UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Genetic analysis of the U5 snRNA and the 3' splice site in yeast

Permalink
https://escholarship.org/uc/item/6vm3j1x6

Author
Patterson, Bruce,

Publication Date
1989

Peer reviewed|Thesis/dissertation

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

Genetic Analysis of the U5 snRNA and the 3' Splice Site in Yeast

by

Bruce Patterson

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

GENETICS

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Date JUN 11 1989 University Librarian

Degree Conferred: . . . .

I

support

try my

was 2000

DEDICATION

I dedicate this thesis to my parents, whose unwavering love and support during my life have given me the strength and confidence to try my hand at anything. I've never felt far from home even when I was 2000 miles away from it.

My

intermina

not perm

contribut

you know

those mos

barring s

amusing ex

First

immediate

with me fo

input and

were amazi

made the d

(I'm still

Roy Pa

senior grad

genetics.

California

though he k

Paulson be

mentors in

know about

Milligan, be

taught me ab

ACKNOWLEDGEMENTS

My career as well as this thesis are brought to you by an interminable list of individuals, and space and general interest do not permit me to ennumerate all of them or their various contributions. So for any who are left out of the printed set here, you know who you are and so do I. What follows is a brief list of those most prominent in making the whole experience possible and, barring some notable exceptions, a tremendously worthwhile and amusing experience.

First, I thank Christine, who despite experience with my immediate predecessor, took me on as a student and rode out the storm with me for seven years. Her patience (yes, that does go first), input and forbearance from the four deadly words ("I told you so") were amazing, and her support during the aforementioned "dark times" made the difference between this thesis and a job cooking burgers (I'm still not sure if I should thank her for that...).

Roy Parker taught me everything I know about being a jerk and a senior graduate student (redundancy?), and everything he knows about genetics. Dan Frank introduced me to the carbonated version of fine California white wine and was a consummate comrade-in-arms, even though he knows I'm better looking than he is. John Chant and Bob Paulson belong in the same sentence together as my housemates and mentors in the arts manly. Paul Siliciano taught me everything I know about impressing chicks by the naming of engine parts. John Milligan, besides being the most easily suckered post-doc I ever know taught me about the fall from Sr. Grad. Stud. to junior postdoc. Evi

Strauss
I just
neither
taught
Michelle
establis
moms-awa
wasn't u
advice,
the main
this tom
vegetable
Wrischnik
Joe Cout
female k
individua
dance on

The
quarters
made me
re-live
students
I did. M
their wil
its exis
alternate

Strauss taught me everything I know about being a sensitive male, but I just can't seem to remember it.... Sean Burgess taught me that neither the 60s nor naked horseback riding are dead. Cammie Lesser taught me about the role of suicide pills in a liberal education. Michelle Marie Haltiner Jones Julie taught me how to dress, and established Mothers Against Bad Belts. And speaking of moms-away-from-mom, there is of course Lucita Esperas, who, when she wasn't upbraiding my tortoise-like pursuit of women gave me much fine advice, even if it was in Pilipino. Joanna Gilbert was essential to the maintenance my sanity (such as it is) throughout the creation of this tome (in which she played no small part) and made me eat my vegetables, for which she tells me I will one day thank her. Lisa Wrischnik almost got me a Helen Hay, and was a fine boat chaufette. Joe Couto never actually coached me through an encounter of the female kind but would have. Joe Tamm was about the mellowest individual I've ever met who wasn't asleep, even if he did belly dance on occassion.

The Guthrie Lab as an entity has been a home for me for three quarters of a decade, and the people in it, named and otherwise, have made me one of the few people I know who, without hesitation, would re-live their graduate career. I thank a whole host of rotation students for making me feel smart and for doing more of my work than I did. My thesis committee bailed me out of several tight spots, and their willingness to view this work on short notice was essential to its existence. The people of the 9th floor West here have alternately tolerated my enthusiasms and enlivened my stay here

immeasurably. For a bunch of people I had to go to work with, it sure seemed like we were having fun....

Genetic Analysis of the U5 snRNA and the 3' Splice Site in Yeast

Bruce Patterson

ABSTRACT

We have undertaken an analysis of the role of the U5 small nuclear ribonucleoprotein (snRNP) in pre-mRNA splicing in yeast. We began by showing that the gene encoding the U5 snRNA (SNR7) is essential for viability. By depleting cells for the U5 snRNA, we demonstrated that it is required for splicing in vivo. In the absence of the U5 snRNA, transcripts for the spliced genes we examined accumulate primarily as unspliced precursor. However, the actin pre-mRNA, and to a lesser extent, the CYH2 transcript undergo the first step of the splicing reaction. These results, combined with results from mammalian splicing systems, led us to focus on possible roles of the U5 snRNP in the second step of splicing.

In order to further characterize the U5 snRNA we constructed several point mutations in it, some of which confer a ts growth defect on the cell. Our inability to detect a splicing defect in these cells led us to employ a genetic screen to detect factors related in function to the U5 snRNA. We have shown that several of these gene products are involved in pre-mRNA splicing.

We have also investigated the role of the polypyrimidine stretch in yeast splicing. We have shown that this element is involved in efficient 3' splice site utilization. By establishing a 3' splice site competition assay, we have set the stage for a further analysis of the factors, including the U5 snRNP, involved in 3' splice site selection and utilization in yeast.

INTRODUC

CHAPTER

CHAPTER I

CHAPTER I

CHAPTER I

CHAPTER V

CHAPTER VI

EPILOGUE..

TABLE OF CONTENTS

PAGE

INTRODUCTION..1

CHAPTER I: An Essential Yeast snRNA with a U5-like Domain is Required for Splicing in vivo............9

CHAPTER II: A Novel Genetic Screen to Identify U5-Related Splicing Factos....................................59

CHAPTER III: Architecture of Fungal Introns: Implications for Spliceosome Assembly..........................83

CHAPTER IV: A Role for the Polyuridine Stretch in Yeast Splicing is Revealed by Employing a Branchsite Mutation..118

CHAPTER V: 3' Splice Site Requirements in Yeast.............128

CHAPTER VI: Spliceosomal snRNAs............................168

EPILOGUE...221

CHAPTER

Table I

CHAPTER

Table I

LIST OF TABLES

PAGE

CHAPTER III

Table I Possible Helical Structures in *Saccharomyces cerevisiae* Introns....105

CHAPTER IV

Table I β-galactosidase Assays of Strains Containing actin-*lac*Z Fusions with 3' Splice Site Alterations....122

CHAPTER

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

CHAPTER I

Figure 1 S

Figure 2 C

Figure 3 S

Figure 4 S

CHAPTER II

Figure 1 B

Figure 2 A

Figure 3 A

Figure 4 P

CHAPTER IV

Figure 1 A

Figure 2 S

LIST OF FIGURES

PAGE

CHAPTER I

Figure 1 DNA and RNA Sequence of snR7......................14

Figure 2 Mapping the ends of snR7 by S1 Analysis...........17

Figure 3 Establishing Conditional Synthesis of snR7........21

Figure 4 Analysis of Transcripts Produced by (pGAL1)-SNR7..25

Figure 5 Analysis of in vivo splicing phenotypes of (pGAL1)-SNR7....................................29

Figure 6 Northern Analysis of Actin Splicing in Cells Carrying (pGAL1)-SNR7............................33

Figure 7 Comparison of snR7 with U5 snRNAs.................37

Figure 8 Northern Analysis of RNAS in other Fungi which Contain Homologies to the snR7-U5 Loop Domain....40

CHAPTER II

Figure 1 Strategy for Identification of slu^{ts} alleles......64

Figure 2 Growth Phenotypes of slu^{ts}........................69

Figure 3 Splicing Phenotypes of slu^{ts} alleles..............73

Figure 4 Splicing Phenotypes of slu1-1....................75

CHAPTER III

Figure 1 Branchpoint to AG Spacings in Fungal Introns......87

Figure 2 Analysis of Conserved Sequences in 3' Splice Sites in 3'L Introns..................................90

Figure 3 Analysis of 5' Junction to Branchpoint Spacings in Fungal Introns......................101

Figure 4 Possible Helical Structures in *Saccharomyces cervisiae* Introns...............................103

CHAPTER IV

Figure 1 Alterations at the 3' Splice Site of the C259-lacZ Fusion...............................120

Figure 2 Splicing Phenotypes of actin-lacZ fusions with

3' Splice Site Alterations........................125

CHAPTER V

Figure 1 Alterations in the 3' Splice Site of the Yeast Actin Gene..136

Figure 2 Splicing of the +0 Series.........................141

Figure 3 Splicing of the +A and +T Series..................143

Figure 4 Splicing of the +AA and +TT Series................146

CHAPTER VI

Figure 1 U4 Alignment......................................173

Figure 2 U6 Alignment......................................177

Figure 3 U5 Alignment......................................183

Figure 4 U1 Alignment......................................188

Figure 5 U2 Alignment......................................193

Figure 6 Consensus Structures of the U snRNAs..............202

CHAPTER

REQUIRED

SUMMARY

Yeas
for viabi
RNAs have
metazoan
these snR
biochemica
the condi
sequences
SNR7 gene
marked acc
intron-cont
also accumu
snR7 with l
to the mamm
requires the

CHAPTER I: AN ESSENTIAL YEAST snRNA WITH A U5-LIKE DOMAIN IS REQUIRED FOR SPLICING IN VIVO

SUMMARY

Yeast contains at least 24 snRNAs, many of which are dispensable for viability. We recently demonstrated that a small subset of these RNAs have a functional binding site for the Sm antigen, a hallmark of metazoan snRNAs involved in mRNA processing. Here we show that one of these snRNAs, snR7, is required for growth. To determine the biochemical basis of lethality in cells lacking snR7, we engineered the conditional synthesis of snR7 by fusing the snRNA coding sequences to the yeast GAL1 control region. Cells depleted for the SNR7 gene product by growth on glucose for five generations show marked accumulation of unspliced mRNA precursors from the four intron-containing genes tested. In some cases, intron-exon 2 lariats also accumulate. We have identified a ∩70 nucleotide domain within snR7 with limited sequence-specific but striking structural homology to the mammalian snRNA U5. We conclude that in yeast mRNA splicing requires the function of a U5-like snRNA.

INTRODUCTION

Nuclear mRNA splicing is mediated by a group of snRNAs which associate with the Sm antigen (see Maniatis and Reed (1987) for a recent review). These Sm snRNAs comprise the most abundant class of snRNAs in metazoans. Their role in splicing has been investigated primarily by the use of in vitro systems. In particular, oligonucleotide-directed RNase H cleavage has been used to demonstrate specific requirements for U1, U2, U4 and U6 snRNAs (Kramer et al., 1984; Black et al., 1985; Krainer and Maniatis, 1985; Berget and Robberson, 1986; Black and Steitz, 1986). Ribonuclease protection assays have suggested that U1 (Mount et al., 1983) and U2 (Black et al., 1985) snRNPs bind to the sites of 5' cleavage and branch formation respectively. In the case of U1, recognition of the 5' splice site has recently been shown to involve base-pairing (Zhuang and Weiner, 1986); a similar mechanism has been proposed for U2 (Black et al., 1985). No apparent binding site has been identified for the U4/U6 snRNP. Finally, U5 has been implicated in the recognition of the 3' splice site (Chabot et al., 1985; Gerke and Steitz, 1986; Tazi et al., 1986). Due to the resistance of the U5 snRNA to RNase H-mediated cleavage, however, its participation could only be inferred from the similarity in resistance to micrococcal nuclease digestion between the U5 snRNP and a 3' splice site binding activity (Chabot et al., 1985).

Several years ago we began a study of the snRNAs from Saccharomyces cerevisiae to take advantage of the ability to combine

modern genetics with traditional biochemical approaches (Wise et al., 1983). Our primary goal was to identify the snRNAs likely to be involved in splicing so that specific molecular models could be explicitly tested. The use of full-length heterologous clones as hybridization probes failed to reveal candidates for specific snRNA analogues. Most surprisingly, the first three single copy snRNA (SNR) genes we tested proved to be dispensable for growth (Tollervey et al., 1983; Tollervey and Guthrie, 1985; Guthrie et al., 1986), as have three additional SNR genes since then (Parker et al., in preparation). From the foregoing, we concluded 1) that the yeast snRNAs involved in essential roles such as splicing comprise only a subset of the total population and 2) that the homology of S. cerevisiae RNAs to their presumptive metazoan counterparts is relatively limited at the primary sequence level.

Two experiments have suggested that another yeast snRNA, snR7, may play a role in pre-mRNA splicing. We previously sought to identify yeast snRNAs which contain a binding site for the highly conserved Sm antigen, a hallmark of metazoan snRNPs involved in mRNA processing. We have recently demonstrated that, following microinjection into Xenopus eggs or oocytes (which stockpile the Sm antigen in their cytoplasms), at least two snRNAs from Saccharomyces cerevisiae, snR7 and snR14, become immunoprecipitable with anti-Sm antibodies (Riedel et al., 1987). These same snRNAs are found physically associated with the spliceosome, as is a third species, snR20 (Pikielny and Rosbash, 1986; J. Abelson, personal communication; Riedel et al., in preparation).

In this work, we investigate the role of snR7 in vivo. Specifically, we have tested the prediction that SNR7 encodes an snRNA that performs an essential role in mRNA splicing. We first show that snR7 does contain a consensus Sm binding site (see, for example, Branlant et al., 1982), the sequence AUUUUUUG in a potentially single-stranded region. Second, we demonstrate that this gene product is indispensable for growth. Third, in order to determine the molecular basis of the lethal phenotype of null mutations of SNR7, we have made the synthesis of snR7 conditional by fusing the snRNA coding sequences to the GAL1 control region. By shifting from galactose- to glucose-containing media, we show that cells which are substantially depleted of snR7 show a marked inhibition of splicing. Based on the pattern of accumulated intermediates, we suggest that snR7 participates in more than one step in the splicing pathway. Finally, we have identified an internal domain of snR7 (ntes. 59-129) which exhibits limited sequence-specific but striking structural homology to U5.

RESULTS

THE SNR7 LOCUS PRODUCES TWO TRANSCRIPTS

We previously identified snR7 as a capped, nuclear RNA of approximately 220 nucleotides (Wise et al., 1983); it was found to be encoded by a single gene and cloned as described in these initial studies. The gene was localized to a 600 basepair fragment of the genomic clone by Southern hybridization using cDNA made from size fractionated yeast RNA (130-600 ntes.) as a probe (data not shown). The sequence of the 600 basepair ClaI-HindIII fragment containing the SNR7 coding region is shown in Figure 1, along with relevant restriction sites.

We determined that the 600 bp fragment encodes two RNAs, and that a labelled cDNA from one RNA also detected the second species on Northern blots (data not shown), indicating that they were extensively colinear. To determine the 5' and 3' termini of the two transcripts, we performed S1 nuclease protection experiments. As shown schematically in Figure 2D, we used uniformly labelled probes which had one endpoint within the transcribed sequences (the HpaII site, data not shown) and one endpoint extending in either the 5' (probe A) or 3' (probe B) direction. Digestion of probe A annealed to total yeast RNA gives only a single digestion product (Figure 2A), indicating that both transcripts share a common 5' end. This observation has been confirmed by primer extension analysis (see Figure 4B).

FIGURE 1:

The top l
which con
snR7. The
(Sanger e
snR7 infe
found in s
5' and 3'
homology t
"TATA" box
of simila
Restrictio

FIGURE 1: DNA AND RNA SEQUENCE OF snR7.

The top line shows the sequence of 600 basepair ClaI-HindIII fragment which contains all sequences required for wild-type expression of snR7. The DNA sequence was determined by the Sanger dideoxy method (Sanger et al., 1977). The bottom line shows the RNA sequence of snR7 inferred from the DNA sequence. The additional nucleotides found in snR7(L) but not in snR7(S) are indicated in lower case. The 5' and 3' ends of snR7 were determined as described in the text. The homology to the Sm antigen binding site is boxed. The presumptive "TATA" box (identified on the basis of its sequence and the presence of similar sequences 5' of other yeast snRNAs) is underlined. Restriction sites referred to in the text are also indicated.

