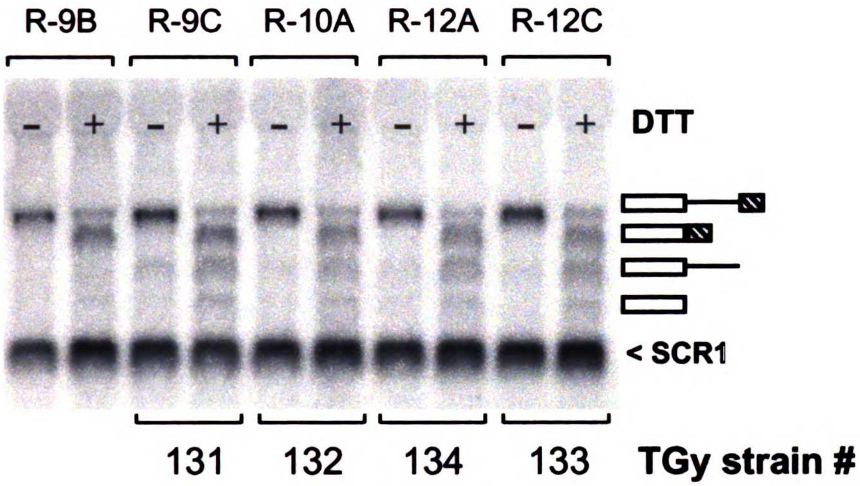

R-9B
R-9C
R-10A
R-12A
R-12C
– + – + – + – + – +
DTT
< SCR1
131
132
134
133
TGy strain #

***Figure IV-16 C2-Sor*$^+$ *strains are defective for nonstop mediated mRNA decay***

*SKI2, Δski2, rlg1-100,* and two *C2-Sor*$^+$ strains, Sor-43.1 and Sor-65.1 were transformed with a *URA3* CEN/ARS plasmid carrying a nonstop allele of the *HIS3* gene (ns-HIS) or a vector control. Transformed and untransformed strains were struck out onto YPD media, SD-uracil, or SD-histidine-uracil. The genotypes of each strain are indicated along the left. Only strains defective for nonstop mediated mRNA decay (such as *Δski2)* should grow in the absence of histidine in the media.

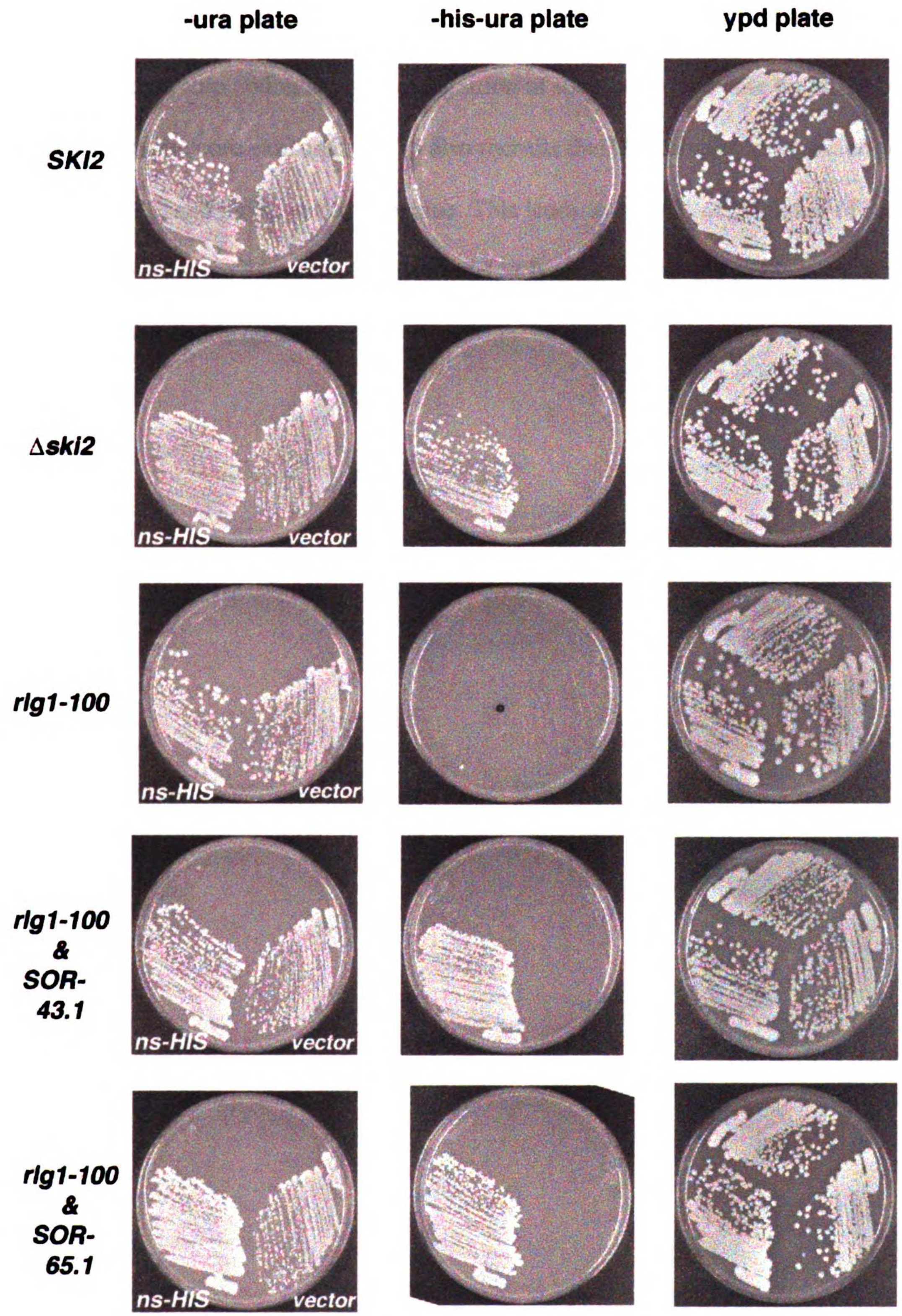

-ura plate
-his-ura plate
ypd plate
SKI2
Δski2
rlg1-100
rlg1-100 & SOR-43.1
rlg1-100 & SOR-65.1
ns-HIS
vector

***Figure IV-17 Model for nonstop mediated mRNA decay in yeast***

Model proposed by van Hoof et al. (2002). Stalling of the ribosome at the 3'end of an mRNA lacking a stop codon leads to recognition of the ribosome A-site by Ski7p bound to the cytoplasmic core exosome. Ski7p also recruits the *SKI* helicase complex composed of Ski2p, Ski3p, and Ski8p to the ribosome. This leads to 3' to 5' degradation of the nonstop mRNA by the exosome. In this model, it is unclear how the ribosome releases the stalled polypeptide chain.

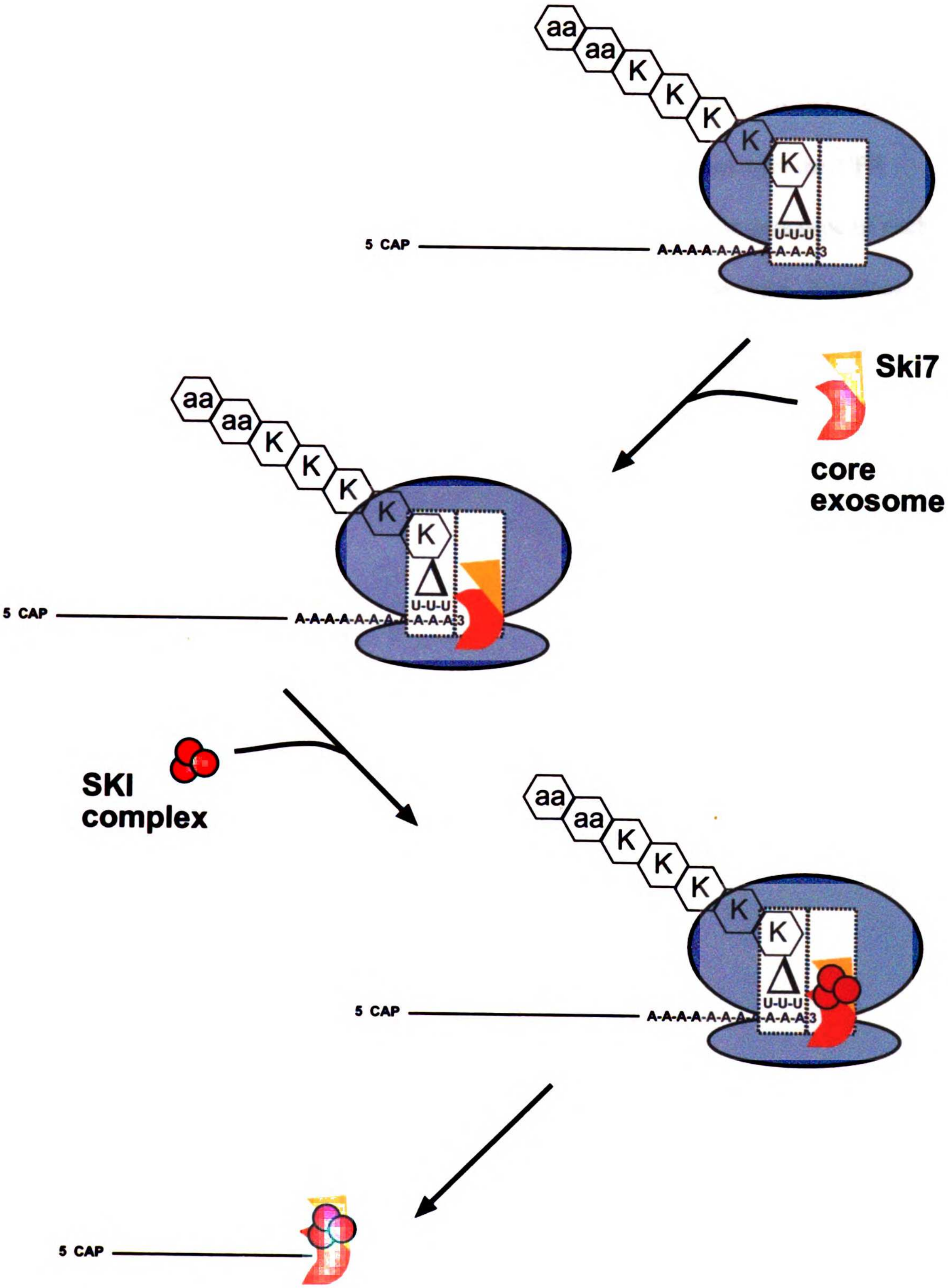

aa
aa
K
K
K
K
K
U-U-U
5 CAP
A-A-A-A-A-A-A-A-A-A3
Ski7
core exosome
SKI complex

***Figure IV-18  Growth of diploid strains deleted for SKI2, SKI3, or SKI8***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. –trp, SD-tryptophan media; -ino+Tm, synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 µg/ml. The genotype of each strain is indicated on the left. The strain numbers are indicated on the right.  * indicates lack of strain designation. All three *SKI* genes were deleted by integration of a DNA cassette carrying the *TRP1* gene.

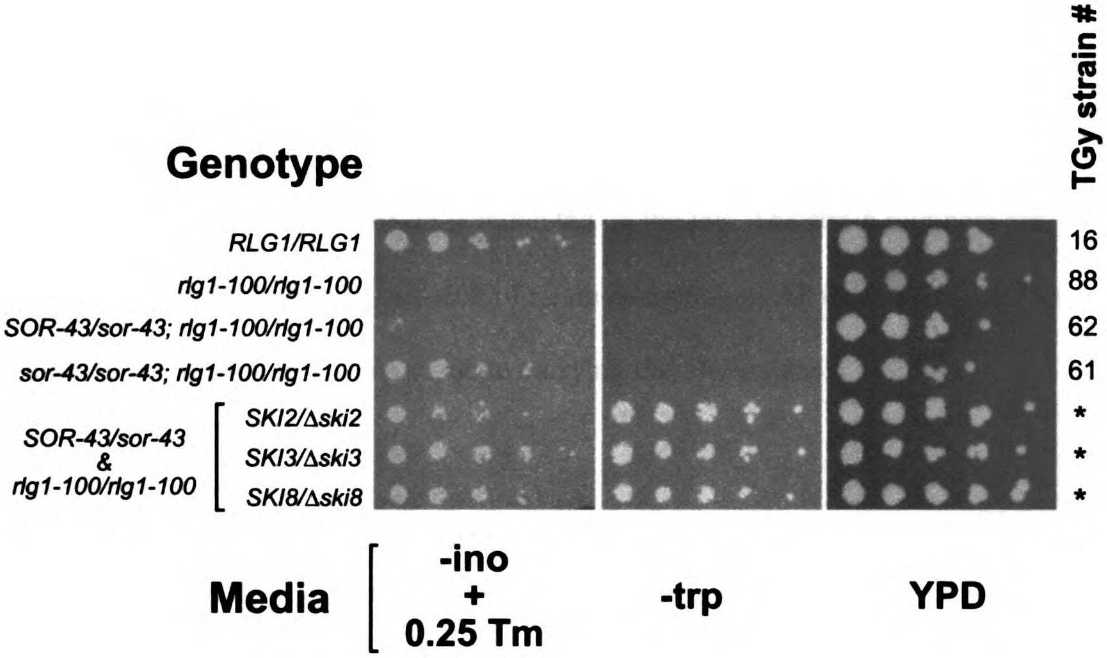
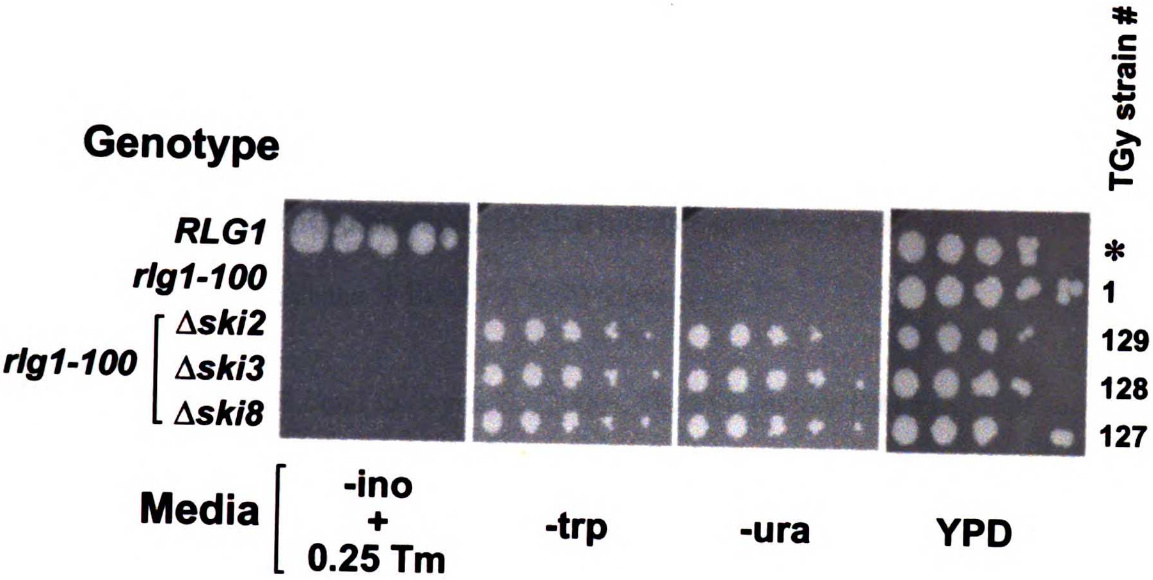