ClaI

ATCGATGACAA

ATTTGAGTTCG

AAATTTTGTAGA

GTTTCTTTCCTC

10
GCAGCTTTACAG
GCAGCUUUACAG

70
CCTATAGAACTT
CCUAUAGAACUU

130
TCCATAGAAACA
UCCAUAGAAACA

190
TGCCCTTTTTCT
UGCCCUUUUUCU

TTTGATTCATGA

ATTAACTCATTG

```
ClaI
|
ATCGATGACAAAGGATAATGGGTAGAGTCTGGCACTCCTACCCTAAATTGTTAACTTCCT

ATTTGAGTTCGTGGTGTTAGTATTCTCATCACGATTAACGAATATGAAAAAAAAAATTGA

AAATTTTGTAGAAACGGAGTGCTCAGTATAAAAAGCGCATAGTAAGACTTTTTGTTAAAT

                     AhaIII                                         AhaIII               1
GTTTCTTTCCTCCTATATATTTTAAAATACTTTTCTTTCTTTTTGTTTTAAAACCTGCAA
                              HincII                                          m3GpppAA
         10        20        30|       40        50        60
GCAGCTTTACAGATCAATGGCGGAGGGAGGTCAACATCAAGAACTGTGGGCCTTTTATTG
GCAGCUUUACAGAUCAAUGGCGGAGGGAGGUCAACAUCAAGAACUGUGGGCCUUUUAUUG

                                                  HpaII
         70        80        90       100       110|      120
CCTATAGAACTTATAACGAACATGGTTCTTGCCTTTTACCAGAACCATCCGGGTGTTGTC
CCUAUAGAACUUAUAACGAACAUGGUUCUUGCCUUUUACCAGAACCAUCCGGGUGUUGUC

        130       140       150       160       170
TCCATAGAAACAGGTAAAGCTGTCCGTTACTGTGGGCTTGCATATTTTTTGGAACTTTTC
UCCAUAGAAACAGGUAAAGCUGUCCGUUACUGUGGGCUUGCAU[AUUUUUUG]GAACUuuuc

        190       200       210
TGCCCTTTTTCTCAATGAGTAAGGAGGGCGTTTATCTTTTCTATTTTATTTTTCTACAAA
ugcccuuuuucucaaugaguaaggagggcguOH

TTTGATTCATGAGTCCATGGAACAAATATATAGAACTCATTTTCTTGTTTCGGTATATAT
                                                 HindIII
                                                   |
ATTAACTCATTGAAGCATTTGCTTTTATCCTGACTGCATTTTTAGGAAGCTT
```

Use of the probe extending in the 3' direction indicates that the transcripts from the locus have two different 3' endpoints (Fig 2B). The size difference between these two protected fragments (approximately 35 nucleotides) corresponds to the difference in size of the two transcripts identified by Northern hybridization analyses described above. To rule out the possibility that the two transcripts differed by the removal of an intron located at or near the HpaII site, we made an S1 probe which spanned the entire 600 basepair fragment (probe C). As shown in Figure 2C, total RNA protects two fragments of the sizes detected by Northern analysis, indicating that both transcripts are colinear with the DNA.

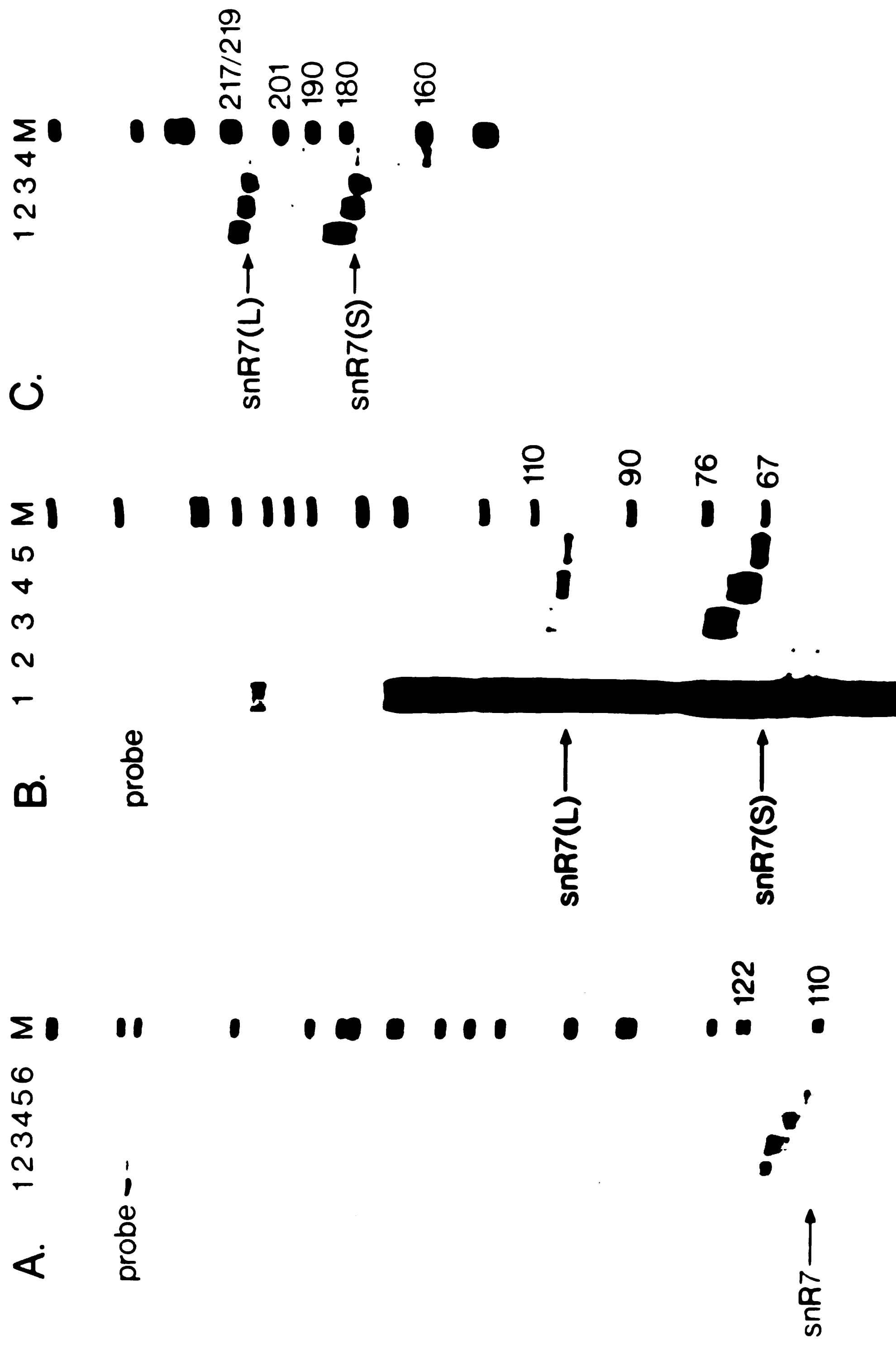
We conclude that the SNR7 locus encodes two transcripts, of ∩214 and ∩179 ntes., which differ only in the length of their 3' termini (see Figure 1). We have designated them snR7 Long (snR7(L)) and Short (snR7(S)), respectively. For convenience, we will use "snR7" to denote both RNAs unless otherwise noted.

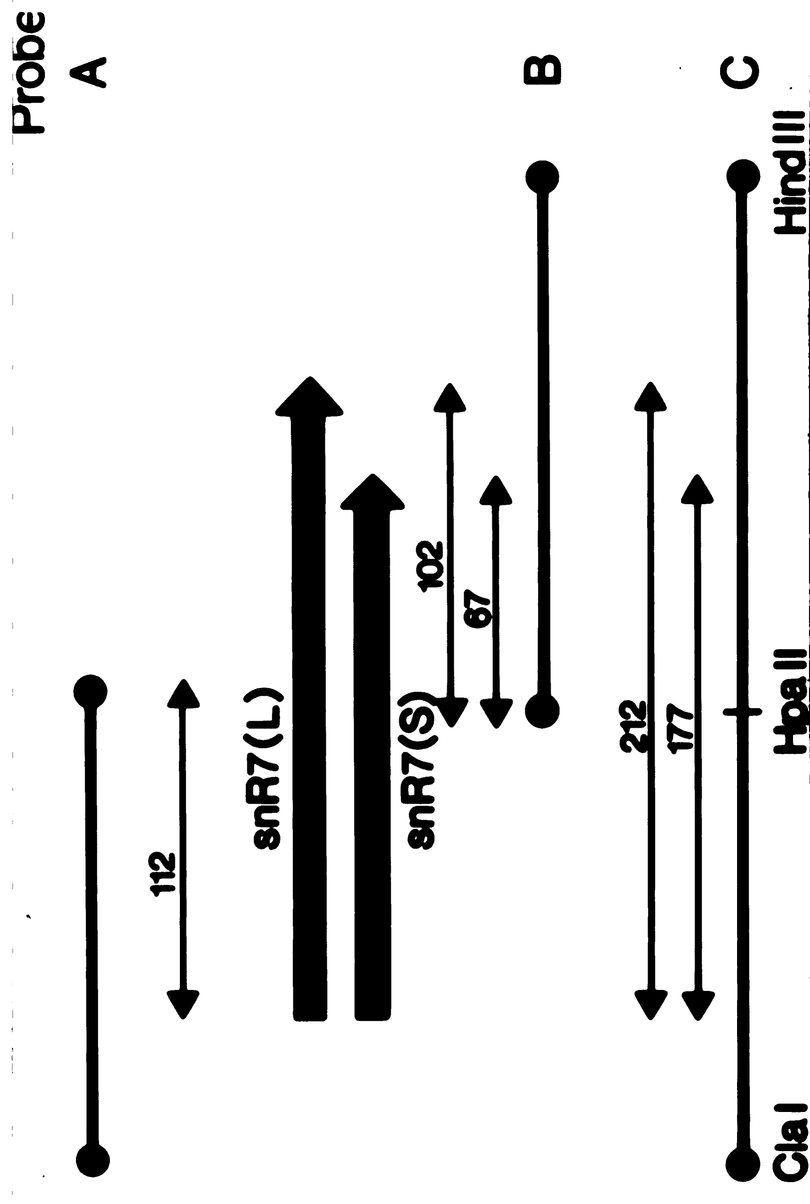
SNR7 IS AN ESSENTIAL GENE

To investigate the role of snR7 we first asked whether the gene (SNR7) that encodes it is essential. We did this by the method of one-step gene replacement (Rothstein, 1983) in diploid cells. As described in Experimental Procedures we created a disrupted allele of SNR7 by introduction of the yeast TRP1 gene into the HincII site of SNR7, creating a 1400 basepair insertion between nucleotides 33 and 34 of the SNR7 coding region. We refer to this allele as snr7::TRP1.

FIGURE 2: MAPPING THE ENDS OF snR7 BY S1 ANALYSIS.

Total RNA from strain (YP52xYP54) was subjected to S1 analysis as described in Experimental Procedures and electrophoresed on a 6% sequencing gel. The probes used were uniformly labelled single-stranded DNA synthesized as described in Experimental Procedures. The probes anneal to the coding strand from the ClaI site to the HpaII site (probe A, Figure 2A), the HpaII site to the HindIII site (probe B, Figure 2B) or the Cla I site to the HindIII site (probe C, Figure 2C). The probes, snR7(L) and snR7(S) and the observed protected fragments are shown schematically in D. Probe A and Probe B overlap by several nucleotides, hence the sum of their sizes are greater than the actual length of the molecule. Sizes shown are our best estimates based on gel mobilities, and are therefore not precise. Units of S1 added were as follows: for Fig 2A and 2B, lane1=2.5, lane 2 = 7.5, lane 3 = 22, lane 4 = 67, lane 5 = 200 and lane 6 = 600. for Fig 2C, lane 1 = 5 units, lane 2 = 28, lane 3 = 140, and lane 4 = 700.


A.
probe
1 2 3 4 5 6 M
122
110
snR7
B.
probe
1 2 3 4 5 M
110
90
76
67
snR7(L)
snR7(S)
C.
1 2 3 4 M
217/219
201
190
180
160
snR7(L)
snR7(S)


Probe
A
B
C
112
snR7(L)
snR7(S)
102
67
212
177
Cla I
Hpa II
Hind III

We performed the gene replacement in a diploid cell and by Sourthern analysis, confirmed the replacement of one chromosomal copy of the SNR7 locus with the disrupted gene. When this heterozygous diploid was sporulated and dissected, we observed only two viable spores per tetrad. In all cases, these spores were Trp- , indicating that the disrupted copy of the SNR7 gene was unable to support growth. Microscopic examination of the inviable spores revealed that they had undergone three to four cell divisions prior to cessation of growth. This indicates that SNR7 is not required for germination and that spores unable to synthesize snR7 successfully complete several cell cycles. These cells did not exhibit a uniform terminal phenotype.

To prove that the lethal phenotype was the specific consequence of the absence of snR7, we asked whether viability could be restored by providing snR7 in trans. We constructed a HIS3-centromere plasmid carrying the 600 basepair ClaI-HindIII fragment shown in Figure 1. This plasmid (pBWHis600) was transformed into diploids containing a single copy of the disrupted gene, snr7::TRP1, and transformants were sporulated and dissected. Trp+, His+ haploid cells were recovered (data not shown), indicating that cells containing the chromosomal disruption of SNR7 were viable in the presence of the complementing plasmid.

ESTABLISHING CONDITIONAL SYNTHESIS OF snR7

To determine the biochemical basis of lethality due to the absence of snR7, we sought a conditional SNR7 allele. The strategy we chose was to put the synthesis of snR7 under the control of a

repressible/inducible promoter. Our rationale was that by shutting down the synthesis of snR7 in a controlled fashion, we could examine the effect of depletion of the snRNA on RNA processing. If snR7 is an essential splicing factor, the depletion of the snRNA by dilution should result in an increase in precursors or intermediates in the splicing pathway.

Construction of a GAL1-SNR7 Gene Fusion

We chose the yeast GAL1 promoter because it has been extensively characterized (Johnston and Davis, 1984) and has been used successfully for other applications of this type (see, for example, Fried et al., 1985). As shown in Figure 3A, we fused the GAL1 regulatory sequences up to but not including the GAL1 RNA start site to snR7-encoding sequences beginning just 5' of the snR7 RNA start site. We expected that the resultant fusion, designated (pGAL1)-SNR7, would allow the synthesis of snR7 in the presence of galactose and its repression in the presence of glucose.

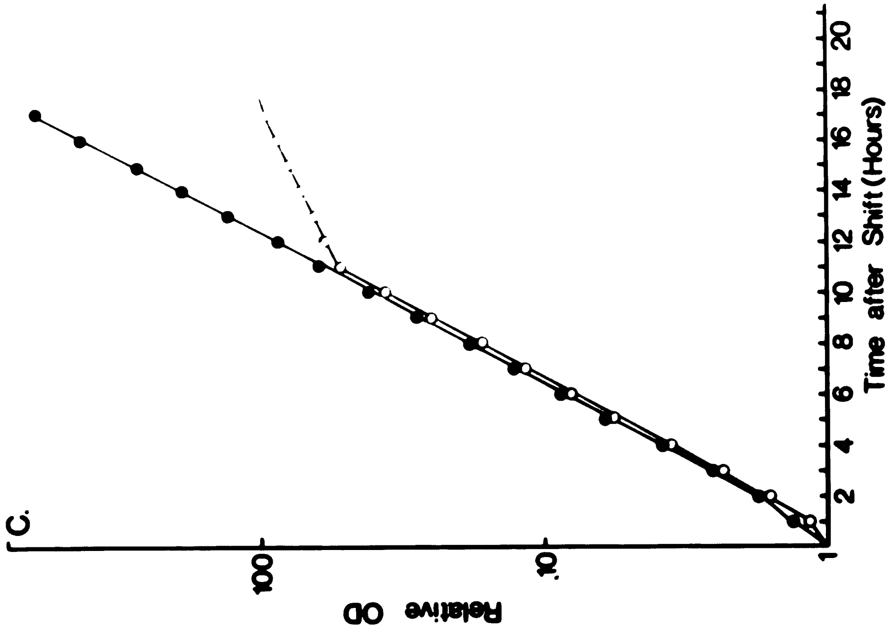
The promoter substitution was inserted into a HIS3-centromere vector (pBWHis3) to create pBWHisG7 (Figure 3B). This plasmid was transformed into heterozygous SNR7/snr7::TRP1 diploids, and His+ transformants were sporulated and dissected on galactose-containing medium. Tetrads with three and four viable spores were recovered after sporulation, indicating that the fusion construction could provide snR7 activity in trans. Analysis of nutritional markers demonstrated that cells containing snr7::TRP1 (Trp+) were recovered

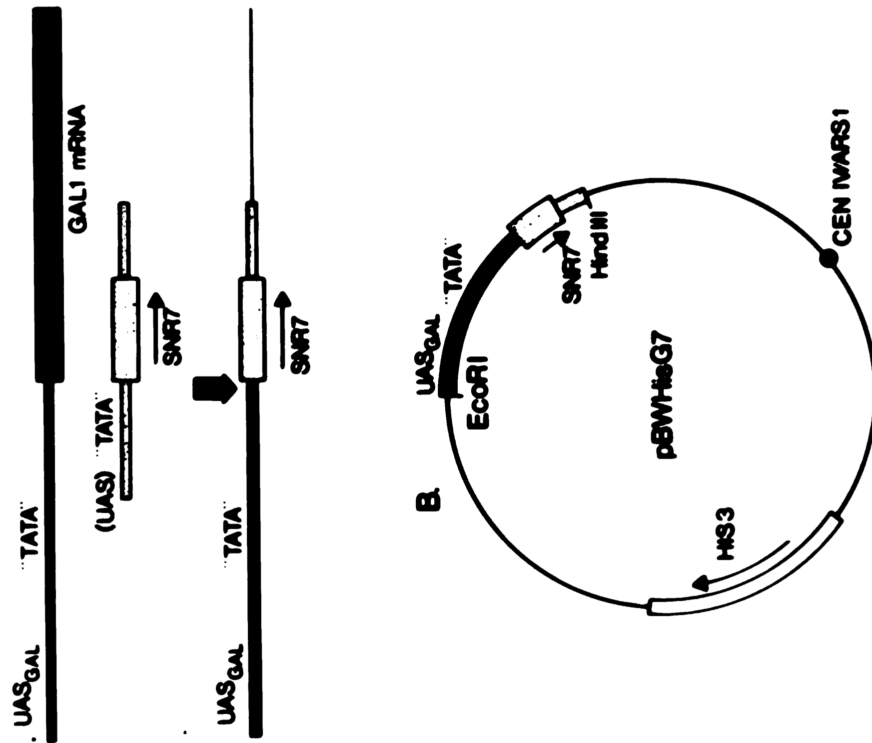
FIGURE 3: ESTABLISHING CONDITIONAL SYNTHESIS OF snR7.

A) Strategy: The general strategy of the construction of a repressible allele of SNR7, (pGAL1)-SNR7, is shown. The UAS and untranscribed regions of the yeast GAL1 gene, up to but not including the GAL1 RNA start site, were fused to sequences just upstream of snR7-encoding sequences. For complete details of this construction, see Experimental Procedures.

B) pBWHisG7: In order to introduce the (pGAL1)-SNR7 allele into yeast, we cloned the fragment shown in Figure 3A into a yeast His3 centromere vector (see Experimental Procedures) resulting in the plasmid depicted.

C) Growth Curve: To determine the kinetics of the effect of snR7 depletion on growth, we followed logarithmically growing cells containing the (pGAL1)-SNR7 allele after shifting them from galactose- to glucose-containing media. Cell doublings from the time of shift (time 0) are shown (cells were diluted in order to maintain logarithmic growth). Isogenic wild-type cells were also shifted and followed similarly. The arrow marks the point at which cells were taken to prepare RNA (see text).


C.
100
10
1
Relative OD
2 4 6 8 10 12 14 16 18 20
Time after Shift (Hours)


UAS$_{GAL}$
TATA
GAL1 mRNA
(UAS)
TATA
SNR7
UAS$_{GAL}$
TATA
SNR7
B.
UAS$_{GAL}$
TATA
EcoRI
pBWHisG7
SNR7
HindIII
CEN IV/ARS1
HIS3

only in the presence of the plasmid bearing (pGAL1)-SNR7 (His+). Moreover, haploids whose only functional copy of SNR7 was (pGAL1)-SNR7 (i.e., His+, Trp+) were inviable on glucose, suggesting that we had successfully established conditional synthesis of snR7.

We monitored the growth rate in liquid culture of cells containing only (pGAL1)-SNR7 to determine the effects of depletion of snR7 on cell doubling times. Cultures were grown on galactose to mid-log phase, shifted to glucose, and their subsequent growth followed by monitoring A600 of the cultures. A representative growth curve is shown in Figure 3C. It should be noted that cells in which the promoter substitution is the only source of snR7 have doubling times in galactose indistinguishable from those of otherwise isogenic cells. This suggests that snR7 produced using the GAL1 promoter is functional and supports wild-type growth rates; we will return to this point later (see Discussion). Following the shift to glucose, we reproducibly observe a break in logarithmic growth between 11 and 12 hours, corresponding to 5.5 to 6 generations of growth in glucose. In contrast, wild-type cells continue to grow logarithmically for at least 24 hours after the shift.

Conditional Expression of (pGAL1)-SNR7

We next examined the actual depletion of snR7 in cells containing the glucose repressible allele of SNR7, (pGAL1)-SNR7, on glucose media. RNA was extracted from cells grown continually on galactose and from cells which had been grown on glucose for 10 hours (i.e., 5 generations after the shift). Equal amounts of RNA from

wild-type and fusion-containing cells were examined by Northern analysis. We used an end-labelled oligonucleotide complementary to nucleotides 10 to 28 of snR7 as a probe. An end-labelled oligonucleotide hybridizing to snR17 (Riedel et al., 1986) was also added to provide an internal control (Figure 4A, band "C")

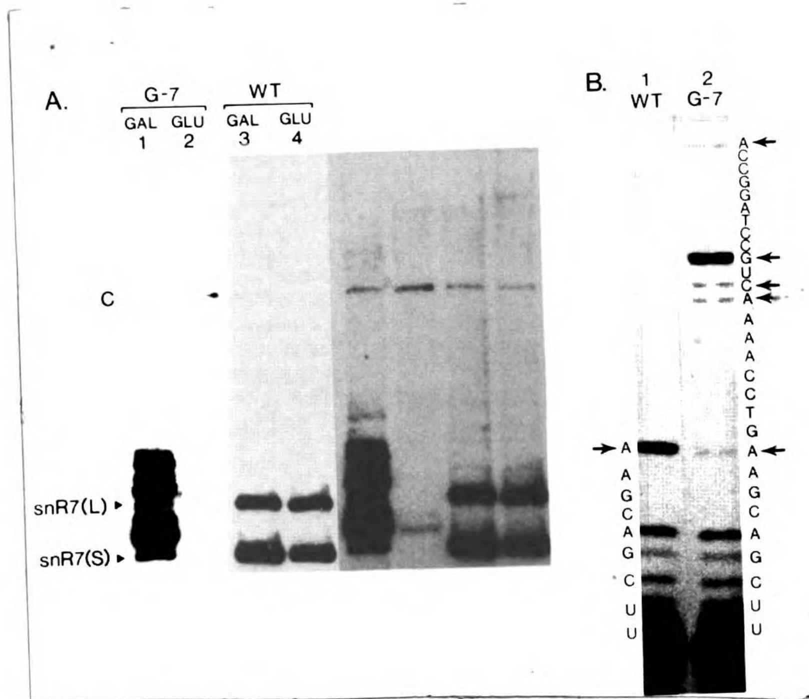
The results of this experiment are shown in Figure 4A. Lanes 3 and 4 contain RNA from control cells grown on galactose and glucose, respectively. Note the two essentially equimolar bands corresponding to snR7(L) and snR7(S), described earlier. The level of these transcripts, relative to the internal control, shows no apparent variation in response to carbon source. Lane 1 contains RNA produced from cells carrying (pGAL1)-SNR7 grown on galactose. When this lane is compared to its control (lane 3, wild-type cells grown on galactose), several important differences can be seen. 1) There is a net increase in transcripts with SNR7-derived sequences in cells containing (pGAL1)-SNR7; by densitometry we estimate this increase to be 2.5 fold. 2) The majority of transcripts in fusion-containing cells are not identical to wild-type snR7(L) or snR7(S) in size. Rather, these molecules appear to comprise two families of related species which are larger than snR7(S) and snR7(L), respectively. 3) Transcripts of the appropriate size for wild-type snR7(S) and snR7(L) are detectable, but they are present at significantly lower levels than wild-type.

The effect of shifting cells containing (pGAL1)-SNR7 to glucose for 10 hours is shown in lane 2 (Figure 4A). There is a dramatic decrease in the amount of snR7-based transcripts relative to the internal standard; we estimate this to be on the order of 500-fold.

FIGURE 4: ANALYSIS OF TRANSCRIPTS PRODUCED BY (pGAL1)-SNR7.

A) Northern analysis: The products of the (pGAL1)-SNR7 allele were assayed by Northern analysis using a primer annealing to nucleotides 10 to 28 of wild-type snR7. RNA was made from cells containing only the fusion (lanes 1 and 2) or from isogenic wild-type cells (lanes 3 and 4). Cells were either grown continuously on galactose (Lanes 1 & 3) or shifted to glucose medium for 10 hours (lanes 2 & 4) prior to RNA extraction. The positions of the wild-type snR7 bands are indicated, as is the position snR17 provided as an internal control (C).(the oligo to snR17 was synthesized from sequence information kindly provided by J. Hughes and G. Cesareni (pers. comm.)).

B) Mapping 5' ends of transcripts produced by (pGAL1)-SNR7: To determine the precise 5' end of the aberrant transcripts, an end-labelled oligonucleotide annealing to nucleotides 10 to 28 of wild-type snR7 was used for primer extension analysis of total RNA from (pGAL1)-SNR7 containing and wild-type cells grown on galactose- or glucose-containing media. cDNAs were analyzed on a 10% sequencing gel. Lane 1 represents primer-extended RNA from wild-type cells grown on galactose medium, while lane 2 represents (pGAL1)-SNR7 cells grown similarly. Primer extension products just larger than the primer do not represent real 5' ends in that we do not detect them with a primer situated within the middle of snR7 (data not shown).


A.
G-7
WT
GAL
GLU
GAL
GLU
1
2
3
4
C
snR7(L)
snR7(S)
B.
1
WT
2
G-7
A
C
C
G
G
A
T
C
C
G
U
C
A
A
A
A
C
C
T
G
A
A
G
C
A
G
C
U
U
A
A
G
C
A
G
C
U
U

The only clearly visible signal corresponds to the predominant band among the multiple novel species produced during growth on galactose (lane 1). Thus, while many aberrantly sized transcripts are generated from (pGAL1)-SNR7, they all appear to be under the control of the GAL1 promoter.

Structure of Transcripts from (pGAL1)-SNR7

Because we had substituted GAL1 control sequences for those normally found in SNR7, we thought that the aberrant transcripts generated by this construction probably arose from heterogeneous transcription initiation, which appears to be characteristic of non-snRNA polymerase II transcripts in yeast (see, for example, Guarente and Mason (1983)). To distinguish this possibility from a model in which the size differences in the snR7 transcripts produced from the promoter substituted gene are due to variation in 3' ends, we mapped the 5' ends of these molecules. We performed primer extension analysis of RNA from wild-type and fusion-containing cells, using an oligonucleotide primer which anneals to nucleotides 10-28 of snR7. The results of this experiment are shown in Figure 4B. Consistent with the data from S1 nuclease protection experiments described earlier (see Figure 2), snR7 transcripts normally initiate at a single site (Figure 4B, lane 1). As we anticipated, snR7 transcripts made using the GAL1 promoter initiate at the wild-type site as well as at sites located 9, 10, 12, 13 and 21 nucleotides upstream of this position. The predominant product initiates at position -12. These observations are consistent with the pattern of

transcripts revealed by Northern hybridization analysis (Figure 4A) and argue that the size heterogeneity can be accounted for by differences in the 5' termini. In addition, we demonstrated by immunoprecipitation with anti-trimethylguanosine antibodies (cf. Riedel et al., 1986) that all of the 5'-extended species are trimethylguanosine capped (data not shown).

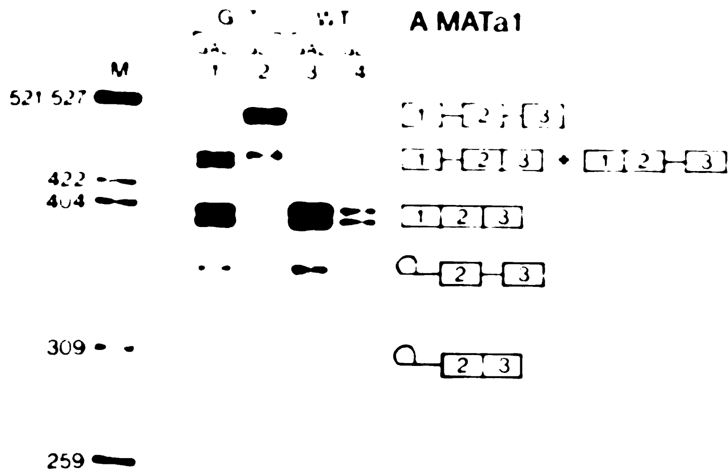
IN VIVO SPLICING REQUIREMENT FOR snR7

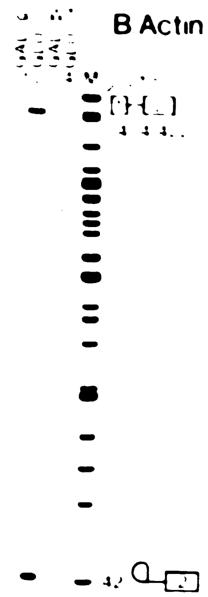
The preceding experiments established a means for exploring the role of snR7 in vivo. By analysing RNA extracted from cells dependent on the (pGAL1)-SNR7 promoter substitution as the sole source of snR7, we can directly compare RNA processing pathways, in particular pre-mRNA splicing, in the presence and absence of this essential snRNA. Our strategy was to monitor, by primer extension and Northern analyses, transcripts from a variety of spliced genes. As described in a previous section, we extracted RNA from wild-type cells and from those carrying (pGAL1)-SNR7; cells were grown continuously on galactose (Gal), or shifted to glucose for 10 hours prior to RNA extraction (Glu). We chose this time point because we knew that the pool of snR7 would be severely reduced (see Figures 4A and 4B), yet cells could continue to grow at a normal rate for another generation (see Figure 3C). In this way we hoped to minimize the impacts on cellular processes anticipated after the break from logarithmic growth.

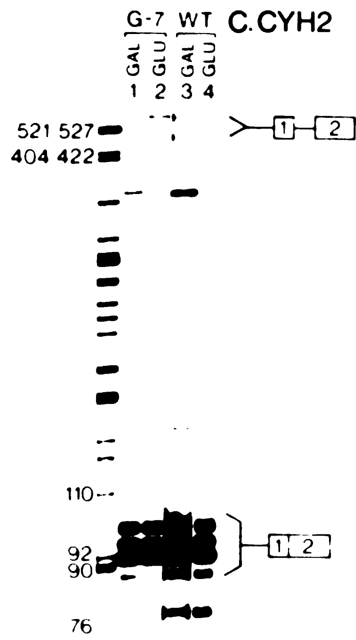
Figure 5A shows the analysis of transcripts from the MATa1 gene, which contains two introns (of roughly equal size). The oligonucleotide primer anneals to sequences in the third exon,

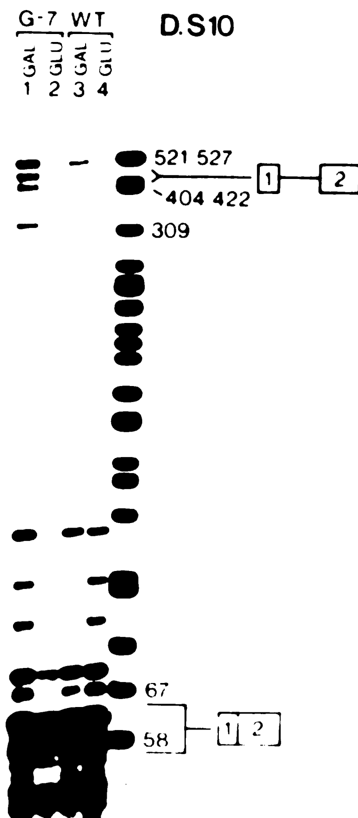
FIGURE 5: ANALYSIS OF IN VIVO SPLICING PHENOTYPES OF (pGAL1)-SNR7.

Primer extension analysis was used to assay the accumulation of mRNA splicing precursors and intermediates in cells containing the (pGAL1)-SNR7 allele grown either continuously on galactose media (lane 1) or shifted to glucose media for 10 hours (lanes 2). As controls, isogenic wild-type cells were grown under identical conditions (lanes 3 & 4, respectively). RNA was extracted from cells and poly-A+ enriched (see Experimental Procedures). 5 ug of pA+ RNA was then annealed and extended with a primers specific for MATa1 (Figure 5A), actin (Figure 5B), CYH2 (Figure 5C) or S10 (Figure 5D) RNAs. The CYH2 and S10 primers anneal to the 5' region of exon II of their respective genes. The MATa1 primer anneals to the 5' region of exon III, and the actin primer is intron-specific, annealing between the branchpoint and AG dinucleotide. The structure and expected positions of the RNA species detected by each of these primers are shown beside each gel. Markers are end-labelled pBR325 cut with HpaII. Unmarked primer extension products do not correspond in size to any expected intermediates, and are therefore most probably the result of reverse transcriptase stalling.