Genotype
TGy strain #
RLG1/RLG1
16
rlg1-100/rlg1-100
88
SOR-43/sor-43; rlg1-100/rlg1-100
62
sor-43/sor-43; rlg1-100/rlg1-100
61
SOR-43/sor-43
&
rlg1-100/rlg1-100
SKI2/Δski2
*
SKI3/Δski3
*
SKI8/Δski8
*
Media
-ino
+
0.25 Tm
-trp
YPD

***Figure IV-19 Growth of haploid rlg1-100 strains deleted for SKI2, SKI3, or SKI8***

Serial dilutions (from left to right, least to most dilute) of each strain were plated onto the indicated media and grown at 30°C for 1 to 2 days. –trp, SD-tryptophan media; -ura, SD-uracil media;-ino+Tm, synthetic dextrose complete media lacking inositol and supplemented with the UPR inducing drug tunicamycin to a final concentration of 0.25 µg/ml. The genotype of each strain is indicated on the left. The strain numbers are indicated on the right. * indicates lack of strain designation. All three *SKI* genes were deleted by integration of a DNA cassette carrying the *TRP1* gene.


Genotype
TGy strain #
RLG1
*
rlg1-100
1
rlg1-100
Δski2
129
Δski3
128
Δski8
127
Media
-ino
+
0.25 Tm
-trp
-ura
YPD

# ***REFERENCES***

Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S., and Katada, T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. EMBO J. *20*, 4684-4693.

Benard, L., Carroll, K., Valle, R. C., Masison, D. C., and Wickner, R. B. (1999). The ski7 antiviral protein is an EF1-alpha homolog that blocks expression of non-Poly(A) mRNA in Saccharomyces cerevisiae. J Virol *73*, 2893-2900.

Bonnerot, C., Boeck, R., and Lapeyre, B. (2000). The two protein Pat1p (Mrt1p) and Spb8p interact *in vivo*, are required for mRNA decay, and are functionally linked to Pab1p. Mol Cell Biol *20*, 5939-5946.

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. EMBO *19*, 1661-1671.

Butler, J. S. (2002). The yin and yang of the exosome. Trends Cell Biol *12*, 90-96.

Cao, D., and Parker, R. (2001). Computational modeling of eukaryotic mRNA turnover. Rna *7*, 1192-1212.

Clark, M. W., and Abelson, J. (1987). The subnuclear localization of tRNA ligase in yeast. J Cell Biol *105*, 1515-1526.

Cox, J. S., and Walter, P. (1996). A novel mechanism for regulating activity of a trancription factor that controls the unfolded protein response. Cell *87*, 391-404.

Diehn, M., Eisen, M. B., Botstein, D., and Brown, P. O. (2000). Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nat Genet *25*, 58-62.

Fischer, N., and Weis, K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. Embo J *21*, 2788-2797.

Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258-2261.

Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. Embo J *18*, 3119-3132.

Guthrie, C., and Fink, G. R. (1991). Guide to yeast genetics and molecular biology (San Diego, Academic Press).

Hsu, C. L., and Stevens, A. (1993). Yeast cells lacking 5' to 3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol. Cell. Biol. *13*:4826-4835.

Kaufman, R. J. (2002). Orchestrating the unfolded protein response in health and disease. J Clin Invest *110*, 1389-1398.

Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical

PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Maquat, L. E. (2002). Molecular biology. Skiing toward nonstop mRNA decay. Science *295*, 2221-2222.

Mori, K., Ogawa, N., Kawahara, T., Yanagi, H., and Yura, T. (2000). mRNA splicing-mediated C-terminal replacement of transcription factor Hac1p is required for efficient activation of the unfolded protein response. Proc Natl Acad Sci U S A *97*, 4660-4665.

Muhlrad, D., Decker, C. J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. Mol Cell Biol *15*, 2145-2156.

Ng, D. T., Spear, E. D., and Walter, P. (2000). The unfolded protein response regulates multiple aspects of secretory and membrane protein biogenesis and endoplasmic reticulum quality control. J Cell Biol *150*, 77-88.

Patil, C., and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Curr Opin Cell Biol *13*, 349-355.

Phizicky, E. M., and Fields, S. (1995). Protein-protein interactions: Methods for detection and analysis. Microbiol Rev *59*, 94-123.

Puglisi, J. D., Blanchard, S. C. and Green, R. (2000) Approaching translation at atomic resolution. Nat Struct Biol *7*, 855-861.

Ramirez, C. V., Vilela, C., Berthelot, K., and McCarthy, J. E. G. (2002). Modulation of Eukaryotic mRNA Stability via the Cap-binding Translation Complex eIF4F. J Mol Biol *318*: 951-962.

Ruegsegger, U., Leber, J. H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103-114.

Searfoss, A. M., and Wickner, R. B. (2000). 3' poly(A) is dispensable for translation. Proc Natl Acad Sci U S A *97*, 9133-9137.

Schwartz, D. and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of yeast mRNAs. Mol Cell Biol *19*, 5247-5256.

Schwartz, D. and Parker, R. (2000). mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E. Mol Cell Biol *20*, 7933-7942.

Schwartz, D., Decker, C. J., and Parker, R. (2003). The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. RNA *9*, 239-251.

Sidrauski, C., Chapman, R., and Walter, P. (1998). The unfolded protein response: an intracellular signalling pathway with many surprising features. Trends Cell Biol *8*, 245-249.

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response [see comments]. Cell *87*, 405-413.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell *90*, 1-20.

Sikorski, R. S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19-27.

Steiger, M., Carr-Schmid, A., Schwartz, D. C., Kiledjian, M., and Parker, R. (2003). Analysis of recombinant yeast decapping enzyme. RNA *9*, 231-238.

Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000) Yeast Sm-like proteins function in mRNA decapping and decay. Nature *404*, 515-518.

Toh, E. A., Guerry, P., and Wickner, R. B. (1978). Chromosomal superkiller mutants of Saccharomyces cerevisiae. J Bacteriol *136*, 1002-1007.

Toh, E. A., and Wickner, R. B. (1980). "Superkiller" mutations suppress chromosomal mutations affecting double-stranded RNA killer plasmid replication in saccharomyces cerevisiae. Proc Natl Acad Sci U S A *77*, 527-530.

Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249-258.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.

Vasudevan, S., Peltz, S. W., and Wilusz, C. J. (2002). Non-stop decay--a new mRNA surveillance pathway. Bioessays *24*, 785-788.

Wickner, R. B. (1993). Host control of yeast dsRNA virus propagation and expression. Trends Microbiol *1*, 294-299.

Widner, W. R., and Wickner, R. B. (1993). Evidence that the SKI antiviral system of Saccharomyces cerevisiae acts by blocking expression of viral mRNA. Mol Cell Biol *13*, 4331-4341.

Wilusz, C. J., Wormington, M., and Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol *2*, 237-246.

# Appendix A

## *The sequence of tRNA ligase:*
## *speculations on structure and function*

### *The primary sequence of tRNA ligase: speculations on structure and function*

The tRNA ligase of *Saccharomyces cerevisiae* is a 95KD multifunctional protein with adenylylate synthetase, polynucleotide kinase, and 2',3'-cyclic phosphodiesterase activities. By protease digestion and deletion analysis, these enzymatic activities have been mapped to three distinct regions of the protein (Figure A-1) (Apostol et al., 1991; Xu et al., 1990). All three domains participate enzymatically during *in vitro* splicing of pre-tRNA and *HAC1* mRNA substrates (Abelson et al., 1998; Gonzalez et al., 1999; Greer et al., 1983; Phizicky et al., 1986). Following endonucleolytic cleavage of the RNA substrate by tRNA endonuclease or Ire1p, the cyclic phosphodiesterase domain of tRNA ligase opens up the 2',3'-cyclic phosphate at the end of the 5' exon to yield a 2'-phosphate. The kinase domain phosphorylates the 5'-hydroxyl at the end of the 3' exon using the gamma phosphate from GTP or ATP. The adenylylate synthetase domain reacts with ATP to form a covalent ligase-AMP intermediate. The AMP is then transferred to the 5'-phosphate on the 3' exon to form a high energy 5'-5' phosphoanhydride bond. Ligation occurs with the concomitant release of the AMP activating group from the RNA substrate, leaving behind a 2' phosphate at the splice junction.

Though yeast-like tRNA ligase activities have been described in plants (Gegenheimer et al., 1983; Schwartz et al., 1983) and mammals (Zillmann et al., 1991), to date, the amino acid sequences of only eight tRNA ligase genes have been identified; they are all encoded by yeasts (Figure A-2). Notwithstanding the lack of tRNA ligase homologues, regions within the three tRNA ligase domains resemble other functionally related proteins. The adenylylate synthetase domain of tRNA ligase is a good deal similar to the corresponding N-terminal domain of T4 ligase (Apostol et al., 1991; Koonin and

Gorbalenya, 1990). During the ligation reactions catalyzed by these two enzymes, a conserved residue, lysine-99 of T4 ligase and the equivalent residue, lysine-114 of tRNA ligase, is transiently adenylylated (Thogersen et al., 1985; Xu et al., 1990). More recently, the adenylylate synthetase domain of tRNA ligase was identified as a member of the covalent nucleotidyl transferase (CNT) superfamily (Figure A-3) (Aravind and Koonin, 1999; Shuman, 1996). This protein superfamily includes ATP- and NAD-dependent DNA ligases, mRNA capping enzymes, and RNA ligases characterized by the presence of five to six conserved motifs. Crystal structures of five members reveal a common fold that brings motifs I, III, IIIa, IV, and V together at the active site (Hakansson et al., 1997; Lee et al., 2000; Odell et al., 2000; Singleton et al., 1999; Subramanya et al., 1996). Thus, despite amino acid sequence divergence outside of the conserved motifs, these covalent nucleotidyl transferase proteins share a common fold that positions the conserved motifs at the ligase active site (Aravind and Koonin, 1999; Shuman, 1996). The structure of these proteins has been described as "pincher-like;" it is composed of two subdomains that form an electro-positive cleft between them, in which the ligation substrate is thought to bind. ATP binds in a pocket within this cleft as well (Shuman, 1996).

Of the three tRNA ligase domains, the adenylylate synthetase domain is the most highly conserved across the eight species of yeast (Figure A-2), with 25% (99/396) identical amino acids. These conserved regions are mostly associated with the five conserved CNT motifs. The H148Y mutation in the UPR-defective *rlg1-100* allele of ligase maps to adenylylate synthetase domain, just upstream of motif III (Figure A-3) (Sidrauski et al., 1996). In addition, two temperature sensitive-mutations causing defects in pre-tRNA splicing map between motifs III and IIIa: F170S (*rlg1-10*) and T180I (*rlg1-*

*4*) (Phizicky et al., 1992). Residues in motif III contact the ribose ring of ATP whereas those in motif IIIa contact the purine ring (Shuman, 1996). The possible significance of this clustering of mutations remains to be seen; however, it would be interesting to learn if *rlg1-100* strains are temperature-sensitive for growth in the absence of UPR-induction and if *rlg1-4* and *rlg1-10* strains are UPR-defective.

Whereas the similarities between T4 ligase and tRNA ligase extend over a region of 100 amino acids, only three small pockets of homology to other kinases are discernable in the kinase domain of tRNA ligase. The first of these, Nuc A, corresponds to a Walker A nucleotide binding motif akin to that found in T4 polynucleotide kinase (T4 PNK) and adenylate kinase; the second, Nuc B, corresponds to a divergent Walker B nucleotide binding motif, similar to that found in T4 PNK (Apostol et al., 1991; Koonin and Gorbalenya, 1990; Walker et al., 1982). Amino acids from both Walker motifs contribute to the T4 PNK kinase active site (Galburt et al., 2002; Wang et al., 2002). The third region of homology, PNK 1, is similar to a portion of T4 PNK that also forms part of the kinase active site (Apostol et al., 1991; Galburt et al., 2002; Koonin and Gorbalenya, 1990; Wang et al., 2002). Of these three regions, the Nuc A and Nuc B are the most highly conserved in sequence across the eight species of yeast (Figure A-2). The structures of the kinase domain of T4 PNK and adenylate kinase superimpose readily; thus it is tempting to speculate that the structure of the tRNA ligase kinase domain might easily superimpose on the structures of these two kinases as well.

The 2',3'-cyclic phosphodiesterase domain of tRNA ligase was recently identified as a member of the 2H phosphodiesterase protein superfamily which includes members found in archaebacteria, eubacteria, and eukaryotes (Mazumder et al., 2002). These proteins are predicted to adopt a common fold typified by the structure of the *Arabidopsis* ADP-ribose 1",2" cyclic phosphodiesterase (Figure A-5) (Hofmann et al., 2002; Hofmann et al., 2000). This protein family is named for two conserved active site histidine residues each found in the motif Hh(S/T)h, where h is a hydrophobic residue. The histidine and serine/threonine residues of this motif are essential for phosphodiesterase function (Hofmann et al., 2000; Nasr and Filipowicz, 2000). All eight yeast tRNA ligase genes have two of these Hh(S/T)h motifs as predicted for members of this phosphoesterase family (Figure A-5).

We also performed protein database searches using various portions of the tRNA ligase amino acid sequence. The only significant match found was with the putative DEAD-box related helicase, Ski2p. Ski2p shares sequence similarity with portions of the kinase and cyclic phosphodiesterase domains of tRNA ligase (Figure A-6). DEAD-box RNA helicases are defined by the presence of seven to eight conserved motifs (de la Cruz et al., 1999; Linder et al., 1989; Linder et al., 2001). Motifs I and II are respectively Walker A and B nucleotide binding motifs. The first four and the final three to four helicase motifs group together to fold into two discrete domains; thus the fully folded helicase is composed of two domains linked together to form a cleft for NTP binding. The region of homology shared between Ski2p and tRNA ligase includes the first domain but not the second domain of these helicases. In addition, the amino acids within the four motifs that make up the only shared helicase domain are not well conserved in tRNA ligase (Figure A-6). Thus it seems unlikely that tRNA ligase is active as a helicase.

Ski2p is involved in non-stop mediated RNA degradation, a process by which mRNAs lacking stop codons are degraded (Frischmeyer et al., 2002; van Hoof et al.,

2002; Vasudevan et al., 2002). More specifically, mRNAs in the midst of being translated are preferentially destroyed by this pathway. *HAC1* mRNA is polysome associated when it is spliced by Ire1p and tRNA ligase in the cytoplasm (Ruegsegger et al., 2001). Perhaps the region of shared homology mediates contact with the ribosome (or ribosome associated factors) and Ski2p during non-stop mediated mRNA degradation or tRNA ligase during *HAC1* mRNA splicing.

### *T4 PNK: a model for tRNA ligase:tRNA interactions?*

T4 PNK is composed of two domains of comparable size linked by a flexible hinge (Galburt et al., 2002). The kinase domain is located in the N-terminal half of the protein and the 3'-phosphatase domain in the C-terminal half. T4 PNK is one of the most widely used enzymes in molecular biology laboratories. However, its role in the T4 bacteriophage lifecycle is generally less widely known. When a bacterium becomes infected with T4 phage, the bacterium responds by cleaving its own tRNAs presumably to prevent the production of T4 phage proteins. In turn, T4 phage has developed a way to get around this suicide response; T4 repairs the cleaved tRNAs using T4 PNK and T4 ligase (Amitsur et al., 1987). At the site where the tRNA was cleaved, T4 PNK phosphorylates the 5'-hydroxyl, opens up the 2',3'-cyclic phosphate, and then removes the resulting 3'-phosphate-- all to produce a substrate appropriate for ligation by T4 ligase. As an exciting consequence of the recent crystal structure of T4 PNK, Galburt and colleagues (Galburt et al., 2002) have proposed a model for how the polynucleotide kinase and 3'-phosphatase active sites of T4 can simultaneously interact with the cleaved tRNA termini (Figure A-7). T4 PNK forms a homotetramer, and this subunit arrangement is the key to the binding model proposed. In the model, the kinase domain of one monomer cooperates with the phosphatase domain of a second monomer. The distances and orientation between the two active sites in the two monomers is about right for binding the two tRNA termini. This is not the case for the two active sites within a single monomer.

The biological substrates for T4 PNK and tRNA ligase are virtually the same: a tRNA cleaved at or adjacent to the anticodon loop and bearing 5'-OH and 2',3'-cyclic phosphate termini at the cleavage site. In fact, tRNA ligase will bind a cleaved pre-tRNA about 5 times more strongly than it will bind an intact pre-tRNA (Apostol and Greer, 1991). And, like T4 PNK, the kinase domain of tRNA ligase is just N-terminal to the phosphodiesterase domain. Thus, it is tempting to propose that tRNA ligase might interact with tRNA much like T4 PNK does. With this appealing model in mind, we must nevertheless recall the significant differences between the two proteins. As outlined above, the primary sequence of the kinase domains of tRNA ligase and T4 PNK are similar in three small regions associated with the kinase active site. However, the phosphodiesterase domains of these two proteins appear to be members of two distinct families of phosphodiesterases characterized by differing protein folds and active site residues. The 2',3'-cyclic phosphodiesterase domain of tRNA ligase is a member of the 2H phoshophodiesterase family defined by the presence of two essential histidine active site residues (Mazumder et al., 2002). The 3'-phosphatase domain of T4 PNK is a member of the haloacid dehalogenase (HAD) family of phosphotransferases defined by the presence of the active site motif DxDxT (Galburt et al., 2002; Wang and Shuman, 2002). In addition, a separate protein, T4 ligase, catalyzes the final ligation step in the T4 system, whereas ligation of cleaved pre-tRNAs is catalyzed completely by tRNA ligase. In the pre-tRNA splicing reaction, it is almost as if the T4 ligase were fused just N-terminal to the kinase domain to produce tRNA ligase. Thus if we were to model the binding interactions between tRNA ligase and a cleaved pre-tRNA, we must consider how the adenylylate synthetase domain fits in. In addition, we need to determine if ligase behaves as a monomer or oligomer in solution. As seen for T4 PNK, the active sites of adjacent monomers may work together on a single substrate as opposed to the active sites within the single subunit

### ***Future directions***

Yeast tRNA ligase was first described in the literature twenty years ago (Greer et al., 1983). Despite these past two decades, there is still much we do not know or understand about this fascinating, multifunctional protein. The recent bioinformatic explosion has provided us with more clues as to how ligase performs its functions by making it more and more facile to uncover functionally significant yet subtle primary sequence homologies with other more thoroughly characterized enzymes. What can we do with this information? We should at the very least determine if the amino acids predicted to be in the active site of each of the three domains of tRNA ligase are indeed required for the catalytic function of each domain. This can easily be tackled by introducing mutations at these sites and testing the function of resulting protein *in vitro*. In addition, given that there are crystal structures available for members of the covalent nucleotidyl transferase family, the polynucleotide kinase family, and the 2H phosphoesterase family, we should pursue solving the structure for tRNA ligase. These structures can be used to aid in solving the structures of individual tRNA ligase domains by molecular replacement. Having the structure of the entire protein should provide clues as to how the active sites of all three domains access the ends of the RNA substrate to be ligated. In addition, structures of tRNA ligase bound to tRNA, ATP, GTP, or covalently attached to AMP have the potential to greatly increase our understanding of this protein and lead to testable hypothesis. Do the domains change orientation or shape upon binding to any of these ligands? What is the role, if any, of the amino acids that link the three domains? Is substrate channeled from one domain's active site to another? In lieu of actual crystal structures, homology models of each of the domains could be constructed. Again, inspection of these models should likely lead to testable hypothesis regarding the mechanism by which tRNA ligase coordinates the activities of its three domains to bring about RNA ligation.

***Figure A-1   The domains of tRNA ligase***

**(A)** Approximate domain boundaries as defined by proteolytic and deletion analysis (Apostol et al., 1991; Xu et al., 1990).

**(B)** Regions of homology to other proteins within each domain. The adenylylate synthetase domain is a member of the covalent nucleotidyl transferase family (Aravind and Koonin, 1999). Nuc A and Nuc B refer to nucleotide binding motifs found in T4 polynucleotide kinase (PNK) and adenylate kinase (Apostol et al., 1991; Koonin and Gorbalenya, 1990). PNK 1 refers to a third region of homology with T4 PNK (see text). The cyclic phosphodiesterase domain is a member of the 2H phosphoesterase protein family (Mazumder et al., 2002) and also is similar to a portion of the Ski2 protein.

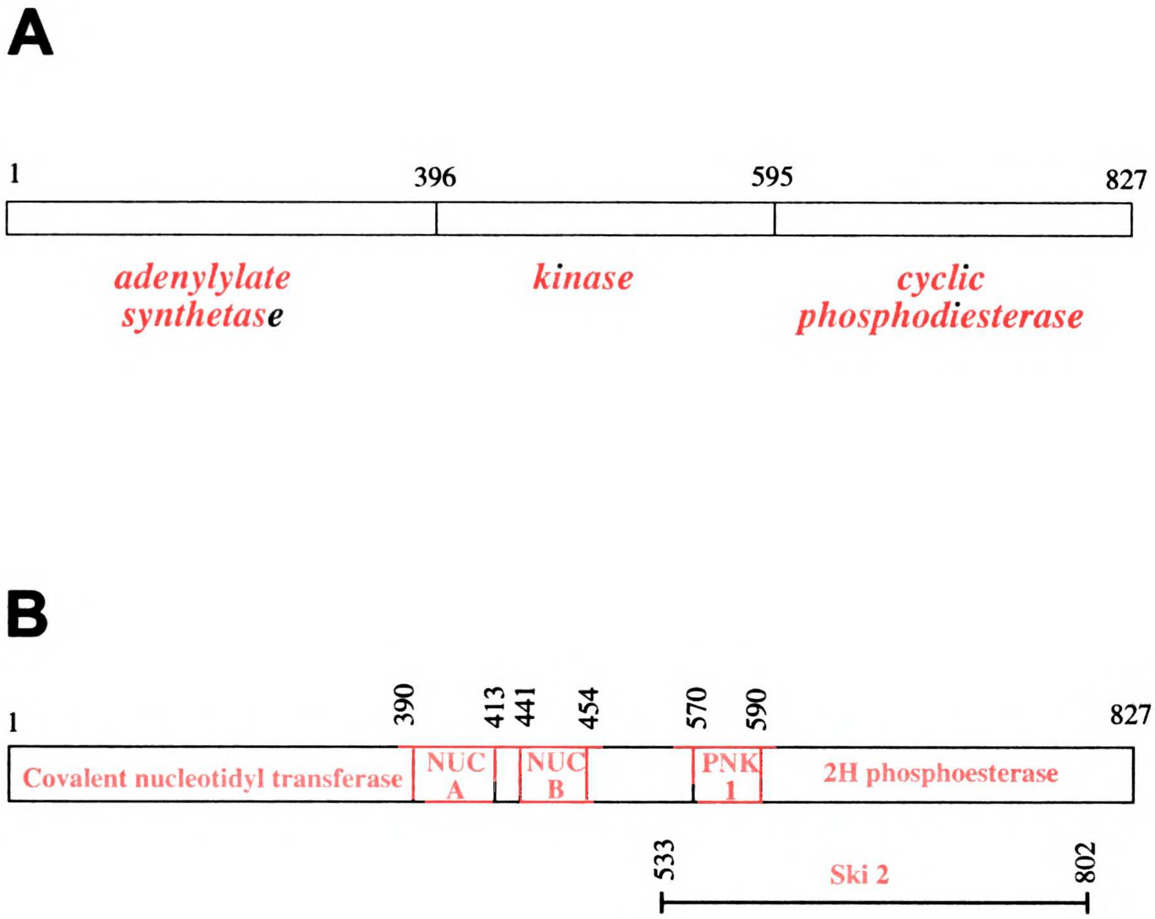
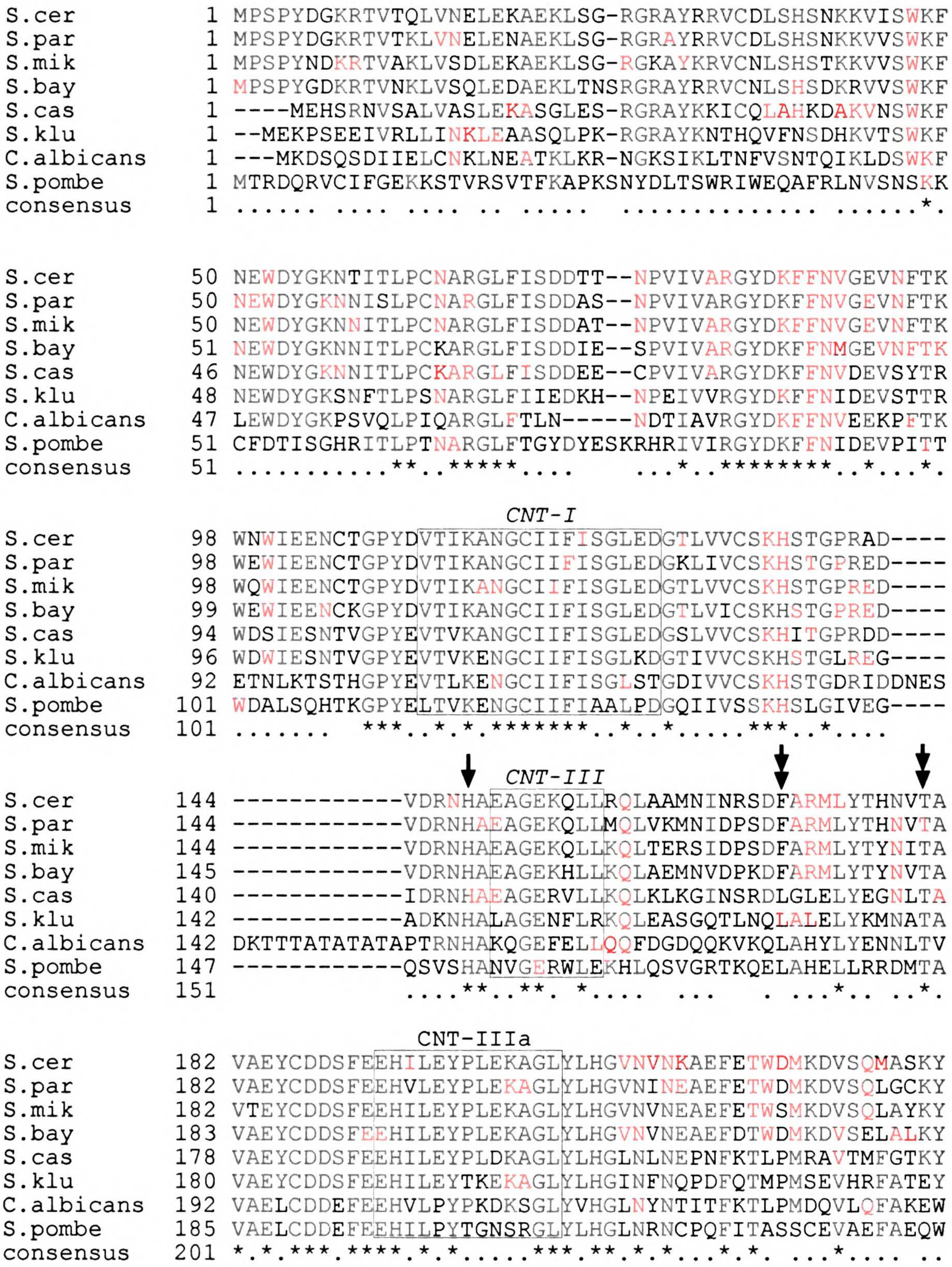