A MATa1
M
521 527
422
404
309
259


B Actin


G-7
WT
C.CYH2
GAL
GLU
GAL
GLU
1 2 3 4
521 527
404 422
110
92
90
76


G-7
WT
D.S10
GAL
GLU
GAL
GLU
1 2 3 4
521 527
404 422
309
67
58

allowing the detection of all splicing intermediates and products. As shown, the predominant extension product in wild-type cells (lane 3 (Gal), lane 4 (Glu)) is generated from mature MATa1 mRNA. Although pre-mRNA containing one of the two introns can be detected, the ratio of these primer extension products to one another shows no variation in response to carbon source.

Interestingly, (pGAL1)-SNR7 cells grown continually on galactose (Figure 5A, lane 1) contain noticably greater amounts of pre-mRNA retaining both introns, as well as partially spliced precursor, than do similarly grown wild-type cells. While this suggests that the presence of fusion-derived snR7 confers a partial splicing block, the predominant form of the MATa1 transcript in these cells is, nonetheless, fully matured mRNA. In contrast, depletion of the snRNA by shift to glucose (Figure 4A, lane 2) results in a dramatic accumulation of unspliced pre-mRNA, and no fully spliced mRNA is visible. We conclude that the splicing of MATa1 precursor is completely dependent upon the presence of snR7.

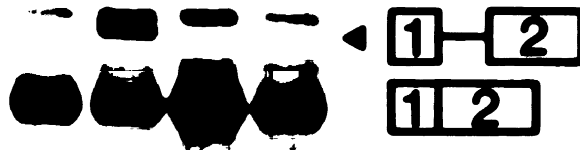
Figure 5B shows a similar analysis of actin mRNA splicing. In this case, however, the oligonucleotide primer anneals to the intron, and therefore detects only precursor and lariat-containing species. The splicing of actin pre-mRNA is normally very efficient, since we reproducibly do not detect precursor or lariat intermediate in wild-type cells (Figure 5B, lanes 3 (Gal), 4 (Glu)) by this sensitive technique. In contrast, (pGAL1)-SNR7 cells shifted to glucose (lane 2) contain significant amounts of pre-mRNA and lariat intermediate (we know that the lariat-containing species includes the lariat intermediate because we also detect it using an actin exon 2 primer

(data not shown)) . As we observed with MATa1, the presence of the fusion in cells grown continually on galactose is also associated with a slight increase in precursor (which is visible only upon lengthy exposure of the autoradiogram). To confirm each of these observations, these same RNAs were subjected to Northern analysis using an end-labelled oligonucleotide probe specific for the second exon of actin; in addition, we included a probe for URA3 mRNA as an internal control. Figure 6 shows that, as expected, wild-type cells (lanes 3,4) contain only mature actin mRNA, while significant amounts of precursor are accumulated in cells depleted of snR7 by carbon source shift (lane 2). Again, a faint band of precursor is visible on suitably exposed autoradiograms in RNA derived from fusion-containing cells grown on galactose (lane 1). From these results we conclude that efficient splicing of actin pre-mRNA also requires snR7, although the observed effects of depletion by carbon source shift are not as quantitative as those observed with MATa1.

Similar conclusions can be drawn for the splicing of two ribosomal protein transcripts derived from intron-containing genes, CYH2 and S10 (Figures 5C, 5D). Once again, primer extension analysis of RNA from cells depleted of snR7 (lane 2) reveals a significant increase in the level of unspliced pre-mRNA compared to wild-type (lanes 3, 4). As with MATa1 and actin pre-mRNAs, the products of the promoter substitution appear to confer a partial inhibition on the splicing of S10, since (pGAL1)-SNR7 cells grown continually on galactose accumulate unspliced S10 precursor (visible on darker exposures). Because the splicing of pre-mRNA from CYH2 is relatively inefficient even in wild-type cells (Figure 5C, lanes 3, 4), however,

FIGURE 6 NORTHERN ANALYSIS OF ACTIN SPLICING IN CELLS CARRYING (pGAL1)-SNR7.

To confirm the results seen by primer extension, the same polyA+ RNA as that used in Figure 5 was glyoxylated and run on a 1.5% agarose gel. The gel was blotted to HyBond and probed with an oligonucleotide annealing to exon II of actin. An oligonucleotide annealing to the URA3 mRNA was included as an internal control. The actin RNA products are as indicated. The probe also anneals non-specifically to contaminating 18S and 23S rRNAs as indicated.


G-7
WT
GAL
GLU
GAL
GLU
1
2
3
4
23S
18S
1
2
1
2
URA3

it is difficult to evaluate whether the 5'-extended snR7 transcript makes an additional contribution to the accumulation of precursor in the case of this splicing pathway (Figure 5C, lane 1). Finally, the depletion of snR7 by carbon source shift is also associated with the accumulation of an extension product corresponding to the CYH2 lariat intermediate (visible only in longer exposures). We do not detect a lariat intermediate of S10.

snR7 CONTAINS A DOMAIN WITH HOMOLOGY TO U5

The relationship of snR7 to previously characterized metazoan snRNAs is of interest for a number of reasons. In particular, the identification of an analogue should provide the opportunity to compare our in vivo results with predictions based on data from in vitro experiments in mammalian systems. To date, the only yeast snRNA with known homology to a metazoan U-snRNA is snR20. Ares (1986) found that the first 53 nucleotides of the yeast RNA (LSR1 by his nomenclature) are 81% homologous to mammalian U2, and that structural homology continued over the next 50 residues. Because the yeast RNA extends for an additional kilobase, while U2 from metazoans is less than 200 nucleotides in length, Ares suggested that snR20 might comprise a "poly-snRNP" also encoding U4/U6 and U5 domains. While this is an appealing concept, Ares also pointed out that the regions of sequence homology to U4, U5 and U6 were weak and did not readily fit available structural models.

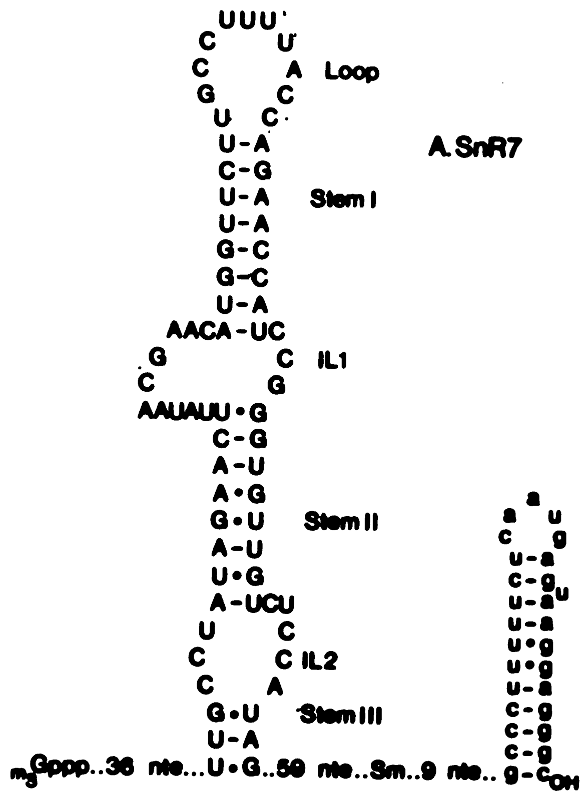
Computer-assisted searches comparing the snR7 sequence to those of known mammalian snRNAs failed to reveal any stretches of homology comparable to that shared by U2 and snR20. Based on the conclusions

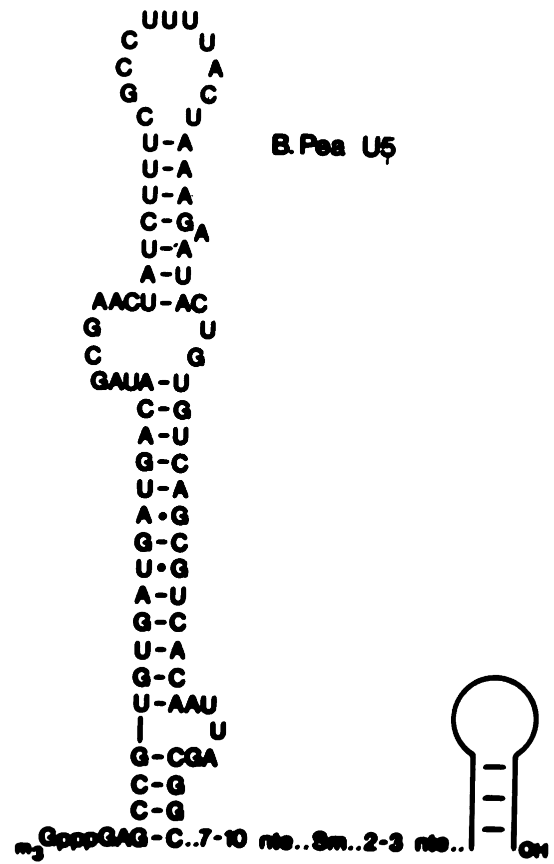
derived from extensive phylogenetic comparisons of ribosomal RNAs (Noller, 1984), we searched for shorter homologies predicted to occur in loops or single-stranded regions of snR7 and the respective U snRNA. The only homology which satisfied this criterion was a 9/9 match between snR7 and U5. Notably, the homology lies within an extraordinarily conserved 11 nucleotide loop in U5; this sequence is identical in pea, tetrahymena, dinoflagellates, chicken, rat, and man (Reddy, 1986)(Figure 7). Moreover, the two nucleotide changes observed for snR7 are at the ends of this loop and are conservative (pyrimidine-pyrimidine transitions). As shown in Figure 7A, when these 11 nucleotides of snR7 are placed in a loop, it is possible to generate a structure which has striking overall similarities to the compound helix which underlies this loop in all sequenced U5 RNAs; the domains from pea (Figure 7B), and rat (Figure 7C) are shown for comparison.

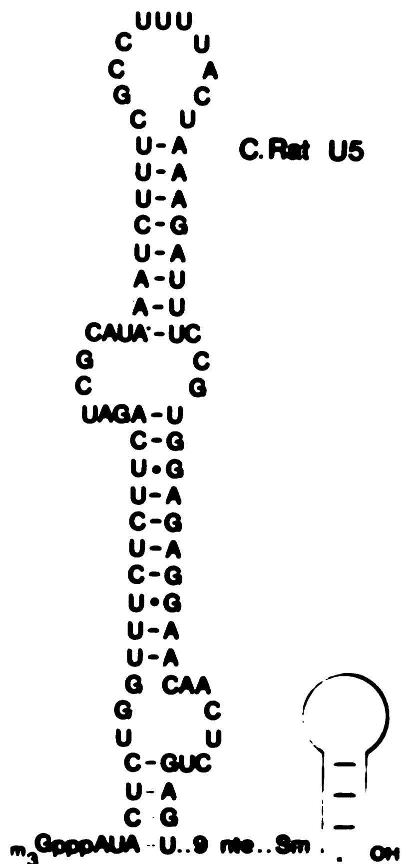
For purposes of discussion, the central U5 domain can be divided into several sub-structures. As shown in the composite (Figure 7D), the conserved loop (Loop) is supported by two major stems (I, II) which are separated by an internal loop (IL1); a second internal loop (IL2) separates Stem II from a third, short stem (Stem III). In snR7, as in all other U5 molecules studied, the Loop is 11 nucleotides and Stem I is comprised of 8 base pairs (except pea, where an A residue is unpaired (Figure 7C)). The sizes of the remaining sub-domains are more variable phylogenetically. The most conserved of these is the uppermost internal loop (IL1). The 5' side of this loop contains 7-10 nucleotides; notably, the snR7 sequence is identical to pea in 7 of 8 positions. Moreover, the 3' side of this loop in snR7 is CCG, as it

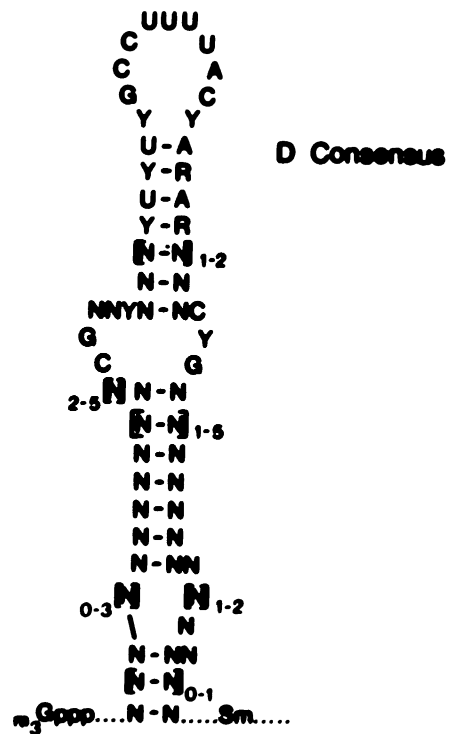
FIGURE 7 COMPARISON OF snR7 WITH U5 snRNAs.

A postulated structure for snR7 is shown (Figure 7A) along with U5 snRNAs from pea (Figure 7B) and rat (Figure 7C). The rat structure is as published (Krol et al., 1981), while the pea structure has been slightly re-drawn from the one published (Krol and Ebel, 1983) in order to maximize structural homology to snR7 and known U5 snRNAs. The locations of the Sm antigen binding sites are also indicated. For convenience, we have termed the structures of the stem as shown. Figure 7D represents a consensus U5 snRNA constructed from the RNAs shown.


Loop
A. SnR7
Stem I
IL1
Stem II
IL2
Stem III


B. Pea U5


C. Rat U5

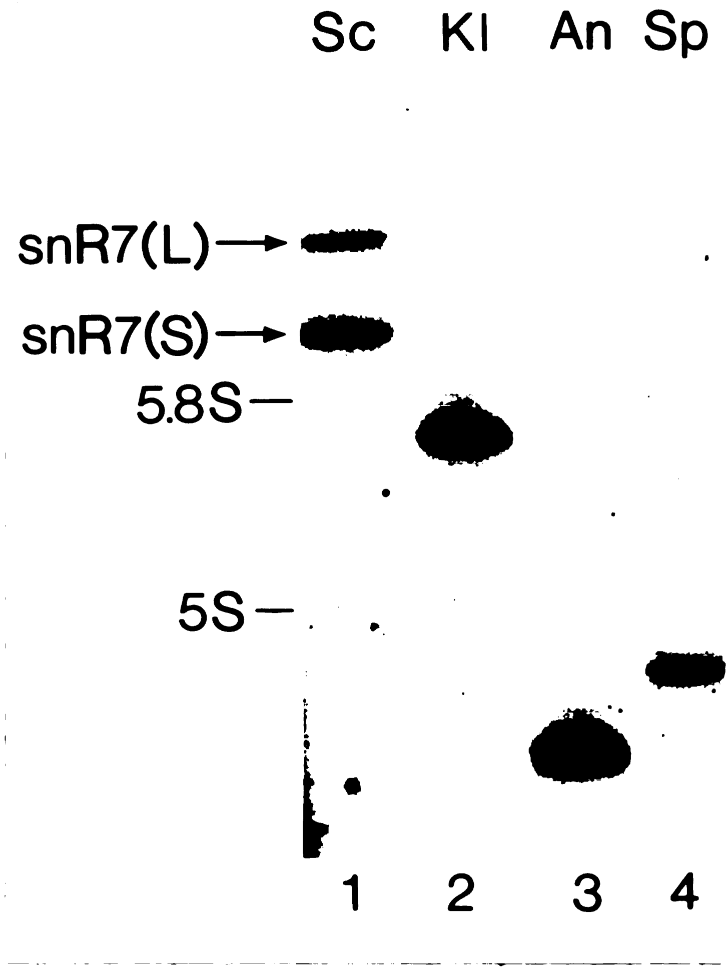

D Consensus

is in all organisms except pea, which has a conservative C to U change.

This striking correlation between the positions of conserved nucleotide sequence and conserved structural features makes a strong case for a phylogenetic relationship between snR7 and U5. As a further test of this hypothesis, we asked whether we could identify possible U5-like RNAs in other fungi using an oligonucleotide complementary to the loop sequence from snR7 as a hybridization probe in Northern analyses. As shown in Figure 8, signals can be detected in RNA from Kluyveromyces lactis, a closely related yeast, as well as from two distantly related fungi, Schizosaccharomyces pombe and Aspergillus nidulans. Under the conditions employed, we would expect a match of at least 11 contiguous nucleotides, suggesting that the pyrimidine transitions at the base of the Loop (cf. Figure 7A with 7B,C) in snR7 may be conserved in other lower eukaryotes.