A
1
396
595
827
adenylylate synthetase
kinase
cyclic phosphodiesterase
B
1
390
413
441
454
570
590
827
Covalent nucleotidyl transferase
NUC A
NUC B
PNK 1
2H phosphoesterase
533
Ski 2
802

***Figure A-2  tRNA ligase genes of yeast***

ClustalW (Baylor College of Medicine, http://searchlauncher.bcm.tmc.edu/multi-align/Options/clustalw.html) and Boxshade (European Molecular Biology Network, www.che.emnet.org/software/Box_form.html) computer programs were used to align the tRNA ligase genes of eight species of yeast. Protein sequence data was down loaded from the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces/), the Washington University Genome Sequencing Center (http://www.genome.wustl.edu/projects/yeast/), and the CandidaDB Web Server (http://genolist.pasteur.fr/CandidaDB/). Abbreviations are: S.cer, *Saccharomyces cerevisiae;* S.par, *Saccharomyces paradoxus*; S.mik, *Saccharomyces mikatae*; S.bay, *Saccharomyces bayanus*; S.cas, *Saccharomyces castelli*; S.klu, *Saccharomyces kluyveri*; C.albicans, *Candida albicans*; S.pombe, *Schizosaccharomyces pombe*. Consensus sequence symbols: (.) indicates a residue that is conserved or is replaced by a related residue in at least four of the species of yeast. (*) indicates a residue that is conserved in all eight species. CNT-I through CNT-V are conserved motifs found in the covalent nucleotidyl transferase (CNT) protein family (see Fig. A2-3). Nuc A, Nuc B, and PNK I are regions of homology to T4 PNK (see Fig. A2-4). The 2H-1 and 2H-2 refer to two motifs found in the active site of 2H phosphoesterase family proteins (see Fig. A2-5). Just upstream of the 2H-2 motif, we have highlighted in bold what we believe to be the 2H-2 motif for C. albicans that was missed by the ClustalW program. The single headed arrow marks the location of the H148Y mutation found in the UPR-defective *rlg1-100* yeast. The double headed arrows mark the locations of two separate mutations that each cause temperature-sensitive defects in pre-tRNA splicing: F107S (*rlg1-10*) and T180I (*rlg1-4*).



```
S.cer         1 MPSPYDGKRTVTQLVNELEKAEKLSG-RGRAYRRVCDLSHSNKKVISWKF
S.par         1 MPSPYDGKRTVTKLVNELENAEKLSG-RGRAYRRVCDLSHSNKKVVSWKF
S.mik         1 MPSPYNDKRTVAKLVSDLEKAEKLSG-RGKAYKRVCNLSHSTKKVVSWKF
S.bay         1 MPSPYGDKRTVNKLVSQLEDAEKLTNSRGRAYRRVCNLSHSDKRVVSWKF
S.cas         1 ----MEHSRNVSALVASLEKASGLES-RGRAYKKICQLAHKDAKVNSWKF
S.klu         1 --MEKPSEEIVRLLINKLEAASQLPK-RGRAYKNTHQVFNSDHKVTSWKF
C.albicans    1 ---MKDSQSDIIELCNKLNEATKLKR-NGKSIKLTNFVSNTQIKLDSWKF
S.pombe       1 MTRDQRVCIFGEKKSTVRSVTFKAPKSNYDLTSWRIWEQAFRLNVSNSKK
consensus     1 ...... .... .... .. .....  .............. ......*.

S.cer        50 NEWDYGKNTITLPCNARGLFISDDTT--NPVIVARGYDKFFNVGEVNFTK
S.par        50 NEWDYGKNNISLPCNARGLFISDDAS--NPVIVARGYDKFFNVGEVNFTK
S.mik        50 NEWDYGKNNITLPCNARGLFISDDAT--NPVIVARGYDKFFNVGEVNFTK
S.bay        51 NEWDYGKNNITLPCKARGLFISDDIE--SPVIVARGYDKFFNMGEVNFTK
S.cas        46 NEWDYGKNNITLPCKARGLFISDDEE--CPVIVARGYDKFFNVDEVSYTR
S.klu        48 NEWDYGKSNFTLPSNARGLFIIEDKH--NPEIVVRGYDKFFNIDEVSTTR
C.albicans   47 LEWDYGKPSVQLPIQARGLFTLN-----NDTIAVRGYDKFFNVEEKPFTK
S.pombe      51 CFDTISGHRITLPTNARGLFTGYDYESKRHRIVIRGYDKFFNIDEVPITT
consensus    51 ...........**..*****....    ...*..********..*...*.

                                   CNT-I
S.cer        98 WNWIEENCTGPYDVTIKANGCIIFISGLEDGTLVVCSKHSTGPRAD----
S.par        98 WEWIEENCTGPYDVTIKANGCIIFISGLEDGKLIVCSKHSTGPRED----
S.mik        98 WQWIEENCTGPYDVTIKANGCIIFISGLEDGTLVVCSKHSTGPRED----
S.bay        99 WEWIEENCKGPYDVTIKANGCIIFISGLEDGTLVICSKHSTGPRED----
S.cas        94 WDSIESNTVGPYEVTVKANGCIIFISGLEDGSLVVCSKHITGPRDD----
S.klu        96 WDWIESNTVGPYEVTVKENGCIIFISGLKDGTIVVCSKHSTGLREG----
C.albicans   92 ETNLKTSTHGPYEVTLKENGCIIFISGLSTGDIVVCSKHSTGDRIDDNES
S.pombe     101 WDALSQHTKGPYELTVKENGCIIFIAALPDGQIIVSSKHSLGIVEG----
consensus   101 .......  ***..*.*.*******..*..*.....***..*....

                                ↓                    ↓         ↓
                                   CNT-III
S.cer       144 ------------VDRNHAEAGEKQLLRQLAAMNINRSDFARMLYTHNVTA
S.par       144 ------------VDRNHAEAGEKQLLMQLVKMNIDPSDFARMLYTHNVTA
S.mik       144 ------------VDRNHAEAGEKQLLKQLTERSIDPSDFARMLYTYNITA
S.bay       145 ------------VDRNHAEAGEKHLLKQLAEMNVDPKDFARMLYTYNVTA
S.cas       140 ------------IDRNHAEAGERVLLKQLKLKGINSRDLGLELYEGNLTA
S.klu       142 ------------ADKNHALAGENFLRKQLEASGQTLNQLALELYKMNATA
C.albicans  142 DKTTTATATATAPTRNHAKQGEFELLQQFDGDQQKVKQLAHYLYENNLTV
S.pombe     147 ------------QSVSHANVGERWLEKHLQSVGRTKQELAHELLRRDMTA
consensus   151             ....**..**. *....  ...   . ...*.. ..*.

                             CNT-IIIa
S.cer       182 VAEYCDDSFEEHILEYPLEKAGLYLHGVNVNKAEFETWDMKDVSQMASKY
S.par       182 VAEYCDDSFEEHVLEYPLEKAGLYLHGVNINEAEFETWDMKDVSQLGCKY
S.mik       182 VTEYCDDSFEEHILEYPLEKAGLYLHGVNVNEAEFETWSMKDVSQLAYKY
S.bay       183 VAEYCDDSFEEHILEYPLEKAGLYLHGVNVNEAEFDTWDMKDVSELALKY
S.cas       178 VAEYCDDSFEEHILEYPLDKAGLYLHGLNLNEPNFKTLPMRAVTMFGTKY
S.klu       180 VAEYCDDSFEEHILEYTKEKAGLYLHGINFNQPDFQTMPMSEVHRFATEY
C.albicans  192 VAELCDDEFEEHVLPYPKDKSGLYVHGLNYNTITFKTLPMDQVLQFAKEW
S.pombe     185 VAELCDDEFEEHILPYTGNSRGLYLHGLNRNCPQFITASSCEVAEFAEQW
consensus   201 *.*.***.****.*.*.....***.**.*.*...*.*. ...*.... ..
```

```
                                                   CNT-IV
S.cer        232 GFRCVQCITSNTLEDLKKFLDNCSATGSFEGQEIEGFVIRCHLKST----
S.par        232 GFRCVQCTSLKTLDDLKKFLDNCSATGSFQGQEIEGFVIRCHLKST----
S.mik        232 GFKHTQSFTLNTLKDLKEFLENCSTSGSFQGQEIEGFVIRCHLENT----
S.bay        233 GFKRIQSFKINTIEDLKKLFDECSASGSFQGQEIEGFVIRCHLKSN----
S.cas        228 GFKNINAFIKDDVGSLKVFLEDCAVNGSFNGEEIEGFVIRCKSRDT----
S.klu        230 NFKKIDFFTKPDVASLREFLEDCATKGSYKGMEIEGFVVRCQNKD-----
C.albicans   242 GFKYVSYLTYDNADELFKFLHKCSETGTYNGREIEGFVIRCHRQSHTNGD
S.pombe      235 GFMKVSSFFMDSIHELKAFLENASKDGKWNNRAIEGFVIRCHSDHSSLEQ
consensus    251 .*.... ... ....*........ .*......*****.**.....

                       CNT-V
S.cer        278 --EKPFFFKYKFEEPYLMYRQWREVTKDYIS-NKSRVFKFRKHKFITNKY
S.par        278 --GKPFFFKYKFEEPYLMYRQWREVTKDYIS-NKSRVFKFRKHKFITNKY
S.mik        278 --GKPFFFKFKFEEPYLMYRQWREVTKDYIS-KKSRLFKFKKHKFITNKY
S.bay        279 --GKPFFFKYKFEEPYLMYRQWREVTKDYIS-NNSRVFKFKKHKFITNKY
S.cas        274 --GDSFFFKYKFEEPYLMYRQWREVTKDYIT-NKSRVFKFRKHKLVTNKY
S.klu        275 --GNAFFFKYKFEEPYLMYRQWREVTKEYIK-NKSRVFRFKKHKFITNKY
C.albicans   292 TDGDCFFFKYKFEQPYLLYRQFREVTKQLLNGTPINSIKIKKNKPITKKY
S.pombe      285 QSSNDFFFKYKFPEPYGMFRQWREVTKMLIS---GKKPSYTKYKKVTAEY
consensus    301   ...****.**..**...**.*****.... .........*.*..*..*

S.cer        325 LDFAIPILESSPKICENYLKGFGVIELRNKFLQSY----GMSGLEILNHE
S.par        325 LDFVIPILESSPKICEDYMKGFGVIKLRNRFLGSY----GMSGLEILNHE
S.mik        325 LDFIIPILESSPKICEDYMKGFGVIELRNKFLQSY----GMSGLEILNHE
S.bay        326 LDFVIPILDSSPKIREDYMKGFGVIGLRTMFLKSY----GMSGLEILNHE
S.cas        321 LDFVIPILDNDSRLCEEYMRGFGIIKLRNMFLQSL----GMSGMEILNHE
S.klu        322 LDFVIPLLNENPQLCDEYMRGFGVIKLRNAFLKSY----GMSGFEILNQE
C.albicans   342 LQFVEKLFEQEPEIARNFENGFDIIKVRQLFLQSLNETNGMNLLSIDSEL
S.pombe      332 ITFCDKKFKEDEDAKRLYMSNKGIISLRDEFLVLS---------------
consensus    351 ..*.....................*..*. **...    ...........

                                             Nuc A
S.cer        371 KVAELELKNAIDYDKVDERTKFLIFPISVIGCGKTTTSQTLVNLFPDSWG
S.par        371 KVAELELKNAIDYDRVDEHTKFLIFPISVIGCGKTTTSQILVNLFPESWG
S.mik        371 KVVELELRNAIDYEKVDEHTKFLIFPISVIGCGKTTTSQTLVNLFPESWG
S.bay        372 KVAELELRNAIDYDKVDEHTKFLIFPISVIGCGKTTTSQTLTNLFPESWG
S.cas        367 KVMQLELKNSIDYDKVDEHTKFLVFPIAVIGCGKTTTSVALTSLFPDSWG
S.klu        368 RLEELELQNAIDYDKVDERTKFLFIPIATIGCGKTTTALTLQNLFPNTWG
C.albicans   392 SDQLKNLALANGNEGLSTTTKYIFVPIATIGCGKTTVFNTLNNLFPQ-WT
S.pombe      367 --KLDLMHLSVSNDNDCGKEFTLLVPIATIGCGKTTVAKILEKLFGWPVV
consensus    401 .. ......................** .*******....* .**.....

                                          Nuc B
S.cer        421 HIQNDDITGKDKSQLMKKSLELLSKKEIKCVIVDRNNHQFRERKQLFEWL
S.par        421 HIQNDDITGKDKSQLMKKSLELLSKREIKCVIVDRNNHQFRERKQLFGWL
S.mik        421 HIQNDDITGKDKSQLMKKSLELLSKKEVKCVIVDRNNHQFRERQQLFDWL
S.bay        422 HIQNDDITGKDKSQLLKKSLELLAKQDMKCVIVDRNNHQLRERKQLFEWL
S.cas        417 HIQNDDITGKDKSKLMKKSLELLADPNIKCVIVDRNNHQFRERKQLFEWL
S.klu        418 HIQNDDITGKDKSMLMKRSLELLSRDEIKAVIVDRNNHQYRERKQLFQWF
C.albicans   441 HIQNDNISKKAKLKICDLTLLALEDDDQSVVLFDRNNSASRERRQIFTTI
S.pombe      415 QNDNLPSGKGGPKRFAKAIIEEFRNG-HSVVFADRNNHISNMR------S
consensus    451 ...*..................... ....*..****.....*.......
```

```
S.cer         471 NELKEDYLVYDTNIKVIGVSFAPYDKLS-EIRDITLQRVIKRGNNHQSIK
S.par         471 DELKEDYLAYDTNIKVIGVSFAPYDKLS-EIRDITMQRVIRRGNNHQSIK
S.mik         471 DELKEDYLAYDTNIKVIGVSFAPYDKLP-EIKDITMRRVINRGNNHQSIK
S.bay         472 DELKEDYLAYDTNIKVIGISFAPYDKLS-KIREITMQRVIKRGNNHQSIK
S.cas         467 DELKEDYLPYDTNIKVIALSFASYDDLD-KLKTLTVDRVLERGDAHQTIK
S.klu         468 DELKEDFLPYDANVKVIAFSYLPYNDID-ITSAITLKRILARGDNHQSIK
C.albicans    491 DQKRDEHLDDTVDLKYIAINFIPEDLSEEELWDITYNRVIQRGDNHQSIK
S.pombe       458 -TLQTDILALIDGVRFVALPFKHTPEVP----EFVQNRVLQRGDRHQSIK
consensus     501 .......*...................  ........*.. ** .**.**


S.cer         520 WDELGEKKVVGIMNGFLKRYQPVNLD-KSPDNMFDLMIELDFGQAD----
S.par         520 WDELGEKKVVGIMNGFLKRYQPVNLG-KTPDNMFDLMVELDFEGAD----
S.mik         520 WDELGEKKVVGIMSGFLKRYQPVSLR-RSPDNMFDLMVELDFGEVD----
S.bay         521 LDELGEKKVMGIMSGFLKRYQPCNID-RSPDNMFDLIIELDLSEVD----
S.cas         516 VDQYGKAKVLGIMGGFWKRFQPIIED-KSPDNRFDLIIQLNSLQKD----
S.klu         517 AVTDGEKKVTSIMRGFIQRYQPVNEN-KYPDSLFDMVIHLDVKEEN----
C.albicans    541 SQSD-ENLVESVMKGFIQRYQPINTS-RSPDDQFDHVIHLKLSKDENSSK
S.pombe       503 VSEG-VDKVKAIMNTFYKQYKPFDPAGNKHDANYDDIIELDPLIGS----
consensus     551  .......*...* .*.....*...  ...*...*....*.    .

                             PNK 1
S.cer         565 SSLTNAKQILNEIHKAYPILVPEIPKDDEIETAFRRSLDYKPTVRKIVGK
S.par         565 TSLTNAKQILNEIHKAYPILVPLVPKDDEIDTAFRRSLDYKPTVRKIVGK
S.mik         565 SSLTNAKQILSEIHKAYPILIPEIPKDDEIDASFRKSLDYKPTVQKRVGK
S.bay         566 SSLTNAKQILNELHRAYPILIPEVPTDDKIDEAFKKSLEYKPTVRKIVGK
S.cas         561 SSLMNVKTILKRLHDVYPILIPEIPSEESINIAFQKSLQYKPKMKERKVT
S.klu         562 SSLKNAKRILEALYEKYQILIPTVPSDEEISNAFKKSLEYKPKITKVVKS
C.albicans    589 SSLENVRIIIDDLVQNFPDLIKEKPADELINECFQKALDYKPTFVKNMTA
S.pombe       548 --LENARRIVNYFKKNIPELIPNDPSDDDYAAALNYAVNEYVPTYRKTFG
consensus     601 ..*.*...*..........*....*...... ..................


S.cer         615 GNNNQQKTP-KLIKPTYISAKIENYDEIIELVKRCIASDA----ELTEK-
S.par         615 GNGNQPKTP-KLIKPTYFSANINNYDEIIDLVKESVASDA----DLSEQ-
S.mik         615 GNCNQPKTP-KLIKPTYISAKIESYDEIIDLARQGIANIA----ELSQQ-
S.bay         616 GNGNQQKTP-KLIRPVYISVNVTEYDEIIDSVRQGVANDK----DLSQK-
S.cas         611 NNETVSGTPKKKVRPAYFSVNIKDGTPIKELISNVIEKPSNISEDAQTI-
S.klu         612 NNKSKEKEG--KYKPVYFSANIRVKQSILDEIWKLVDTKRD-KVLSVEG-
C.albicans    639 NTIKKDPTY------YGIAMHYSSILENLEIVSHNEHFQN----------
S.pombe       596 NDSKKIKNKITAEGITGSSTCFKKAPRYFGVLLDRKTVESSLVQVLTIAN
consensus     651  . .. ... ............  .........  .. .     ...

                                  2H-1
S.cer         659 --FKHLLASGKVQKELHITLGHVMSSREKE--AKKLWKSYCNRYTDQITE
S.par         659 --FKYLLTSGKVQKELHITLGHVISSREKE--AKKLWKSYCNRYTDQITE
S.mik         659 --FEDLLTSGKFQNELHITLGHVMSSREKE--GKRMWKSFCKRYTDQITK
S.bay         660 --FEGLLANGKFQKDFHITLGHVMSSREKE--GKKLWKLYLNRYHDQITE
S.cas         660 --LKNLLQNDNFQESFHITLAHVMQGKRGDPAEQELWKAYLKHYKKILKS
S.klu         658 --IEMLFEENKIRPEFHITLCHVITGKNGTKSQKDIWCEFNKRYSKDLVK
C.albicans    673 -------IKSHIQTEFHVTLGHIASSKQDK-AGRVKWKKLVKTLGKGDPN
S.pombe       646 LQWQEAFSRYTLQDSFHITMIHESQKPVNS----RIWEQYLQHMHDKNTT
consensus     701   .  ..  .... ..*.*..*........  ....*. ..........
```

```
S.cer        705 YNNNRIENAQGSGNNQNTQVKTTDKLNFRLEKLCWDEKIIAIVVELSKDK
S.par        705 YNNHRIGNTPAPGTNQDKQLKTTDTLKFRLEKLCWDEKIIAIVVELSGDK
S.mik        705 YNDRLIENTAGTSINQGKSVETTDKLKFRLKKLCWDQKIIAIVVELSGGE
S.bay        706 YNDNHKNNNSSPG---TKFVNTTDKLKFKLNKLCWDDKIIAIVVELSGNE
S.cas        708 YNDKEIPKLIETN----------DFVRFQPKMICWDDKIVSIVIKFPNDE
S.klu        706 TQPLSSRVPKLIK--------TKDVMKFKLDKLLWDDKIVTIVVKSEKEC
C.albicans   715 KPKSALKFFADVK---------LLQIVINTDKLACIKVEILKIYDTNDVL
S.pombe      692 KMGN---------------------ISFRITHLVWDDRVICFRVTMNENS
consensus    751 ..   . .           . ... ...................... .

                                       2H-2
S.cer        755 DGCIIDENNEKIKGLCCQNKIPHITLCKLESGVKAVYSNVLCEKVESAEV
S.par        755 DGCIINANNENVKGLRCQNKIPHITLCKLESGVKAVYSNVLCEKVESAEV
S.mik        755 GGCIIDSNGDNVKELYCQNKVPHITLCKLGSDVKAVYSNVLCEKVELNEI
S.bay        753 TGCLIDVNGDTVDGLKCLNKVPHITVCKIDCDVKAVYSNVLCEKAQSNEA
S.cas        748 --CVSDENGLLVTKLKCANKIPHITVARLQDNVKAVYSNTICKDVVENRD
S.klu        748 --VYDSTNELKVSRLECANDSPHITIGICSKGLNLFTRILCAM-------
C.albicans   756 QSEIEPINKQLHITIGCIPPATAVESNITLEELYDNPDEQELKPDGTYKC
S.pombe      721 -----------VWYGKTCNPQLHITLGTSSSDVKAFESNFLLKKLRWQGD
consensus    801  ..... . .  ... ........... ............... ...

S.cer        805 D-ENIKVVKLDNSKEFVGSVYLNF----
S.par        805 G-ENVKVVKLNSSKEFVANVYLNFYAAL
S.mik        805 D-ERVKVVKLNNAKEFAAHVYLNM----
S.bay        803 QNENVKVLELSNTKEFVGNVCLNF----
S.cas        796 A-TNINHVELVDSPSLTGNVCINL----
S.klu            ----------------------------
C.albicans   806 GDDTLHVFNFDNPDLKLFSQQLFVAYQ-
S.pombe      760 EVDSTDGNVRYLTVLPKIIIEGMLEPVY
```