FIGURE 8 NORTHERN ANALYSIS OF RNAS IN OTHER FUNGI WHICH CONTAIN HOMOLOGIES TO THE snR7-U5 LOOP DOMAIN.

In order to search for RNAs in other fungi which encoded a domain homologous to that found in U5 and which we suggest is present in snR7, we synthesized an oligonucleotide which anneals to the putative loop domain of snR7 (a 15mer which includes 2 nucleotides on both the 5' and 3' sides of the stem shown in Figure 7A). This oligonucleotide was kinased and used to probe a Northern blot containing RNA from S. cerevisiae (Sc), Kluyveromyces lactis (Kl), Aspergillus nidulans (An) and Schizosaccharomyes pombe (Sp). The Northern was washed under conditions which we would predict to demand an 11 contiguous nucleotide match.


Sc
Kl
An
Sp
snR7(L)
snR7(S)
5.8S
5S
1
2
3
4

DISCUSSION

EXPRESSION OF SNR7

The SNR7 Locus Produces Two Related Transcripts

To our knowledge, snR7 is unique among all snRNAs characterized to date in that it exists as two equimolar species of distinctly different size. One species, snR7(L), differs from the other, snR7(S), by 35 additional nucleotides at the 3' terminus. It is interesting to note that these additional residues have the potential to form a stable stem-loop structure, as shown in Figure 7A. We have no data which address the question of whether the two 3' ends are the result of differential termination or processing; nor can we rule out the possibility that snR7(S) comprises a stable degradation product of snR7(L). We are also interested to know whether these transcripts perform identical functions. It seems unlikely that their roles are significantly different, in that both species 1) are capped with trimethylguanosine (Riedel et al., 1986; this work), 2) are associated with the yeast spliceosome (Pikielny and Rosbash, 1986; Riedel et al., in prep.; J. Abelson, pers. comm.), and 3) exhibit virtually identical profiles with respect to sedimentation velocity (Riedel et al., in prep.) and equilibrium density (M. H. Jones and C. Guthrie, in prep.).

snR7 Can Be Synthesized from a Heterologous Promoter

We engineered the conditional synthesis of snR7 by fusing the SNR7 coding sequences to the upstream control region of GAL1 (to create (pGAL1)-SNR7). Haploid cells in which this fusion construct provides the only source of snR7 are viable when grown on galactose.

Synthesis of active snR7 from an mRNA promoter demonstrates several important features of snRNA biogenesis in yeast. Since (pGAL1)-SNR7 generated transcripts are capped with trimethylguanosine we conclude that the signals for cap hypermethylation must lie downstream of the promoter elements, as is the case for higher eukaryotic snRNAs (where a functional Sm antigen binding site has been shown to be both necessary and sufficient for cap hypermethylation)(Mattaj, 1986; de Vegvar et al., 1986; Hernandez and Weiner, 1986). However, in contrast to data obtained in higher eukaryotes, where substitution of a non-snRNA promoter resulted in polyadenylated snRNA transcripts (de Vegvar et al., 1986; Hernandez and Weiner, 1986), promoter substitution does not appear to alter snRNA 3' end generation in yeast: that is, the altered sizes of (pGAL1)-SNR7 transcripts can be completely accounted for by additional nucleotides at the 5' end. It is not clear whether this 5' heterogeneity arises from the introduction of new sequences at the fusion junction or whether it reflects differences between yeast mRNA and snRNA promoters.

snR7 IS A YEAST SPLICING FACTOR

snR7 is Required for Splicing in vivo

To test the prediction that snR7 is required for splicing in vivo, we examined RNA populations in cells depleted of snR7 by carbon source shift. We have shown that cells carrying (pGAL1)-SNR7 are substantially depleted for snR7 five generations following the shift from galactose to glucose. Under these conditions, specific blocks in mRNA splicing have been detected using primer extension and Northern analyses. Our experiments demonstrate that five introns in genes

encoding three different types of proteins (actin, two ribosomal proteins, and the product of the mating type control gene MATa1) require snR7 for efficient splicing. On this basis, we conclude that snR7 is required for the processing of all intron-containing genes in yeast.

For several reasons we also think it is most likely that snR7 is directly involved in the splicing process. First, sequence analysis of snR7 reveals an excellent consensus binding site for the Sm antigen (see, for example, Branlant et al., 1982), the sequence AUUUUUUG in a region likely to be single stranded. Further, we have demonstrated elsewhere that snR7 can assemble with the Sm antigen following microinjection into Xenopus eggs or oocytes (Riedel et al., 1987). As pointed out in the Introduction, the ability to bind the Sm antigen is a hallmark of metazoan snRNAs involved in splicing.

The most compelling evidence that snR7 participates directly in the splicng process comes from the inhibitory effect of elongated snR7 transcripts on splicing. When cells containing (pGAL1)-SNR7 are grown continuously on galactose, we have observed the steady-state accumulation of modest levels of unspliced precursors of certain transcripts; S10 and MATa1 exhibit significant increases compared to wild-type, while actin shows only a marginal effect, and none is discernible for CYH2. In principle, this effect could be due to the decreased concentration of normal length snR7 molecules in these cells. We consider this explanation extremely unlikely, however, because it is inconsistent with the lengthy delay in the onset of growth inhibition and precursor accumulation following carbon source shift. That is, if functional snR7 transcripts were limiting prior to

the shift, they should be rapidly diluted to levels unable to support splicing; in fact, it is not until five or six generations after the shift that increases in unspliced precursors are observed (data not shown) and growth is inhibited (see Figure 3C). Thus we favor the interpretation that the addition of 9-20 nucleotides to the 5' terminus of snR7 is sufficient to perturb the splicing process.

snR7 resembles the metazoan Sm snRNA U5

We have found significant homology between snR7 and the metazoan snRNA U5. This homology is limited at the primary sequence level to a 9 out of 11 nucleotide match between the U5 loop and a putative loop in snR7 (Loop, Figure 7), and several homologies to an internal loop of the U5 compound helix (IL1, Figure 7). Nonetheless, a 70 nte. domain of snR7 can be drawn as a compound helical element that closely resembles structures seen in U5 snRNAs from pea and rat (Figure 7) as well as tetrahymena, dinoflagellates, chicken and man. It is also clear that snR7 must differ in structure from U5 snRNAs identified to date, which contain approximately 45 ntes. outside of the 70 nte. common domain, compared to 110 (snR7(S)) or 145 (snR7(L)) ntes. The "extra" nucleotides in snR7 are found both 5' and 3' of the domain of U5 homology. In this light, the extraordinary size of snR20 (cf. Ares, 1986) might alsobe indicative of generally larger snRNAs in Saccharomyces cerevisiae. Conceivably, this difference reflects a greater reliance on RNA in the construction of the yeast spliceosome than in the mammalian spliceosome.

Despite these potentially significant differences between snR7 and U5 molecules, the phylogenetically conserved domain strongly

suggests that the yeast snRNA shares functional properties with metazoan U5. Recent experiments have begun to shed light on the role of the U5 snRNP in mammalian splicing. The U5 snRNP appears to recognize the region surrounding the 3' splice site by binding to the conserved polypyrimidine stretch, including the AG dinucleotide but excluding the branchpoint region (Chabot et al., 1985). This snRNP-intron interaction, which is apparently mediated by a U5-associated protein (Tazi et al., 1986; Gerke and Steitz, 1986), is likely to be one of the earliest events in intron recognition and splicing (Frendewey and Keller, 1985; Ruskin and Green, 1985; Chabot and Steitz, 1987). In contrast, there is debate as to whether Saccharomyces cerevisiae introns contain an element analogous to the mammalian polypyrimidine stretch (Py(11)NPyAG) (Mount, 1982), with the exception of the conserved PyAG motif. While it is clear that there is not a pyrimidine tract comparable in degree to that found in mammalian introns, the majority of yeast introns exhibit a substantial enrichment for uridines, especially at positions -7 to -13 upstream of the AG (Parker and Patterson, 1987). Independent of these concerns, we note that the MATa1 introns, which lack the uridine-rich tract found in other yeast introns (Parker and Patterson, 1987), exhibit an absolute requirement for snR7, indicating that the role of snR7 in splicing cannot be limited solely to recognition of the U-rich element.

Splicing of Some Introns May Be Initiated Without snR7

In several cases, depletion of snR7 results not only in the accumulation of unspliced precursors, but also of significant amounts

of lariat intermediate. We observe this intermediate for both cyh2 and actin mRNAs, and in the latter case we detect roughly equal amounts of precursor and lariat intermediate. This unexpected observation leads us to suggest that snR7 may not be absolutely required to complete the first step of splicing of these introns. The accumulation of the lariat intermediate suggests that spliceosomes formed on these substrates and 5' cleavage/lariat formation occurred, but that the second step of the reaction, 3' cleavage/exon ligation, did not take place. Since spliceosomes formed upon an individual transcript seem unlikely to undergo dissociation between the two steps of the reaction, this finding argues that these spliceosomes catalyzed the formation of lariat intermediate in the absence of snR7. Further experiments are required to directly test the hypothesis that, in the case of these introns, snR7 is not absolutely required for initiation of the splicing reaction in vivo. In any event, the increase in precursor levels demonstrates that snR7 is required for efficient completion of the first step. Finally, the fact that the lariats which are formed under these conditions are not efficiently converted to mRNA also argues that there is a requirement for snR7 at the second step of the splicing reaction.

EXPERIMENTAL PROCEDURES

MATERIALS

Reverse transcriptase was obtained from Life Sciences. Restriction enzymes were obtained from New England Biolabs. P32-ATP was purchased from ICN. P32-dCTP, Hybond hybridization membranes, and nick translation kits were obtained from Amersham. S1 nuclease was from Sigma. Oligonucleotides were synthesized by the Biomolecular Resource Center, UCSF.

PLASMID CONSTRUCTION

The snr7::TRP1 allele (which we formally designate snr7-3) was created by cloning the HpaI-HindIII fragment containing SNR7 into the HincII-HindIII sites of pUC18, to create pUC-7HH. This plasmid was then linearized with HincII, and a 1400 basepair EcoRI-EcoRI fragment containing the yeast TRP1 gene was inserted. Integration of this allele into the chromosome was achieved by cutting the plasmid (designated pUC-HHiT) with PvuII prior to transformation.

The (pGAL1)-SNR7 allele (formally designated snr7-5) was created by first cloning the 360 basepair AhaIII-HindIII fragment containing the snR7-encoding sequences into pUC18 to replace the HincII-HindIII fragment to create pUC-7AH. The GAL1 upstream sequences were obtained from pBM126 (a plasmid in which the XbaI site of deletion #153 (Johnston and Davis, 1984) had been converted to an BamH1 site by addition of a BamH1 linker) on an EcoRI-BamHI fragment. This fragment was cloned into the EcoRI-BamHI sites of pUC-7AH to create pUCG-7. The EcoRI-PvuII fragment of pUCG-7 was used to replace the

EcoRI-PvuII fragment from the Chloramphenicol gene of pBWHis (a derivative of pBR328 with the His3 gene inserted into the NdeI site and CENIV/ARS1 inserted into the PvuII site near the origin of replication (B. Patterson, unpublished results) to create pBWHisG7.

pBWHis600 was created by cloning the HindIII-ClaI fragment shown in Figure 1 into pUC18 to create pUC-7HC. The gene was then removed on an EcoRI-PvuII fragment and used to replace the EcoRI-PvuII fragment of pBWHis3 to create pBWHis600.

STRAINS

Gene replacements were performed in the isogenic diploid (YP52xYP54)
(a/α, ura3-52/ura3-52, trpl-289/trpl-289, his3-532/his3-532, lys2-801/lys2-801, ade2-101/ade2-101) (the kind gift of Phil Hieter). The haploid strain used for generating growth curves and from which RNA was made was an isogenic derivative of this strain and had the genotype (a ura3-52 trpl-289 his3-532 lys2-801 ade2-101 snr7::TRP1) and contained the plasmid pBWHisG7.

MEDIA

YEP-glucose, -His and -Trp (synthetic media lacking histidine or tryptophan) are described in Sherman et al., 1974 (revised yearly). YEP-galactose is identical to YEP-glucose except that galactose is used instead of glucose.

MEDIA SHIFT EXPERIMENTS

For experiments involving shifts in growth media, cells were grown in YEP-galactose to an A600 of 1, spun down, and resuspended in an equal volume of YEP-glucose. Growth curves were generated by continually diluting cultures as required to keep all A600 readings between 0.2 and 1.0.

S1 NUCLEASE PROTECTION

S1 experiments were performed as described by Parker and Guthrie (1985). DNA probes were generated by first cloning the desired restriction fragment (see Figure 2 legend) into mp10 or mp11. Labelled single-stranded DNA was generated by priming synthesis off of single-stranded m13 DNA using the sequencing primer (17mer), in the presence of labelled dCTP. This DNA was then cut with the appropriate restriction enzyme (either EcoRI or HindIII) to liberate the small fragment containing sequences complementary to snR7, which was then purified on a denaturing 5% polyacrylamide gel.

NORTHERN ANALYSIS

For small nuclear RNA Northerns, RNA samples (25 ug of total RNA) were electrophoresed on 6% polyacrylamide gels containing 7M urea and transferred electrophoretically to Hybond nylon membrane. Hybridizations were performed at room temperature in 6X SSC, 5X Denhardt's solution, 0.2% SDS and 1 million Cerenkov counts/ml of 5' end-labelled oligonucleotide. Filters were washed 3 times for 5 minutes each at room temperature in hybridization solution lacking Denhardt's, and once for 10 minutes at a temperature 10 degrees below the predicted Tm for the oligonucleotide being used.

mRNA Northerns were performed as described by Couto et al. (in prep.).

OLIGONUCLEOTIDES

The following oligonucleotides were employed in this work:

annealing to:		
snR7	5' CCTCCGCCATTGATCTGTA	3'
MATa1 exon 3	GAATTTATTTAGATCTCATACGTTT	
actin intron	CTAAACATATAATATAGACACAAA	
CYH2 exon 2	TGCTTACCGATACGACCTTTA	
S10 exon 2	CCGTTGACTGGGTAAGAAATG	
snR7 "loop"	CTGGTAAAAGGCAAG	

HOMOLOGY SEARCHES

Comparison of snR7 with mammalian U-snRNAs (first round comparisons were to rat snRNAs) were performed using the D3HOM program of David Mount (pers. comm.).

ACKNOWLEDGEMENTS

We thank Mark Johnston for the plasmid pBM126, Phil Hieter for the strain (YP52xYP54) and David Mount for the sequence analysis programs used in this work. We thank Lucita Esperas whose excellent technical assistance was the sine qua non of this work, and Judy Piccini for her after-hours contribution to the preparation of this manuscript. We also thank David Brow, Michele Marie Haltiner Jones, Sandy Johnson, Sandy Wolin, Beth Shuster and Nora Riedel for their comments on the manuscript. This work was supported by grants GM21119 from the National Institutes of Health and DCB-8603926 from the National Science Foundation to CG. BP was supported by NIH institutional training grant GM 07810.

REFERENCES

Ares, M. (1986). U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. Cell 47, 49-59.

Berget, S.M., and Robberson, B.L., (1986). U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. Cell 46, 691-696.

Black, D.L., Chabot, B., and Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell 42, 737-750.

Black, D.L., and Steitz, J.A. (1986). Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. Cell 46, 697-704.

Branlant, C., Krol, Al, Ebel, J.-P., Lazar, E., Haendler, B., and Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4 and U5 RNAs. EMBO J. 1, 1259-1265.

Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science 230, 1344-1349.

Chabot, B., and Steitz, J.A. (1987). Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. Mol. Cell. Biol. 7, 281-293.

Frendewey, D., and Keller, W. (1985). Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42, 355-367.

Fried, H.M., Nam, H.G., Loechel, S. and Teem, J. (1985). Characterization of yeast strains with conditionally expressed variants of ribosomal protein genes TCM1 and CYH2. Mol. and Cell. Biol. 5, 99-108.

Gerke, V. and Steitz, J.A. (1986). A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47, 973-984.

Guarente, L. and Mason, (1983). Heme regulates transcription of the CYC1 gene of S. cerevisiae via an upstream activation site. Cell 32, 1279-1286.

Guthrie, C., Riedel, N., Parker, R., Swerdlow, H. and Patterson, B. in Yeast Cell Biology, J. Hicks, Ed. (Alan R. Liss, Inc. New York, 1986) pp. 301-327.

Hernandez, N., and Weiner, A.M. (1986). Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. Cell 47, 249-258.

Johnston, M. and Davis, R.W. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. Mol. and Cell. Biol. 4, 1440-1448.

Krainer, A.R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell 42, 725-736.

Kramer, A., Keller, W., Appel, B., and Luhrmann, R. (1984). The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. Cell 38, 299-307.

Krol, A., and Ebel, J.-P. (1983). U1, U2 and U5 small nuclear RNAs are found in plants cells. Complete nucleotide sequence of the U5 snRNA family from pea nuclei. Nucleic Acids Res. 11, 8583-8594.

Krol, A., Gallinaro, H., Lazar, E., Jacob, M. and Branlant, C. (1981). The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. Nucleic Acids Res. 9:769-787.

Maniatis, R., and Reed, R. (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature, in press.

Mattaj, I.W., (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell 46, 905-911.

Mount, S.M. (1982). A catalogue of splice junction sequences. Nucleic Acids Res, 10, 459-472.

Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J.A. (1983). The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell 33, 509-518.

Noller, H.F. (1984). Structure of ribosomal RNA. Ann Rev Biochem 53, 119-162.

Parker, R., and Guthrie, C. (1985). A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. Cell 41, 107-118.

Parker, R., and Patterson, B. (1987). Architecture of fungal introns: implications for spliceosome assembly. In New Perspectives on the Molecular Biology of RNA, B. Dudock, ed. (Academic Press), in press.

Pikielny, C.W., and Rosbash, M., (1986). Specific snRNAs are associated with yeast spliceosomes. Cell 45, 869-877.

Reddy, R. (1986). Compilation of small RNA sequences. Nucleic Acids Res. 14(suppl.), r61-r72.

Riedel, N., Wolin, S., and Guthrie, C. (1987). A subset of yeast snRNA's contains functional binding sites for the highly conserved Sm antigen. Science 235, 328-331.

Riedel, N., Wise, J.A., Swerdlow, H., Mak, A. and Guthrie, C. (1986). Small nuclear RNAs from Saccharomyces cerevisiae: unexpected diversity in abundance, size and molecular complexity. Proc. Natl. Acad. Sci. USA 83, 8097-9001.

Rothstein, R.J. (1983), One-step gene disruption in yeast. Meth. Enzymol. 101, 202-211.

Ruskin, B., and Green, M.R. (1985). Specific and stable intron-factor interactions are established early during in vitro pre-mRNA splicing. Cell 43, 131-142.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sherman, F., Fink, G.R., and Lawrence, C.W. (1974, revised annually). Methods in yeast genetics. Cold Spring Harbor Laboratory, New York.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. (1986). A Protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47, 755-766.

Tollervey, D. and Guthrie, C. (1985). Deletion of a yeast small nuclear RNA gene impairs growth. EMBO J. 4, 3873-3878.

Tollervey, D., Wise, J.A. and Guthrie, C. (1983). A U4-like small nuclear RNA is dispensable in yeast. Cell 35, 753-762.

de Vegvar, H.E.N., Lund, E. and Dahlberg, J.E., (1986). 3' End formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. Cell 47, 259-266.

Wise, J.A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E.J., and Guthrie, C. (1983). Yeast contains small nuclear RNAs encoded by single copy genes. Cell 35, 743-751.

Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell 46, 827-835.

CHAPTER II: A NOVEL GENETIC SCREEN TO IDENTIFY U5-RELATED SPLICING FACTORS

The first real breakthrough in our efforts to comprehend the process of nuclear pre-mRNA splicing came with the proposal that the 5' splice site is recognized by a the small nuclear RNA component (snRNA) of a small ribonucleoprotein particle, the U1 snRNP (Mount et al., 1983). Subsequently, the involvement of three other snRNPs, designated U2, U4/U6 and U5 has been demonstrated. While the U2 snRNA has also been shown to be involved in substrate recognition via basepairing (Parker et al., 1987), the roles of the remaining three snRNPs have remained obscure. The role of the U5 snRNP has been particularly difficult to investigate because it is not amenable to the biochemical techniques used to study other snRNPs in mammalian systems: the U5 snRNP cannot be successfully targeted for oligo-directed RNAse H cleavage (Black et al., 1985), and until recently, no specific antibodies to its proteins have been reported (Jackson et al., 1988).

Nonetheless, several clues to the roles or activities of the U5 snRNP in splicing have been provided. First, one of its proteins, Intron Binding Protein or IBP, binds specifically to the 3' splice site in mammalian introns (Tazi et al., 1986, Gerke and Steitz, 1986). Second, in both yeast and mammals, U5 can form a substrate-independant association with the U4/U6 snRNP (Cheng and Abelson, 1987, Konarska and Sharp, 1987, Lamond et al., 1988). Further, in both systems, these two snRNPs join the spliceosome subsequent to the U1 and U2 snRNPS, and perhaps simultaneously with

one another (Cheng and Abelson, 1987, Konarska and Sharp, 1987, Lamond et al., 1988). Finally, in yeast we have shown that in vivo depletion of the U5 snRNA impairs splicing, although in some cases the first step of splicing occurs in the absence of the U5 snRNA (Patterson and Guthrie, 1987).

In order to further characterize the functions of the U5 snRNA in splicing, we initially made site-directed point mutations in the conserved loop of the snRNA. Alterations at positions 97, 98 and 99 confer temperature-sensitive growth defects on the cell. However, we have not been able to demonstrate a general splicing defect in these cells, even when grown under non-permissive conditions which result in lethality (our unpublished results).

Our inability to identify a biochemical defect in these mutants has led us to take a novel genetic approach to further our understanding of the U5 snRNP. We reasoned that the absence of a general splicing defect despite generation of a clear growth defect might reflect a failure to splice some subset of yeast messages efficiently. We therefore embarked upon a genetic screen which was designed to address this possibility in three ways: (1) identification of the inefficiently spliced messages, (2) amplification or generalization of the presumptive splicing defect in these cells, and (3) identification of gene products related to or involved in the function of the U5 snRNP. To facilitate analysis of the novel gene products we specifically sought conditional alleles in our screen.

Our screen relies on the following observations: The PRP8 gene product which has been shown to be a component of the U5 snRNP

(Jackson et al., 1988). Its temperature sensitive allele rna8-1 (prp8-1) has been shown to confer a splicing defect at the non-permissive temperature (Larkin and Woolford, 1983). This same temperature sensitive allele is lethal even at the permissive temperature in cells containing certain U5 point mutations as their only source of U5 snRNA (our unpublished results). In other words, while both the U5 alleles and the rna8-1 allele are viable at 25° and lethal at 36° the double mutants are inviable at all temperatures. We suggest that this inviability results from a lethal synergy of the defects in the two gene products, either of which provides sufficient function in a wild type environment. This type of synergistic behavior has been observed in several other systems. Some alleles of SAC1 (a locus isolated as a suppressor of the act1-1 ts allele) are lethal in combination with act1-2. The tub1-1 cs allele is lethal in combination with several alleles of TUB2 and TUB3 (Huffaker et al., 1987). In the cell cycle genes in Schizosaccharomyces pombe, cdc2-3w is lethal in combination with wee1-50^{ts} (Booher and Beach, 1987). Finally, in translation, two complementation groups isolated as suppressors of HIS4 initiation codon mutants are lethal when combined with similar mutations in the β-subunit of eIF-2 (Donahue at al., 1988).

Unfortunately, none of these defects are understood at the molecular level. A useful way to think about the phenomenon is as analogous to the case of a binding site, in which the combination of two single mutations is far more deleterious than predicted by simply adding their effects. Indeed, this effect has been demonstrated for mutations in the different consensus sequences in the yeast rp51A

intron. When mutants at the 5' splice site and TACTAAC box were combined, the effects on splicing were orders of magnitude more drastic than predicted even by multiplying the defects of the two starting mutations (Jacquier et al., 1985).

The useful features of this type of screen are several. First, if it is true that the defect in U5 mutant cells is due to the inefficient splicing of several substrates, these could not be identified by seeking classical gain-of-function mutants because all limiting substrates would have to be healed, requiring multiple genetic events. However a loss-of-function mutant (one which was sufficient to support viability in a wild type cell but not in combination with a mutant U5 allele) in any one of the hypothetical substrates could suffice to make the cell inviable, thus allowing us to isolate such a substrate as a lethal synergy locus. The lethal synergy phenotypes might also arise from a broadening of the initial defect such that it becomes apparent for any splicing substrate examined. Finally, the theoretical target size for mutants showing lethal synergy phenotypes should be much higher than for gain of function suppressors, as the latter require a healing of a specific interaction. Finally, since we are looking for defective molecules we anticipate that this screen will enrich for ts alleles. Such alleles allow us ready access to the phenotypes of the new mutants directly. The precedent that the prp8-1 allele and other synergistic lethality alleles are also ts is reassuring in this regard.

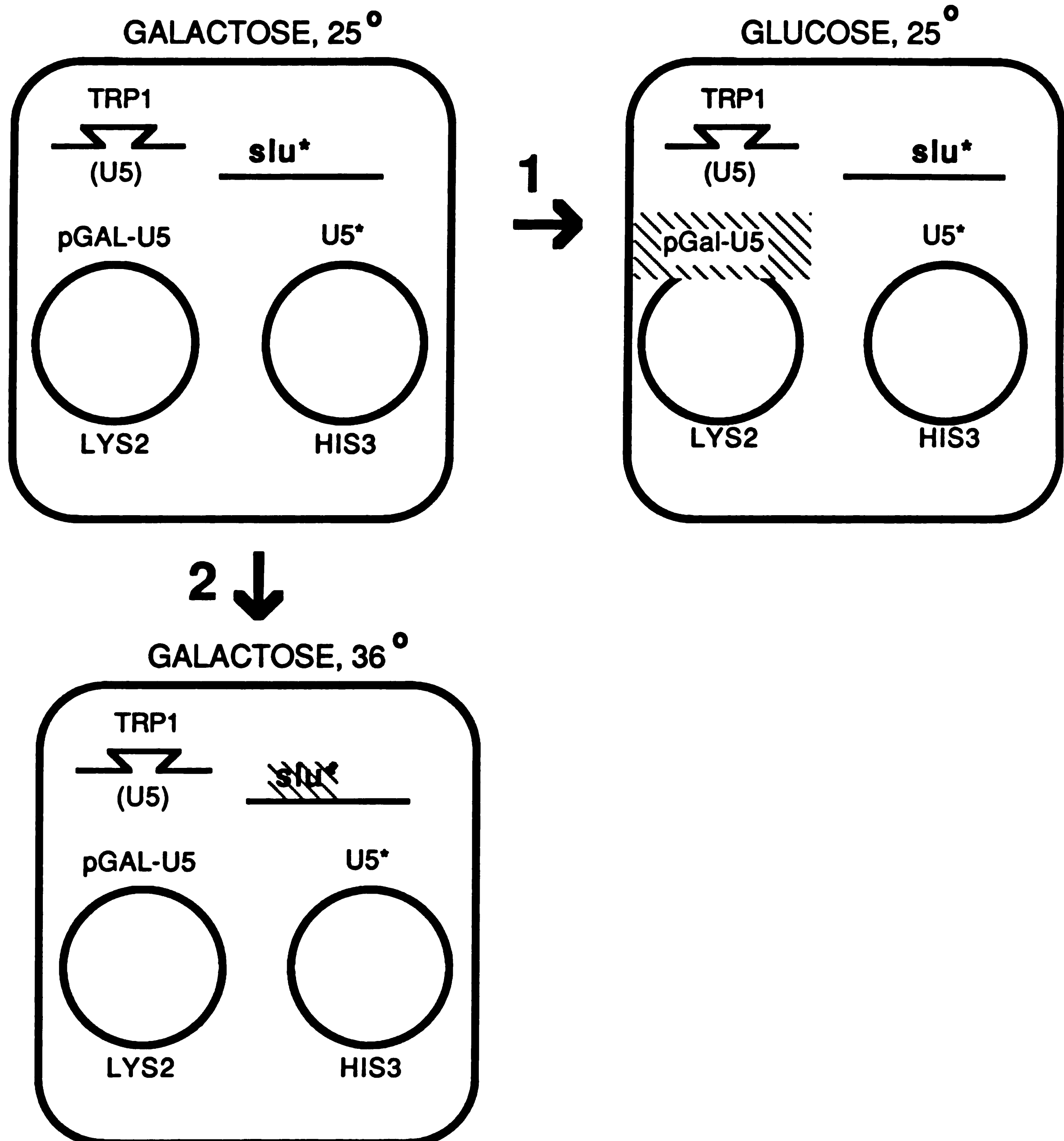
RESULTS

Isolation of Mutants

The strategy we developed is similar to one described by Huffaker et al. (1987). In order to detect the lethal synergy phenotype we first constructed a cell in which the only wild type copy of U5 was under the control of the inducible GAL1 promoter (pGAL1). This was done by disrupting the chromosomal copy of U5 (and marking the disruption with the TRP1 gene) and providing the pGAL1-U5 fusion on a centromere plasmid (which bore the LYS2 gene as a selectable marker). Thus in the absence of any other functional U5, these cells are viable on galactose where the pGAL1-U5 fusion is expressed, but inviable on glucose where the fusion is repressed because no other source of U5 is provided. We performed the screen on the three mutant alleles of U5 which confer growth defects, designated A98, C98 and C97,C99 (this nomenclature parallels that of tRNAs in which the novel nucleotide is given followed by its position in the wild type transcript). These were transformed into the starting parent on YCp50-derived vectors.

The starting parent and strategy for the mutagenesis are shown in Figure I. The unmutagenized parent strains (containing one of the three U5 mutations) are all viable on glucose because the point mutations are sufficient to support growth. The A98 mutation causes a two-fold increase in doubling time, while C98 and C97,C99 increase doubling time by about four-fold. All three alleles showed more severe phenotypes at 36°, but only C98 is lethal at this temperature (B. Patterson and D. Frank, unpublished results). The strains were mutagenized by UV irradiation to 10% survival. The mutagenized cells

FIGURE I. Strategy for identification of $\underline{slu}^{ts}$ (synthetic lethal with U5) alleles. The starting parent is shown. After mutagenesis, the colonies are (1) replica plated to glucose to assay for the synergistic lethal phenotype, and (2) replica plated to galactose at 36° to detect temperature-sensitive alleles. Colonies which fail to grow under both these conditions are selected for further analysis. Diagonal lines indicate repressed or inactive gene products. Asterisks denote mutations in the gene (thus $U5^{*}$ represents U5-A98, U5-C98 or U5-C97,C99).



1. Screen for viability in absence of WT U5

2. Screen for ts phenotype in presence of WT U5

were plated out and the plates wrapped in foil (mutagenesis and outgrowth were done in total darkness to prevent light-mediated repair) and allowed to form single colonies at room temperature on galactose plates, with selection for the plasmids (-Lys and -His) and the chromosomal disruption (-Trp).

In order to screen for mutants displaying the lethal synergy phenotype, cells were replica plated to glucose-containing media at 25°, maintaining selection for the plasmids and disruption. This replica plating was done twice sequentially because depletion of wild type U5 requires 6 generations (Patterson and Guthrie, 1987), producing colonies on the first replica. We then examined the glucose plates for colonies which were inviable. These represented the starting class of mutants for further analysis. Of approximately 2000 colonies screened for each starting point mutation, we recovered 2 mutant colonies from the A98 strain, 7 from the C98, and 3 from the C97,C99. Upon retesting, we retained neither of the A98 derived mutants, 5 of the C98-derived mutants, and all of the C97,C99 mutants.

Characterization of Mutants

Since one of the goals of the screen was to recover conditional alleles, we first screened for ability of our mutants to grow on galactose at 36° (i.e., in the presence of wild type U5, thus assaying for additional defects present in the cells not related to defects arising from the U5 point mutants for detection). Of the original mutants, two derived from C98 and all three of those derived

from C97,C99 displayed no growth or significantly impaired growth at the high temperature. We chose to pursue only these mutants for our studies, and designated them slu1 through slu5, for synergistic lethal with U5. We use the term "synergistic lethal" rather than "synthetic lethal" to distinguish these mutants (in which we seek interacting factors) from the classically described case in which both routes of an alternate pathway are perturbed. slu1, 2 and 3 were isolated from C97,C99-containing background and slu4 and 5 were isolated from a C98-containing background.