***Figure A-3 Secondary structural prediction for the adenylylate synthetase domain of tRNA ligase, a member of the covalent nucleotidyl transferase protein family***

The amino acid sequence of the adenylylate synthetase domain of tRNA ligase was aligned with the predicted secondary structure (2ary CNT) and consensus primary sequence (con CNT) for proteins in the covalent nucleotidyl transferase (CNT) family (Aravind and Koonin, 1999). CNT-I through CNT-V are conserved motifs found in CNT proteins. As demonstrated in the structure of T7 DNA ligase, residues in the CNT-I, CNT-III, CNT-IIIa, and CNT-V motifs contact the∝–phosphate, the ribose sugar, the purine ring, and the ∝–phosphate of bound ATP respectively (Shuman, 1996). The conserved lysine to which AMP is covalently attached is highlighted in red. The single headed arrow marks the location of the H148Y mutation found in UPR-defective *rlg1-100* yeast. The double headed arrows mark the locations of two separate mutations that each cause temperature-sensitive defects in pre-tRNA splicing: F107S (*rlg1-10*) and T180I (*rlg1-4*).

```
                                  CNT-I
2ary CNT                   bbbbbbb  bbbbbbbb
S.cer        98 WNWIEENCTGPYDVTIKANGCIIFISGLEDGTLVVCSKHSTGPRAD----

con CNT                    hhhp KhpG  h h    s


                                ↓                     ↓         ↓
                                  CNT-III
2ary CNT                        bbbbbbbbbb
S.cer       144 ------------VDRNHAEAGEKQLLRQLAAMNINRSDFARMLYTHNVTA

con CNT                           hpuEhh


                             CNT-IIIa
2ary CNT                  bbbbbbbbbbaaaaaa
S.cer       182 VAEYCDDSFEEHILEYPLEKAGLYLHGVNVNKAEFETWDMKDVSQMASKY

con CNT                   phhh  hb    p


                                                  CNT-IV
2ary CNT                                      aa   bbbbbb
S.cer       232 GFRCVQCITSNTLEDLKKFLDNCSATGSFEGQEIEGFVIRCHLKST----

con CNT                                           -Ghlhp


                      CNT-V
2ary CNT             bbbbbb
S.cer       278 --EKPFFFKYKFEEPYLMYRQWREVTKDYIS-NKSRVFKFRKHKFITNKY

con CNT                hKhp
```

**K** = AMP attachment site

a = alpha helix
b = beta sheet

b = big residues (F, I, L, M, V, W, Y, K, R, E, Q)
h = hydrophobic residues (A, C, F, I, L, M, V, W, Y)
l = aliphatic residues (L, I, A, V, M)
p = polar residues (D, E, H, K, N, Q, R, S, T)
s = small residues (A, C, S, T, D, N, V, G, P)
- = negatively charged residues (D, E)
G = glycine
K = lycine
E = glutamate

***Figure A-4 Small islands of homology within the kinase domain of tRNA ligase***

Primary sequence alignments of tRNA ligase, T4 PNK kinase, and adenylate kinase highlight **(A)** a putative Walker A type and **(B)** a putative Walker B type nucleotide binding motif in tRNA ligase (Apostol et al., 1991; Koonin and Gorbalenya, 1990; Walker et al., 1982). A third region of similarity between tRNA ligase and T4 PNK is highlighted in **(C)** (Apostol et al., 1991).

## A

```
S.cer tRNA ligase       390  TKFLIFPISVIGCGKTTTSQTLVN
T4 PNK                    1  MKKIILTIGCPGSGKSTWAREIIA
Adenylate kinase          6  KSKIIFVVGGPGSGKGTQCEKIVQ
                                .*  .   * ****    ..

WALKER A Motif:                      GxxGxGK(S/T)
```

## B

```
S.cer tRNA ligase       444  SKKEIK-CVIVDRNNH
T4 PNK                   75   GDSVKGVIISDTNLN
                                 .*  ....* *

S.cer tRNA ligase       444  SKKEIK-CVIVDRNNH
Adenylate kinase        108   KIGQPTLLYVDAGPE
                              *      ...*

WALKER B Motif:                     hhhhD
```

## C

```
S.cer tRNA ligase       570  AKQILNEIHKAYPILVPEIPK
T4 PNK                  124  VKRNSKRGTKAVPIDVLRSMY
                             .*. .    ** ** *  .
```

***Figure A-5 Secondary structural prediction for the 2',3'-cyclic phosphodiesterase domain of tRNA ligase, a member of the 2H phosphoesterase protein family***

The amino acid sequence of the 2',3'-cyclic phosphodiesterase domain of tRNA ligase was aligned with the predicted secondary structure (2ary 2H) and consensus primary sequence (con 2H) for proteins in the 2H phosphoesterase family (Mazumder et al., 2002). The two 2H phosphoesterase family active site motifs found in tRNA ligase are highlighted in red.

```
2ary 2H                         bbbbbbbb aaaaaaaaaaaaaa
S.cer       615 GNNNQQKTP-KLIKPTYISAKIENYDEIIELVKRCIASDA----ELTEK-
con 2H                      h                 h


2ary 2H                          bbbbbb    aaa  aaaaaaaaaa
S.cer       659 --FKHLLASGKVQKELHITLGHVMSSREKE--AKKLWKSYCNRYTDQITE
con 2H                          Hhol               h


2ary 2H                        bbbbbb bbbbbb  bbbbbbbbb aaaaaaaaaa
S.cer       705 YNNNRIENAQGSGNNQNTQVKTTDKLNFRLEKLCWDEKIIAIVVELSKDK
con 2H                             h h     h     hh          h


2ary 2H         aaaaa                bbbbb      aaaaaaaaaaaaaa   b
S.cer       755 DGCIIDENNEKIKGLCCQNKIPHITLCKLESGVKAVYSNVLCEKVESAEV
con 2H                                Hhol             h


2ary 2H         bbb bbbbbbb
S.cer       805 DENIKVVKLDNSKEFVGSVYLNF----
con 2H                h
```

**HITL** = 2H active site motif

a = alpha helix
b = beta sheet

h = hydrophobic residue (L, I, Y, F, M, W, A, C, V)
l = aliphatic residue (L,I, A, V)
o = alcohol group residues (S, T)
H = histidine

***Figure A-6 Primary sequence alignment of tRNA ligase and yeast Ski2p***

The primary sequence of the tRNA ligase protein was used in a BLAST search for homologues. Three regions within the putative RNA helicase Ski2p of *S.cerevisiae* consistently came up as the highest scoring matches. **(A)** The largest region of similarity between the two proteins includes portions of the kinase and phosphodiesterase domains of tRNA ligase (see FigA2-1) and half of what is postulated to be required for Ski2p helicase function (de la Cruz et al., 1999; Tanner and Linder, 2001). The helicase motifs I, Ia, Ib, II, and III of Ski2p are boxed. For comparison, the two putative 2H phosphoesterase active site motifs of tRNA ligase are highlighted in red (see Fig A2-5). For this alignment, the calculated amino acid identity and overall similarity was 22% (61/274) and 40% (111/274) respectively. **(B)** A second region of sequence similarity. Here the calculated amino acid identity and overall similarity was 31% (12/38) and 60% (23/38) respectively. **(C)** A third, very small region of sequence similarity. Here, the calculated amino acid identity and overall similarity was 64% (11/17) and 70% (12/17) respectively.

**A**

```
Ligase    533 NGFLKRYQPVNLDKSPDNMFDLMIELDFGQADSSLTNAKQILNEI-HKAYPILVPEIPKD
              NG  K  + +N     DN  D+ IE +  +      +AK I  EI  +A      +   D
Ski2      217 NGQFKELKQLN---EIDNELDIRIEANEAKLKEEEKSAKSISEEIMEEATEETTADNADD


Ligase    592 DEIETAFRRSLDYKPTVRKIVGKGNNNQQKTPKLIKPTYISAKIENYDEIIELVKRCIAS
               EI+      +D+  T  K V K    +++   ++    ++ KIEN+DE+I    R
Ski2      274 AEIDELLPIGIDFGRT--KPVSKSVPVKKEWAHVVD---LNHKIENFDELIPNPARSWPF

                                       I                           Ia
Ligase    652 DAELTEK--FKHLLASGKVQKELHITLGHVMSSREKEAKKLWKSYCNRYTDQITEYNNNR
              + +  +K    HL     V    H + G  + +    A          YT  I   +N +
Ski2      329 ELDTFQKEAVYHLEQGDSVFVAAHTSAGKTVVAEYAIAMAHRNMTKTIYTSPIKALSNQK

                                              Ib                      II
Ligase    710 IENAQGSGNNQNTQVKTTD-KLNFRLEKLCWDEKIIAIVV----ELSKDKDGCIIDEN--
                + + + ++ N  + T D ++N     L    +I+  ++    +L +D +  I DE
Ski2      389 FRDFKETFDDVNIGLITGDVQINPDANCLIMTTEILRSMLYRGADLIRDVEFVIFDEVHY

                                        III
Ligase    763 -NEKIKGLCCQNKI---P-HITLCKLESGVKAVY
               N++ +G+  +  I   P H+    L +  V   Y
Ski2      449 VNDQDRGVVWEEVIIMLPQHVKFILLSATVPNTY
```

**B**

```
Ligase    782 KLESGVKAVYSNVL-CEKVESAEVDENIKVVKLDNSKE
              +L+S  +  Y+ +L   ++E+  V+E IK    +N+KE
Ski2      799 RLQSQFRLTYNMILNLLRIEALRVEEMIKYSFSENAKE
```

**C**

```
Ligase    457 NHQFRERKQLFEWLNEL
              N QF+E KQL E  NEL
Ski2      217 NGQFKELKQLNEIDNEL
```

***Figure A-7 A model for binding of cleaved tRNA by T4 polynucleotide kinase***

In this hypothetical model, a tRNA cleaved in the anti-codon loop interacts with two different monomers of T4 polynucleotide kinase, one in red, and the other in blue. The kinase domain of the red monomer binds the 5'-hydroxyl end of the cleaved tRNA (highlighted in yellow) while the phosphatase domain of the blue monomer binds the 3'-phosphaste end (highlighted in orange). The distance and orientation between the kinase and phosphatase domains of adjacent monomers is just right to accommodate this interaction. This is not the case for the kinase and phosphatase domains of a single monomer. This figure was kindly provided to us by Barry Stoddard. It can also be found in a paper recently published by Stoddard and his colleagues on the crystal structure of T4 PNK (Galburt et al., 2002).

## ***REFERENCES***

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. J Biol Chem *273*, 12685-12688.

Amitsur, M., Levitz, R., and Kaufmann, G. (1987). Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. Embo J *6*, 2499-2503.

Apostol, B. L., and Greer, C. L. (1991). Preferential binding of yeast tRNA ligase to pre-tRNA substrates. Nucleic Acids Res *19*, 1853-1860.

Apostol, B. L., Westaway, S. K., Abelson, J., and Greer, C. L. (1991). Deletion analysis of a multifunctional yeast tRNA ligase polypeptide. Identification of essential and dispensable functional domains. J Biol Chem *266*, 7445-7455.

Aravind, L., and Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol *287*, 1023-1040.

de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci *24*, 192-198.

Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258-2261.

Galburt, E. A., Pelletier, J., Wilson, G., and Stoddard, B. L. (2002). Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. Structure (Camb) *10*, 1249-1260.

Gegenheimer, P., Gabius, H. J., Peebles, C. L., and Abelson, J. (1983). An RNA ligase from wheat germ which participates in transfer RNA splicing in vitro. J Biol Chem *258*, 8365-8373.

Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. Embo J *18*, 3119-3132.

Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983). Mechanism of action of a yeast RNA ligase in tRNA splicing. Cell *32*, 537-546.

Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. Cell *89*, 545-553.

Hofmann, A., Grella, M., Botos, I., Filipowicz, W., and Wlodawer, A. (2002). Crystal structures of the semireduced and inhibitor-bound forms of cyclic nucleotide phosphodiesterase from Arabidopsis thaliana. J Biol Chem *277*, 1419-1425.

Hofmann, A., Zdanov, A., Genschik, P., Ruvinov, S., Filipowicz, W., and Wlodawer, A. (2000). Structure and mechanism of activity of the cyclic phosphodiesterase of Appr>p, a product of the tRNA splicing reaction. Embo J *19*, 6207-6217.

Koonin, E. V., and Gorbalenya, A. E. (1990). Related domains in yeast tRNA ligase, bacteriophage T4 polynucleotide kinase and RNA ligase, and mammalian myelin 2',3'-cyclic nucleotide phosphohydrolase revealed by amino acid sequence comparison. FEBS Lett *268*, 231-234.

Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. (2000). Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. Embo J *19*, 1119-1129.

Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989). Birth of the D-E-A-D box. Nature *337*, 121-122.

Linder, P., Tanner, N. K., and Banroques, J. (2001). From RNA helicases to RNPases. Trends Biochem Sci *26*, 339-341.

Mazumder, R., Iyer, L. M., Vasudevan, S., and Aravind, L. (2002). Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. Nucleic Acids Res *30*, 5229-5243.

Nasr, F., and Filipowicz, W. (2000). Characterization of the Saccharomyces cerevisiae cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1",2"-cyclic phosphate. Nucleic Acids Res *28*, 1676-1683.

Odell, M., Sriskanda, V., Shuman, S., and Nikolov, D. B. (2000). Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. Mol Cell *6*, 1183-1193.

Phizicky, E. M., Consaul, S. A., Nehrke, K. W., and Abelson, J. (1992). Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. J Biol Chem *267*, 4577-4582.

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986). Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. J Biol Chem *261*, 2978-2986.

Ruegsegger, U., Leber, J. H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103-114.

Schwartz, R. C., Greer, C. L., Gegenheimer, P., and Abelson, J. (1983). Enzymatic mechanism of an RNA ligase from wheat germ. J Biol Chem *258*, 8374-8383.

Shuman, S. (1996). Closing the gap on DNA ligase. Structure *4*, 653-656.

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response [see comments]. Cell *87*, 405-413.

Singleton, M. R., Hakansson, K., Timson, D. J., and Wigley, D. B. (1999). Structure of the adenylation domain of an NAD+-dependent DNA ligase. Structure Fold Des *7*, 35-42.

Subramanya, H. S., Doherty, A. J., Ashford, S. R., and Wigley, D. B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. Cell *85*, 607-615.

Tanner, N. K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell *8*, 251-262.

Thogersen, H. C., Morris, H. R., Rand, K. N., and Gait, M. J. (1985). Location of the adenylylation site in T4 RNA ligase. Eur J Biochem *147*, 325-329.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.

Vasudevan, S., Peltz, S. W., and Wilusz, C. J. (2002). Non-stop decay--a new mRNA surveillance pathway. Bioessays *24*, 785-788.

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. Embo J *1*, 945-951.

Wang, L. K., Lima, C. D., and Shuman, S. (2002). Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. Embo J *21*, 3873-3880.

Wang, L. K., and Shuman, S. (2002). Mutational analysis defines the 5'-kinase and 3'-phosphatase active sites of T4 polynucleotide kinase. Nucleic Acids Res *30*, 1073-1080.

Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990). Domain structure in yeast tRNA ligase. Biochemistry *29*, 6132-6138.

Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1991). Conserved mechanism of tRNA splicing in eukaryotes. Mol Cell Biol *11*, 5410-5416.

# Appendix B

## *The possible role of the NAD-dependent 2'-phosphotransferase, Tpt1p in the unfolded protein response pathway*

## *INTRODUCTION*

Splicing of nuclear encoded pre-tRNAs is ubiquitous in eukaryotes, and is best understood in the budding yeast *Saccharomyces cerevisiae*. About 20% of yeast nuclear tRNA genes carry a 16 to 60 nucleotide long intron one base 3' of the anticodon (Ogden et al., 1984; Trotta et al., 1997). Intron removal is catalyzed by the sequential action of three enzymes: tRNA endonuclease, tRNA ligase, and (NAD)-dependent 2' phosphotransferase (Abelson et al., 1998; Westaway and Abelson, 1995). The cleavage and ligation steps are thought to occur at the nuclear pore, where tRNA ligase has been visualized by immuno-gold electron microscopy (Clark and Abelson, 1987). Splicing begins when tRNA endonuclease cleaves the pre-tRNA to produce 5'-OH and 2',3'-phosphate termini. In a multistep process, the multifunctional tRNA ligase then joins the two exons to produce a tRNA with a 2'-phosphate at the splice junction. In the final step, the splice junction 2'-phosphate is removed by the 2'-phosphotransferase, Tpt1p, which transfers the phosphate to nicotinamide adenine dinucleotide (NAD), forming the small molecule ADP-ribose 1"-2"-cyclic phosphate (abbreviated as Appr>p) in the process (Culver et al., 1994; Culver et al., 1993; McCraith and Phizicky, 1991). Unlike tRNA ligase, the subcellular localization of Tpt1p has not been determined. Its concentration *in vivo* has been estimated to be 1000-5000 molecules per cell, 10-fold greater than that for either tRNA ligase or tRNA endonuclease (Culver et al., 1997; Phizicky et al., 1986; Rauhut et al., 1990). Tpt1p is found in organisms representing the three domains of life. Equivalent biochemical activities have been detected in xenopus oocytes, HeLa cells, and bacteria; Tpt1p orthologues have been identified in mammals, plants, bacteria, and archaea (Culver et al., 1993; Spinelli et al., 1998; Zillman et al., 1992).