It is worth noting that the specificity of this type of lethal synergy is not known, but we have shown that the starting U5 alleles do not display this phenotype in conjunction with ts alleles of RNA (PRP)3,4,5,6 and 7. A further suggestion that the synergistic lethality with the rna8-1 gene product is related to the A98, C98 and C97,C99 comes from the finding that the phenomenon is not observed when U5 mutants bearing alterations in the conserved Sm binding site are placed in an rna8-1 background (M. Jones, pers. comm.).

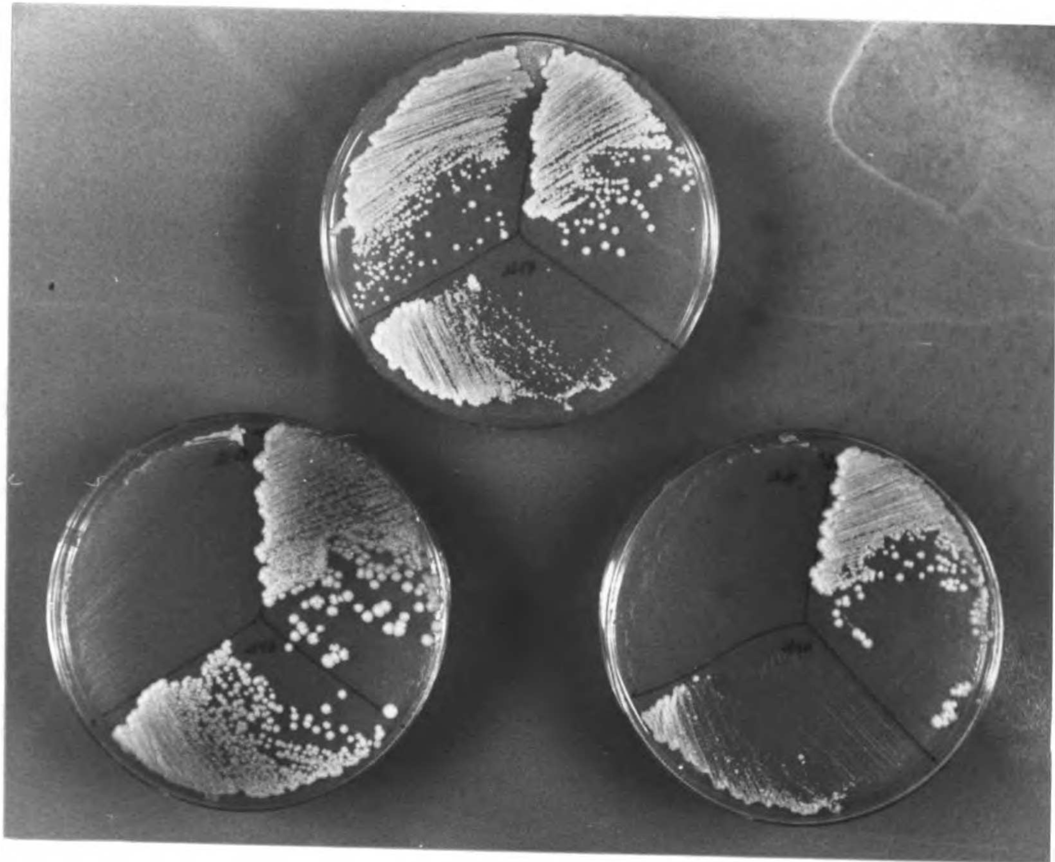
Our first concern in characterization of the mutants was whether the ts phenotype we observed was the result of the same mutation as that causing the synergistic lethal phenotype. To determine this, we crossed the mutants to a cell similar to the starting parent, but of opposite mating type, bearing a HIS3 plasmid with the pGAL1-U5 fusion, and containing no other copy of the U5 point mutants. The resulting diploids were then sporulated and dissected. Analysis of the resulting progeny was complicated in that 1) both chromosomal copies of of the U5 gene were disrupted such that the complementing plasmid had to be inherited for viability, 2) complementation by the

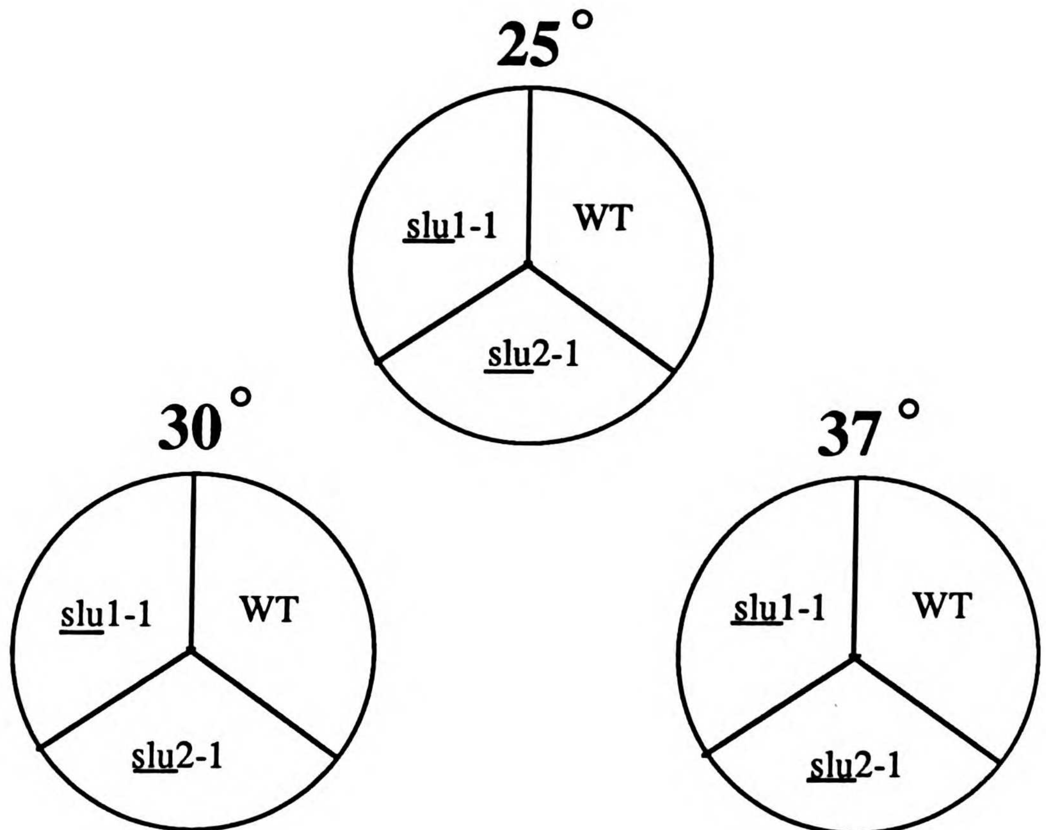
pGAL1-U5 fusion is not 100% during spore germination and outgrowth, and 3) the heavy mutagenesis of the strains resulted in lowered viability of the progeny. Nonetheless, in all cases we observed almost complete co-segregation of the lethal synergy and temperature sensitive phenotypes. For slu1-1 and slu4-1 we performed a second outcross, and in these cases observed complete cosegregation of the two phenotypes.

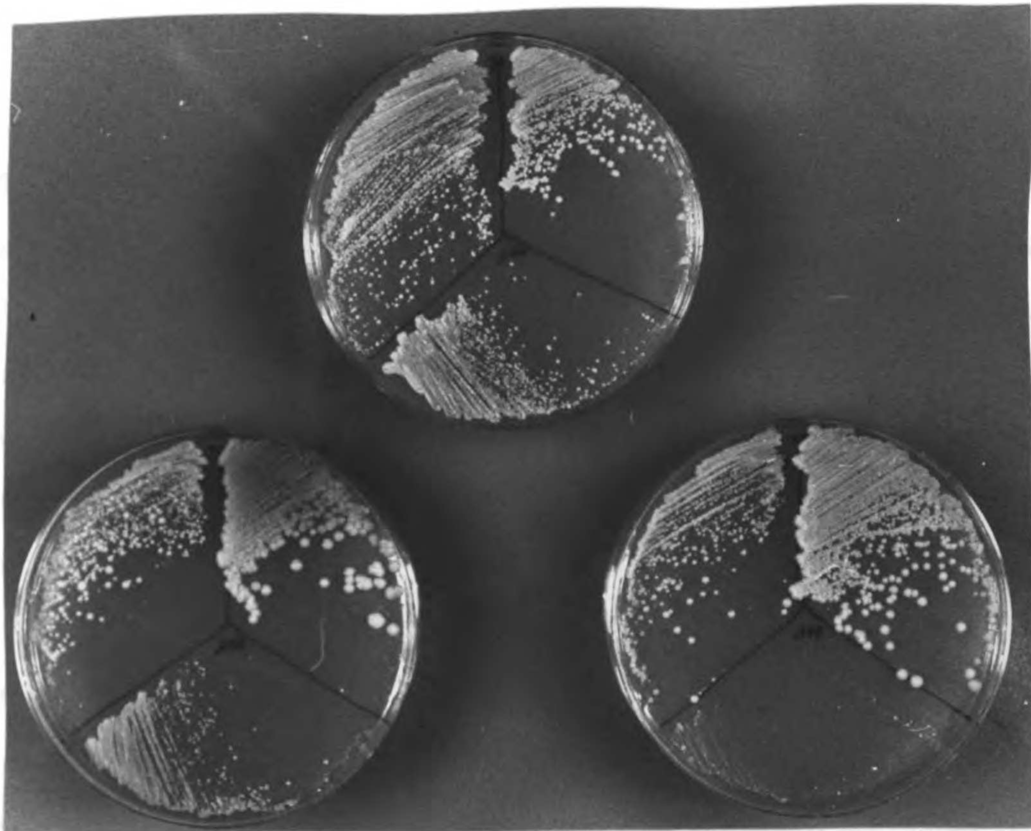
Before proceeding further with our analysis of the mutants, we wanted to assess their effects in a wild-type background. We had two goals here: 1) to examine the mutants in the context of a "naive" U5 gene which had not been present during the mutagenesis, and 2) to further remove from the backgrounds other mutations which resulted from the mutagenesis but were unrelated to the slu^{ts} phenotypes. For this reason, we crossed each of the five strains derived from the segregation analysis to an isogenic wild type cell. For slu1-1,3-1,4-1 and 5-1, this cross was repeated a second time. In the course of this work, slu5-1 was discarded, because it grows very slowly even at room temperature, and higher temperatures only mildly increased this phenotype.

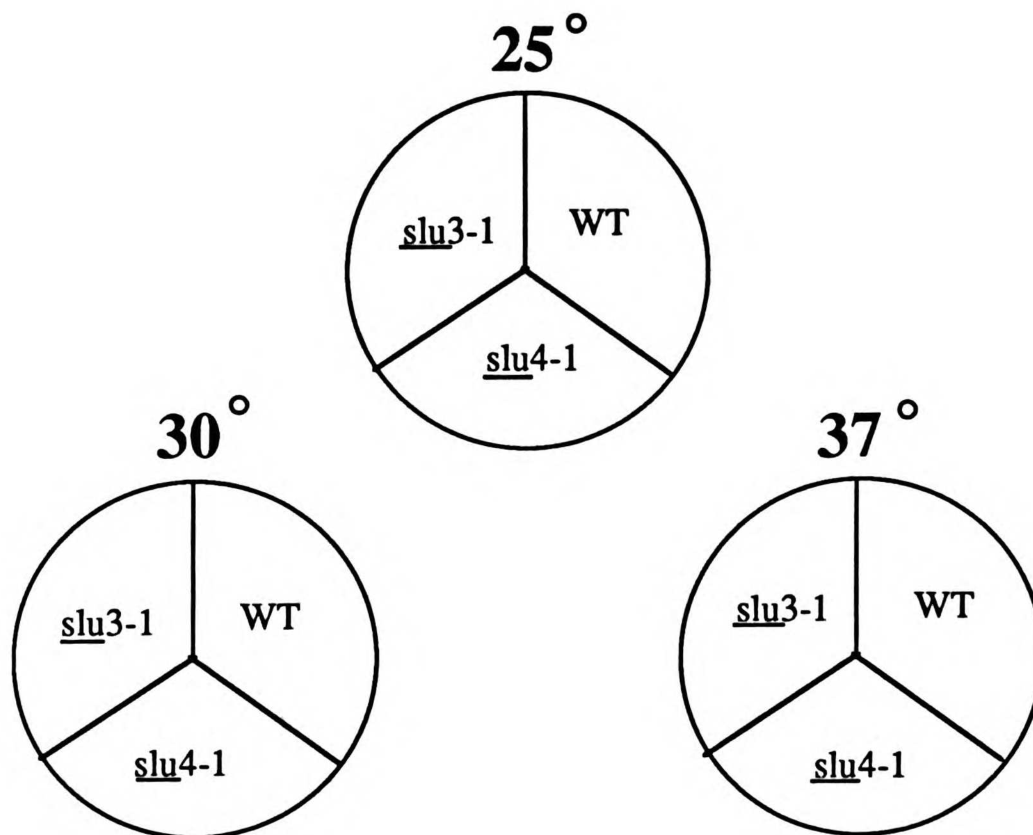
Having isolated the remaining four mutants as reasonably healthy segregants, we wanted to characterize them further. We determined their growth characteristics by examining their ability to form single colonies at 25°, 30° and 37°. The results of this experiment are shown in Figure II. Note that slu1-1 is lethal at 30°, and slu1-1 and slu3-1 are lethal at 36°, while the others are only very impaired.

Figure II. Growth phenotypes of <u>slu</u>^{ts} alleles. Cells were streaked out for single colonies and grown for four days on glucose at the temperatures indicated. A) <u>slu</u>1-1 and <u>slu</u>2-1. B) <u>slu</u>3-1 and <u>slu</u>4-1. WT is an isogenic strain containing a wild-type U5 allele. The mutants are the outcrossed derivatives which also have a wild-type chromosomal U5 allele.




25°
slu1-1
WT
slu2-1
30°
slu1-1
WT
slu2-1
37°
slu1-1
WT
slu2-1





25°
slu3-1
WT
slu4-1
30°
slu3-1
WT
slu4-1
37°
slu3-1
WT
slu4-1

We next set out to assign the mutants to complementation groups. This was done simply by cross stamping opposite mating types of the alleles against one another. We then checked for complementation of the ts defect. In this way, we determined that they indeed represent four different complementation groups. We also sought to determine whether they were alleles of any of the previously characterized rna (prp) genes. Again, we found that they complement ts alleles of RNA (PRP) 2,3,4,5,6,7,8,9 and 11. We therefore conclude that we have identified four complementation groups not previously identified as splicing genes.

At this point, however, we had no evidence that these factors were actually involved in splicing. Given the uncertainty as to what secondary defects might give rise to synergistic lethal phenotypes in conjunction with the U5 point mutations, we sought to assess the splicing phenotypes for the mutants. In order to do so, we extracted RNA from these cells at the non-permissive temperature. First, however, we needed to determine the time of onset of the temperature-induced growth defect in order to examine the RNA just as this defect was becoming manifest. To determine this, cells were grown to an OD_{600} of 0.5 to 1 (log phase) at 25° and shifted to pre-warmed cultures at 36°. Growth rate was followed by monitoring the change in OD_{600}. In all cases, the change in growth rate occurs at about six hours (data not shown), so we selected this timepoint for RNA analysis.