Tpt1p removes the splice junction 2'-phosphate by transferring it to nicotinamide adenine dinucleotide (NAD), in the process, producing the small molecule ADP-ribose 1"-2"-cyclic phosphate (Appr>p) and releasing nicotinamide. When this mechanism was initially worked out, it became apparent that in the absence of Appr>p degradation, the intracellular concentrations of this metabolite were expected to be very high, reaching 10-40 μM (Culver et al., 1993). Thus, a degradation pathway was speculated to exist and soon was found (Culver et al., 1994). Appr>p is converted to ADP-ribose 1"-phosphate (Appr-1"p) by Cpd1p, a cyclic nucleotide phosphodiesterase found in yeast and plants (Genschik et al., 1997; Nasr and Filipowicz, 2000). Interestingly, Cpd1p is related to the 2',3'-cyclic phosphodiesterase domain of tRNA ligase. Both are members of the 2H phosphoesterase protein family (Mazumder et al., 2002; Nasr and Filipowicz, 2000). Deletion or overexpression of the *CPD1* gene results in no discernable phenotype in yeast, suggesting that Appr>p and Appr-1"p lack important roles in yeast growth and physiology (Nasr and Filipowicz, 2000).

Nonetheless, removal of the splice junction 2'-phosphate of spliced tRNAs is required for cell viability (Culver et al., 1997) and appears to be necessary in order to form fully functional tRNAs. Cells in which steady state levels of the 2'-phosphotransferase are reduced accumulate fully spliced tRNAs bearing the 2'-phosphate and an undermodified base at the splice junction (Spinelli et al., 1997). This finding is important on two levels. First it demonstrates that 2'-phosphate removal need not be coordinated with pre-tRNA cleavage and ligation for these latter two processes to take place. Second, it provides a plausible explanation for why a null mutation in the 2'-phosphotransferase is lethal in yeast (Culver et al., 1997). The inability to modify tRNA bases can adversely affect translation and growth in yeast (Bjork et al., 2001; Lecointe et al., 1998; Pintard et al., 2002). Though this likely explains the essential nature of Tpt1p in yeast, it remains to be demonstrated if translation is adversely affected in Tpt1p depleted cells.

The yeast *HAC1* mRNA was the first mRNA discovered that is spliced utilizing a pre-tRNA like mechanism. During the unfolded protein response (UPR), *HAC1* mRNA is cleaved at two sites by the endoribonuclease Ire1p. The two resulting exons are then joined by tRNA ligase via the same multistep process used during pre-tRNA splicing, leaving a 2'-phosphate at the splice junction (Gonzalez et al., 1999; Sidrauski and Walter, 1997). Removal of the intron is required for translation of the *HAC1* mRNA (Chapman and Walter, 1997). It is unclear if Tpt1p removes the splice junction phosphate *in vivo* as it does for spliced pre-tRNAs. Unlike tRNAs, the nucleotides of mRNAs are not normally modified. Thus if removal of the 2'-phosphate is necessary for the biological activity of *HAC1* mRNA, it is unlikely that it involves base modifications. Perhaps the splice junction 2'-phosphate might inhibit translation of *HAC1* mRNA or disrupt the proper subcellular localization of the *HAC1* message if is not promptly removed.

In this study, we set out to investigate if reduction of Tpt1p levels negatively affected the ability of yeast to grow on UPR-inducing media as well as produce Hac1p. To deplete *in vivo* levels of Tpt1p, we used an approach previously taken by Phizicky and colleagues (Spinelli et al., 1997).

## *MATERIALS AND METHODS*

The yeast strains used in this study are listed in Table A-1A. Plasmids pGAL-*TPT1*, pGAL-*tpt1-1*, and pGAL-*tpt1-2* were isolated from yeast strains SC1118, SC1120, and SC1119 obtained from the Phizicky laboratory (Department of Biochemistry and Biophysics, University of Rochester School of Medicine; Rochester, New York). Each plasmid carries a centromere (*CEN IV*), the *URA3* gene, and a *TPT1* gene under control of an inducible galactose promoter (*GAL10*) (Spinelli et al., 1997). The *tpt1-1* allele carries a nonsense mutation at amino acid 23. Strains carrying this plasmid as their only source of the *TPT1* protein have about 5% of the 2'-phosphotransferase activity of strains carrying a plasmid with the wild type gene instead. Presumably read-through of the *tpt1-1* stop codon accounts for the 5% activity (Spinelli et al., 1997). The *tpt1-2* allele is a serine to phenylalanine change at amino acid 15. It produces a more severe phenotype than the *tpt1-1* allele (E. Phizicky, personal communication).

To make the *TPT1* deletion strains used in this study, we first knocked out a single copy of the gene in a diploid strain. We amplified a DNA cassette encoding the *S.pombe HIS5* marker using oligonucleotide primers, each of which carried sequences homologous to the very 5' and 3' ends of the targeted *TPT1* gene (Longtine et al., 1998). This DNA cassette was then transformed into the diploid strain. *TPT1* diploid knock-outs were confirmed by tetrad analysis. A confirmed *Δtpt1::HIS5/TPT1* diploid was transformed with each of the pGAL plasmids. Transformants were sporulated on galatose containing media. We confirmed that the *Δtpt1::HIS5* haploids we obtained required the pGAL plasmids for viability.

Media, Northern blot analysis, and Western blot analysis were prepared or carried out as described in Chapter 4 of this thesis.

## *RESULTS*

The yeast *TPT1* gene encodes the (NAD)-dependent 2'-phosphotransferase. The Phizicky laboratory has isolated various *TPT1* alleles that, while they support cellular growth, are deficient in catalytic activity relative to the wild type gene. We obtained plasmids that carry the two variants, *tpt1-1* and *tpt1-2*, or the wild type *TPT1* gene under transcriptional control of an inducible galatose promoter. These plasmids rescue grow of *TPT1* deletion strains when grown in the presence of galactose (Spinelli et al., 1997).

Our first step was to determine if yeast bearing the catalytically crippled *tpt1* alleles were more sensitive than wild type *TPT1* yeast to growth in the presense of the UPR inducing agent tunicamycin. To this end, we plated the strains onto galactose containing media supplemented with varying amounts of tunicamycin. When plated at tunicamycin concentrations where wild type strains grow robustly, the *tpt1-1* and *tpt1-2* strains grew poorly or not at all (FigureB-1). In particular, the *tpt1-2* strain was as sensitive to growth on tunicamycin plates as a strain carrying the UPR defective *rlg1-100* tRNA ligase allele.

Our next step was to determine if the defective 2'-phosphotransferases affected *HAC1* mRNA splicing and translation of the spliced message. To this end, we exposed mid-log liquid cultures of yeast to DTT for 30 minutes in order to induce the UPR. We then worked up total RNA and protein extracts and analyzed the samples by Northern and Western blot respectively. Wild type strains spliced approximately 80% of their *HAC1* mRNA, (FigureB-2A, lanes 2 and 4), whereas *tpt1-1* and *tpt1-2* spliced approximately 60% and 50% respectively (FigureB-2A, lanes 8 and 6). Simultaneously, *tpt1-1* and *tpt1-2* strains both displayed approximately two fold less total *HAC1* mRNA than wild type strains (FigureB-2A,compare lanes 1, 3, 5, and 7). Thus it appears that defects in the 2'-phosphotransferase reduce *HAC1* mRNA steady state levels as well as *HAC1* mRNA splicing.

2'-phosphotransferase defects also appear to reduce the amount of Hac1p produced when *HAC1* mRNA is spliced following UPR induction. Whereas wild type yeast produced readily detectible amounts of Hac1p (FigureB-2B, lanes 2 and 4), the *tpt1-2* strain made none and the *tpt1-1* strain made a barely detectible amount (FigureB-2B, lanes 6 and 8).

## ***DISCUSSION***

We have demonstrated that yeast defective for (NAD)-dependent 2'-phosphotransferase activity are also sensitive to growth in the presence of the UPR inducing chemical tunicamycin. These same strains also splice *HAC1* mRNA and produce Hac1p to a lesser degree than wild type strains and have lower steady state levels of *HAC1* mRNA than wild type. Thus, reduced growth on UPR-inducing media is likely a consequence of reduced Hac1p levels.

It seems plausible that reduced *HAC1* mRNA levels in combination with decreased *HAC1* mRNA splicing were responsible for the tunicamycin sensitivity and poor Hac1p production. However, we also cannot discount the possibility that global translation levels are significantly reduced in the *tpt1* mutant strains as well, and that this also was a contributing factor. An overall reduction in translation might explain the decrease in *HAC1* mRNA seen in the two mutant strains: perhaps these strains have lower steady state levels of the cellular transcriptional machinery. Likewise, reduced splicing of *HAC1* mRNA could reflect reduced steady state levels of the splicing enzymes Ire1p and tRNA ligase.

These complicated possibilities highlight the difficulty of working with mutations that might affect global translation. The approach we took in these studies cannot adequately tell the difference between reductions in Hac1p production that are a consequence of global cellular defects versus those that are specific to translation of the spliced *HAC1* mRNA itself. We need a different approach to do so.

## *FUTURE DIRECTIONS*

### *Testing the translational block model*

Even with the aforementioned complications, we believe it is possible to determine if the presence of a 2'-phosphate at the splice junction reduces translation of the spliced *HAC1* message *in vivo*. One could do so by designing a *HAC1* gene construct in which a stop codon is placed just upstream of the 5' splice site. Translation of this variant would still require splicing to relieve the translational block imposed by the *HAC1* intron. However, unlike when wild type *HAC1* mRNA is spliced, in this case, the translation stop codon will be located upstream of the splice junction carrying the 2'-phosphate. Thus, unlike the wild type scenario, ribosomes translating the stop codon variant *HAC1* mRNA would be expected to never encounter the splice junction 2'-

phosphate. By comparing the amount of *HAC1* protein made in *tpt1* mutant cells carrying a wild type or stop-codon variant *HAC1* gene, one could determine if the location of the splice junction 2'-phosphate influences the translation of the message.

### *The possible roles of Appr>p and Appr-1"p during the UPR*

We have previously speculated that the removal of the *HAC1* splice junction phosphate might cause the cellular levels of Appr>p to spike during the UPR and that perhaps this spike is somehow integrated into the UPR signaling pathway (Gonzalez et al., 1999). Two other NAD derived small molecules, cyclic ADP-ribose (sADPR) and nicotinic acid adenine dinucleotide phosphate (NAADP), are involved in intracellular signaling. ADPR is thought to stimulate release of $Ca^{++}$ from the ER by binding to the ER ryanodine receptor in mammalian cells; NAADP stimulates release of $Ca^{++}$ from lysosomes (Fill and Copello, 2002; Lee, 2001; Lee, 2003). Thus there is precedent for NAD-derived small molecule signaling in eukaryotes.

The yeast Cpd1p cyclic nucleotide phosphodiesterase converts Appr>p to Appr-1"p. If either small molecule were important to normal cell growth, then an increase or decrease in either molecule should have noticeable affects on cell growth. By deleting the the *CPD1* gene or causing its overexpression, the levels of both of these molecules should change in a reciprocal manner. When this was done in yeast, no discernable change in the growth phenotype of the strains was detected, suggesting that Appr>p and Appr-1"p lack important roles in yeast growth and physiology (Nasr and Filipowicz, 2000). However, this does not preclude a role in the UPR for either of these small molecules, as the deletion and overexpression *CPD1* strains were not tested for growth under UPR inducing conditions. This simple experiment should be pursued.

Vertebrates appear to have two distinct tRNA ligation activities. The first behaves like the yeast tRNA ligase, Rlg1p. The tRNA exons are ligated such that the phosphate that ultimately links the joined exons comes from ATP or GTP and the 2'-phosphate that marks the splice junction comes from the 3'-terminus of the 5' exon (Zillmann et al., 1991). In contrast, the second tRNA ligase activity catalyzes joining such that the phosphate that links the two exons together comes from the 3'-terminus of the 5'exon and the final spliced product lacks a junction 2'-phosphate (Filipowicz and Shatkin, 1983; Laski et al., 1983; Nishikura and De Robertis, 1981). Why have two ligation activities in higher eukaryotes?

We and others have speculated that unlike in yeast, there may be a division of labor in higher eukaryotes: the yeast-like pathway being used predominantly for splicing of *HAC1* mRNA homologues such as Xbp1 and the non-yeast like pathway being used predominantly for pre-tRNA splicing (Urano et al., 2000). A number of factors make this appealing and lead to some interesting predictions. First, in xenopus, pre-tRNA splicing via the non-yeast pathway has been shown to take place in the nucleus (De Robertis et al., 1981; Nishikura and De Robertis, 1981). If UPR-induced Xbp1 mRNA splicing takes place in the cytoplasm, as *HAC1* mRNA splicing does in yeast (Ruegsegger et al., 2001), then we predict that the yeast-like ligase activity will be localized to the xenopus cytosol. Second, Tpt1p orthologues and biochemical activities occur in mammals, frogs, and plants (Culver et al., 1993; Spinelli et al., 1998; Zillman et al., 1992). If indeed pre-tRNA splicing in these higher eukaryotes does not produce spliced tRNAs carrying 2'-

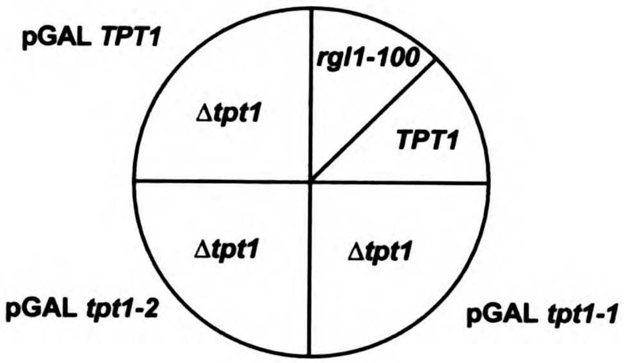
phosphates, then Tpt1p might be used predominantly to dephosphorylate mRNAs spliced via the UPR pathway that do carrying the 2'-phosphastes. If so, this would produce an even greater spike of Appr>p above background than we predict occurs in yeast and might thus be more readily integrated into the UPR signaling pathway in higher eukaryotes than in yeast. It would be very interesting to measure levels of Appr>p during UPR induction in a higher eukaryote, especially during B-cell differentiation when the mRNA for mammalian *HAC1* homologue XBP1 is spliced (Gass et al., 2002; Iwakoshi et al., 2003). It may be possible to adapt a recently described technique that is able to detect the related small molecules ADPR and NAADP with nanomolar sensitivity (Graeff and Lee, 2002a; Graeff and Lee, 2002b).

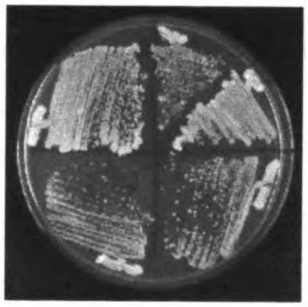
## *Table B-1* *Appendix B plasmid list*

| STRAIN | GENOTYPE |
|---|---|
| W303-1A | *MATa; leu2-3,-112; ura3-1,-112; his3-11,-15; trp1-1; ade2-1; can1-100* |
| CSY228 | same as W303-1A, except *MATα*. CSY228 also known as PWY373 |
| TGy-1 | same as CSY228, except *RLG1* replaced by *rlg1-100* using pTG-1 (see materials and methods, Chapter 3) |
| TGy-113 | same as CSY228 except genomic *HAC1* replaced with *HA-HAC1* using pTG-28 (see materials and methods, Chapter 3) |
| TGy-124 | same as TGy-113 except *Δtpt1::HIS5(S.pombe)*, and carries pGAL-*TPT1* |
| TGy-125 | same as TGy-124 except carries pGAL-*tpt1-2* |
| TGy-126 | same as TGy-124 except carries pGAL-*tpt1-1* |

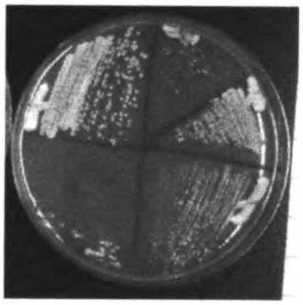
***Figure B-1 Growth of wild type and mutant Tpt1 strains on UPR-inducing media***

Strains were struck out onto synthetic galactose-raffinose media supplemented with 0, 0.125, or 0.166 µg/ml tunicamycin and allowed to grow at 30°C for 2 days. The genotype of each strain and the plasmid each strain carries are denoted inside and outside of the circle respectively.

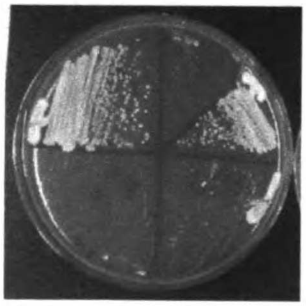

pGAL *TPT1*
*rgl1-100*
Δ*tpt1*
*TPT1*
Δ*tpt1*
Δ*tpt1*
pGAL *tpt1-2*
pGAL *tpt1-1*



**SGR-ura**



**SGR-ura**
**+**
**0.125 Tm**



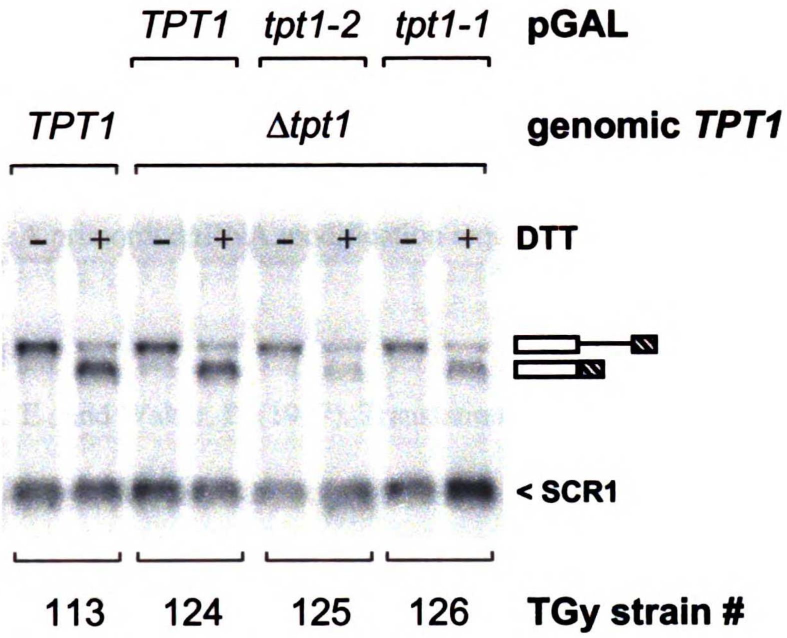
**SGR-ura**
**+**
**0.166 Tm**

***Figure B-2 HAC1 mRNA splicing and Hac1p production in Tpt1 strains***

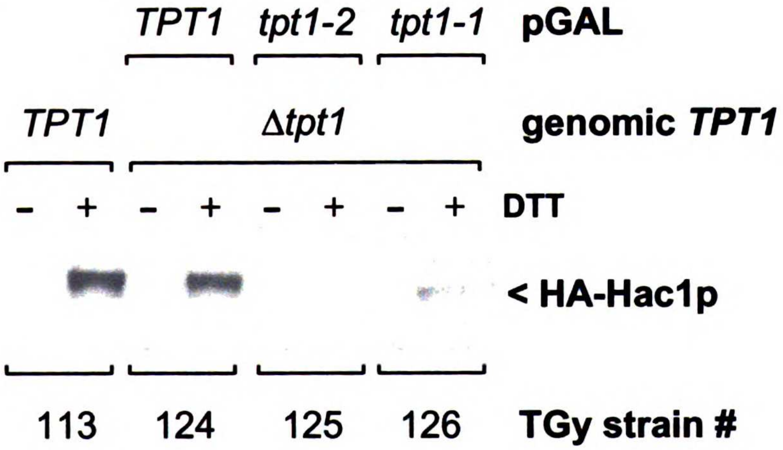
**(A)** Strains were grown to mid-log phase at 30°C and the UPR induced for 30 minutes by the addition of DTT to a final concentration of 8 mM. Northern blots were probed for *HAC1* mRNA and the *SCR1* RNA (included as a loading control). Full length and spliced forms of *HAC1* mRNA are indicated along the right. Strain names (TGy), strain genotype, and the plasmids (pGAL) carried by each are also shown.

***(B)*** Proteins extracts made from the same samples used in the Northern analysis. HA-tagged Hac1p was visualized using an antibody against the HA-epitope.

**A**


*TPT1* *tpt1-2* *tpt1-1* pGAL
*TPT1* *Δtpt1* genomic *TPT1*
- + - + - + - + DTT
< SCR1
113 124 125 126 TGy strain #

**B**


*TPT1* *tpt1-2* *tpt1-1* pGAL
*TPT1* *Δtpt1* genomic *TPT1*
- + - + - + - + DTT
< HA-Hac1p
113 124 125 126 TGy strain #

## ***REFERENCES***

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. J Biol Chem *273*, 12685-12688.

Bjork, G. R., Jacobsson, K., Nilsson, K., Johansson, M. J., Bystrom, A. S., and Persson, O. P. (2001). A primordial tRNA modification required for the evolution of life? Embo J *20*, 231-239.

Chapman, R. E., and Walter, P. (1997). Translational attenuation mediated by an mRNA intron. Curr Biol *7*, 850-859.

Clark, M. W., and Abelson, J. (1987). The subnuclear localization of tRNA ligase in yeast. J Cell Biol *105*, 1515-1526.

Culver, G. M., Consaul, S. A., Tycowski, K. T., Filipowicz, W., and Phizicky, E. M. (1994). tRNA splicing in yeast and wheat germ. A cyclic phosphodiesterase implicated in the metabolism of ADP-ribose 1",2"-cyclic phosphate. J Biol Chem *269*, 24928-24934.

Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., and Phizicky, E. M. (1997). A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. J Biol Chem *272*, 13203-13210.

Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., LaReau, R. D., Turner, D. H., and Phizicky, E. M. (1993). An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. Science *261*, 206-208.

De Robertis, E. M., Black, P., and Nishikura, K. (1981). Intranuclear location of the tRNA splicing enzymes. Cell *23*, 89-93.

Filipowicz, W., and Shatkin, A. J. (1983). Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. Cell *32*, 547-557.

Fill, M., and Copello, J. A. (2002). Ryanodine receptor calcium release channels. Physiol Rev *82*, 893-922.

Gass, J. N., Gifford, N. M., and Brewer, J. W. (2002). Activation of an unfolded protein response during differentiation of antibody-secreting B cells. J Biol Chem *277*, 49047-49054.

Genschik, P., Hall, J., and Filipowicz, W. (1997). Cloning and characterization of the Arabidopsis cyclic phosphodiesterase which hydrolyzes ADP-ribose 1",2"-cyclic phosphate and nucleoside 2',3'-cyclic phosphates. J Biol Chem *272*, 13211-13219.

Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. Embo J *18*, 3119-3132.

Graeff, R., and Lee, H. C. (2002a). A novel cycling assay for cellular cADP-ribose with nanomolar sensitivity. Biochem J *361*, 379-384.

Graeff, R., and Lee, H. C. (2002b). A novel cycling assay for nicotinic acid-adenine dinucleotide phosphate with nanomolar sensitivity. Biochem J *367*, 163-168.

Iwakoshi, N. N., Lee, A. H., Vallabhajosyula, P., Otipoby, K. L., Rajewsky, K., and Glimcher, L. H. (2003). Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. Nat Immunol.

Laski, F. A., Fire, A. Z., RajBhandary, U. L., and Sharp, P. A. (1983). Characterization of tRNA precursor splicing in mammalian extracts. J Biol Chem *258*, 11974-11980.

Lecointe, F., Simos, G., Sauer, A., Hurt, E. C., Motorin, Y., and Grosjean, H. (1998). Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. J Biol Chem *273*, 1316-1323.

Lee, H. C. (2001). Physiological functions of cyclic ADP-ribose and NAADP as calcium messengers. Annu Rev Pharmacol Toxicol *41*, 317-345.

Lee, H. C. (2003). Calcium Signaling: NAADP Ascends as a New Messenger. Curr Biol *13*, R186-188.

Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Mazumder, R., Iyer, L. M., Vasudevan, S., and Aravind, L. (2002). Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. Nucleic Acids Res *30*, 5229-5243.

McCraith, S. M., and Phizicky, E. M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. J Biol Chem *266*, 11986-11992.

Nasr, F., and Filipowicz, W. (2000). Characterization of the Saccharomyces cerevisiae cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1",2"-cyclic phosphate. Nucleic Acids Res *28*, 1676-1683.

Nishikura, K., and De Robertis, E. M. (1981). RNA processing in microinjected Xenopus oocytes. Sequential addition of base modifications in the spliced transfer RNA. J Mol Biol *145*, 405-420.

Ogden, R. C., Lee, M. C., and Knapp, G. (1984). Transfer RNA splicing in Saccharomyces cerevisiae: defining the substrates. Nucleic Acids Res *12*, 9367-9382.

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986). Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. J Biol Chem *261*, 2978-2986.

Pintard, L., Lecointe, F., Bujnicki, J. M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002). Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. Embo J *21*, 1811-1820.

Rauhut, R., Green, P. R., and Abelson, J. (1990). Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. J Biol Chem *265*, 18180-18184.

Ruegsegger, U., Leber, J. H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103-114.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell *90*, 1-20.

Spinelli, S. L., Consaul, S. A., and Phizicky, E. M. (1997). A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. Rna *3*, 1388-1400.

Spinelli, S. L., Malik, H. S., Consaul, S. A., and Phizicky, E. M. (1998). A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. Proc Natl Acad Sci U S A *95*, 14136-14141.

Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. Cell *89*, 849-858.

Urano, F., Bertolotti, A., and Ron, D. (2000). IRE1 and efferent signaling from the endoplasmic reticulum. J Cell Sci *113 Pt 21*, 3697-3702.

Westaway, S. K., and Abelson, J. (1995). Splicing of tRNA Precursors. In tRNA : structure, biosynthesis, and function, D. Söll, and U. RajBhandary, eds. (Washington, D.C., ASM Press), pp. 79-92.

Zillman, M., Gorovsky, M. A., and Phizicky, E. M. (1992). HeLa cells contain a 2'-phosphate-specific phosphotransferase similar to a yeast enzyme implicated in tRNA splicing. J Biol Chem *267*, 10289-10294.

Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1991). Conserved mechanism of tRNA splicing in eukaryotes. Mol Cell Biol *11*, 5410-5416.

UC San Francisco LIBRARY
UNIVERSITY OF CALIFORNIA SAN FRANCISCO

For reference
Not to be taken from the room.

7230652
3 1378 00723 0652