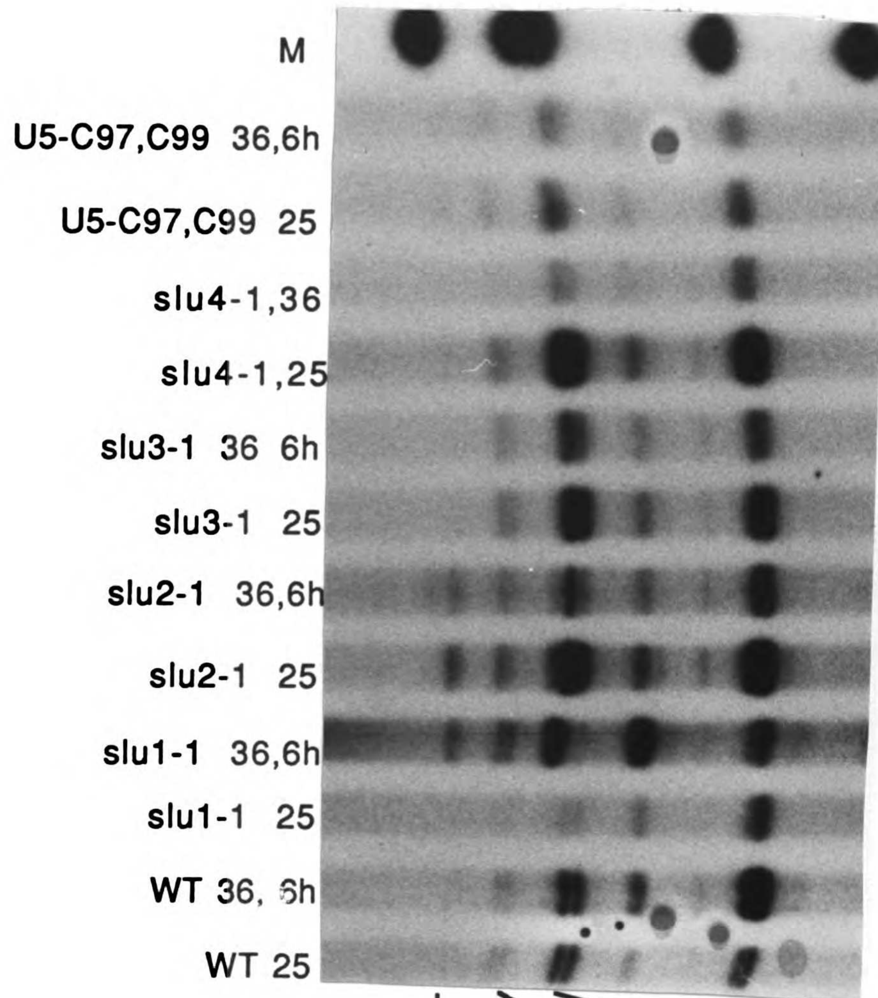
To assay the extracted RNA for splicing defects, we chose to examine the yeast MATa1 gene because it is a very sensitive indicator of splicing defects and depletion for U5 gives rise to a pronounced

splicing defect in the MATa1 transcripts. The results of our primer extension analysis are shown in Figure III. We note that while wild type cells (lanes 1 and 2) and C97,C99 mutants cells (lanes 11 and 12) show no splicing defect at low or high temperatures, slu1-1 exhibits accumulation of precursor and semi-mature (one intron removed) species at the non-permissive temperature. In contrast slu2-1 appears to have a constitutive partial defect in splicing MATa1 transcripts which is not exacerbated at the restrictive temperature. slu3-1 and slu4-1 do not display any defect in splicing MATa1. Since the slu1-1 cells also die at 30°, and to further characterize the MATa1 splicing defect in these cells, we examined RNA made from these cells shift to 36° for 6, 8 or 10 hours and from cells shifted to 30° for 8 or 10 hours. In all cases we observe the same pattern of accumulation of precursor and semi-mature species (Figure IV). We have also examined the splicing of the CYH2 transcript in these cells but do not observe dramatic accumulation of precursor in any of the mutants (data not shown).

DISCUSSION

We have used a novel genetic screen to identify at least two new loci involved in pre-mRNA splicing in yeast. The screen relies on observations from many systems in which partial defects in two related factors result in lethality when combined in a genetic background (Huffaker et al., 1987, Booher and Beach, 1987, Donahue et al., 1988). The use of this type of approach in the forward direction (i.e., identification of new mutations rather than

FIGURE III. Splicing phenotypes of _slu_^{ts} alleles. Cells were grown either at 25° or shifted to 36° for 6 hours. All strains contained wild type copies of the U5 gene excepting the C97,C99 strains, which carried a chromosomal disruption and the mutant U5 allele on a plasmid. The splicing intermediates and products of the _MAT_a1 transcript are indicated. Boxes represent exons and lines represent introns.


M
U5-C97,C99 36,6h
U5-C97,C99 25
slu4-1,36
slu4-1,25
slu3-1 36 6h
slu3-1 25
slu2-1 36,6h
slu2-1 25
slu1-1 36,6h
slu1-1 25
WT 36, 6h
WT 25

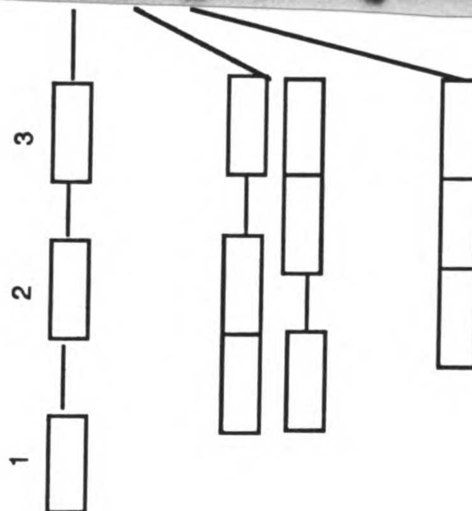
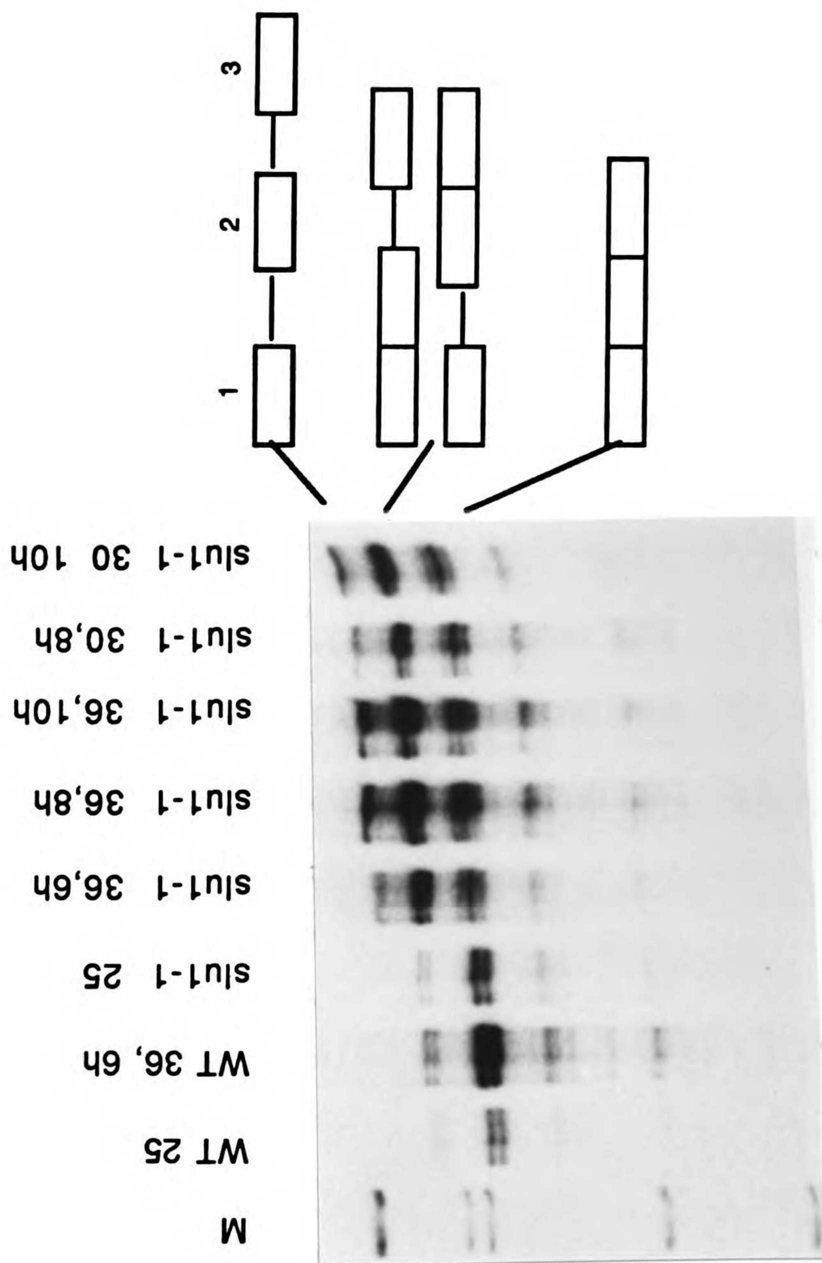

3
2
1

FIGURE IV. Splicing phenotypes of *slu*1-1. *slu*1-1 cells were shifted for the times and temperatures indicated prior to RNA extraction. The transcript is *MAT*a1. Symbols are as in Fig. III.


1
2
3
slu1-1 30 10h
slu1-1 30,8h
slu1-1 36,10h
slu1-1 36,8h
slu1-1 36,6h
slu1-1 25
WT 36, 6h
WT 25
M

observation of synergy among previously characterized mutations) is novel, and it is therefore important to confirm that mutations being uncovered are directly relevant to the pathway being investigated.

We cannot yet say whether the mutations that we have identified are specifically related to U5 function or are more indirectly involved in the splicing process. Nonetheless, we have shown that at least two of the four complementation groups we have recovered affect mRNA splicing, demonstrating the utility of this approach for generating novel splicing mutations. In order to assess more fully the basis of the lethal synergy, the most important next experiment will be to examine pre-mRNA splicing under the "synergistic lethality" conditions. Cells bearing the U5 point mutation and the $\underline{slu}^{ts}$ allele will be assayed for splicing phenotypes in the absence of wild-type U5. We anticipate that the defect under these conditions is similar to that caused by heat inactivation of the $\underline{slu}^{ts}$ gene products, i.e. the same pattern of MATa1 precursor accumulation for slu1-1.

The fact that we observe splicing defects for two of the complementation groups argues that the approach will be generally successful in identifying complementation groups related to the original defects. The fact that the four mutants define four complementation groups indicates that there remain a large number of genes which could be identified by this technique and justifies scaling up the screen to uncover the full set.

The defect that we observe in slu1-1 cells is intriguing. The primary product of MATa1 splicing appears to be a species lacking one but not both introns. As shown in Figure IV, under no conditions do

we observe an absolute splicing defect. Indeed, the uniformity of the defect seen is suggests that heat inactivation of the slu1-1 gene product impairs the efficiency of splicing of the two introns differentially (i.e., one intron is amost completely blocked for splicing while the other is only slightly impaired. This hypothesis could be tested by performing primer extension or Northern analysis with probes specific for each of the two introns. This observation, combined with the failure to observe a significant defect in the splicing of the CYH2 transcript suggests that we are still not observing the actual cause of lethality in these cells (i.e., there is still no general splicing defect) and also supports our thesis that different introns may respond very differently to a given splicing defect. Again, it will be interesting to see what the phenotype of this mutant is under synergistic lethality conditions.

As pointed out in the Introduction, one goal of developing this type of screen was to allow us to identify genetically the hypothetical substrates whose specific inability to be spliced gives rise to the A98, C98 and C97,C99 growth defects. Clearly, the two ts alleles which show splicing defects are not candidates for substrate mutations. Fortunately, the methods required to identify whether the synergistic lethal locus is a splicing substrate are relatively straightforward. To do so, one clones the genes (by complementation of the ts phenotype, or the lethal synergy phenotype) and then characterizes the transcripts produced by the clone. In each case a wild type bank would be employed. Cloning by complementation of the ts phenotype is more straightforward, and clones could then be

re-screened by assaying their ability to alleviate the synergistic lethality phenotype.

True "substrate" clones will encode spliced transcripts, which will accumulate as unspliced precursor in, for example, the rna strains at the non-permissive temperature. Further, these genes should show impaired splicing in the U5 mutants, at the non-permissive temperature in the ts strain from which they were isolated, and a severe defect under synergistic lethal conditions. Thus if clones were nick-translated and used to probe a Northern blot containing RNA from the above mentioned conditions substrates sensitive to defective U5 molecules could be readily identified.

In summary, future directions include: 1) Further characterization of slu1-1 and slu2-1 to define the spectrum of their affects on spliced messages. Characterizing which substrates are and are not susceptible to defects in a given slu^{ts} background could provide clues to its site or mechanism of action. 2) Determination of the relatedness of slu1 through slu4 to the U5 snRNP by looking at splicing under synergistic lethality conditions 3) Investigating whether the gene products alter the structure of the U5 snRNP by examining the stability and density of U5 snRNPs from these strains. 4) Searching for substrate lesions by the methods outlined above.

REFERENCES

Black, D.L., Chabot, B., and Steitz, J.A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell 42: 737-50.

Booher, R. and Beach, D. 1987. Interaction between cdc13+ and cdc2+ in the control of mitosis in fission yeast; dissociation of the G1 and G2 roles of the cdc2+ protein kinase. EMBO J. 6:3441-47.

Cheng, S.-C. and Abelson, J. 1987. Spliceosome assembly in yeast. Genes and Develop. 1:1014-27.

Donahue, T.F., Cigan, A.M., Pabich, E.K. and Valavicius, B.C. 1988. Mutations at a Zn(II) finger motif in the yeast eIF-2B gene alter ribosomal start-site selection during the scanning process. Cell 54:621-32.

Gerke, V., and Steitz, J.A. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47:973-84.

Huffaker, T.C., Hoyt, M.A. and Botstein, D. 1987. Genetic analysis of the yeast cytoskeleton. Ann. Rev. Genet. 21:259-84.

Jackson, S.P., Lossky, M. and Beggs, J.D. 1988. Cloning of the RNA8 gene of Saccharomyces cerevisiae, detection of the RNA8 protein and

demonstration that it is essential for nuclear pre-mRNA splicing. Mol. Cell. Biol. 8:1067-75.

Jacquier, A. Rodriguez, J.R. and Rosbash, M. 1985. A quantitative analysis of the effects of 5' junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. Cell 43:423-30.

Konarska, M.M. and Sharp, P.A. 1987. Interaction between snRNP particles in formation of spliceosomes. Cell49:763-74.

Lamond, A.I., Konarska, M.M., Grabowski, P.J. and Sharp, P.A. 1988. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Genes and Develop. 1:532-43.

Larkin, J.C. and Woolford, J.L. 1983. Molecular cloning and analysis of the CRY1 gene: a yeast ribosomal protein gene. Nucl. Acids Res. 11:403-421.

Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J.A. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell 33:509-18.

Parker, R., Siliciano, P., and Guthrie, C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves basepairing to the U2-like snRNA. Cell 49:220-39.

Patterson, B. and Guthrie, C. An essential yeast snRNA with a U5-like domain is required for splicing in vivo. Cell 49:613-24.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47:755-66.

CHAPTER III: ARCHITECTURE OF FUNGAL INTRONS: IMPLICATIONS FOR SPLICEOSOME ASSEMBLY

INTRODUCTION

The removal of intervening sequences (IVS) by the process of RNA splicing is an elementary step in gene expression. Nuclear mRNA splicing in mammals and in yeast is mediated by a group of small nuclear RNAs complexed with protein (snRNPs) (for reviews, see Maniatis and Reed, 1987; Green, 1986). One role of snRNPs in the splicing process appears to be to assemble with the precursor into a large complex, termed the spliceosome, which is capable of the splicing process. The discovery of self-splicing mitochondrial Group II introns (in which splicing proceeds via the lariat intermediate seen in nuclear mRNA splicing) (Peebles et al., 1986) has suggested the possibility that the assembly of the splicing complex by snRNPs may function primarily to bring the precursor into the proper geometry for a reaction which can then be viewed as being fundamentally autocatalytic (see Cech, 1986).

Two general paradigms have emerged for the mechanism by which snRNPs participate in the assembly of the splicing complex. Considerable evidence supports the conclusion that specific snRNPs recognize and bind conserved sequences within the intron. In particular, the U1 RNA containing snRNP (Mount et al., 1983), the U2 snRNP (black et al., 1985) and the U5 snRNP (Chabot et al., 1985, Gerke and Steitz, 1986; Tazi et al., 1986) bind to the 5' junction, the branch site, and the 3' splice site, respectively. In addition,

the snRNPs are thought to organize the spliceosome by interactions with oneanother, though there are no data that bear directly on this point. Thus, spliceosome assembly, and hence splicing, would appear to be dependent on a variety of both intron-snRNP and snRNP-snRNP interactions.

In principle, the structure of the intron or the spatial arrangement of the conserved intron sequences may promote or influence splicing in two ways. First, if the reaction is fundamentally autocatalytic, an intron with a high degree of the proper secondary and tertiary structure may position the nucleotides involved in bond clearvage and ligation in the proper geometry for the splicing process to occur. Indeed, if the degree of structure is sufficient, such intorns may be able to partially or wholly dispense with some trans interactions. Note that, while some mitochondrial Group II introns can self-splice in vitro, genetic evidence suggests that other members of this class of intron require trans-acting factors (Bertrand et al., 1982, McGraw and Tzagoloff, 1983; Piller et al., 1983; Simon and Faye, 1984). Alternatively, and what seems more likely from our current understanding of nuclear pre-mRNA splicing, the spatial arrangement of the conserved sequence elements may position the snRNPs in the proper geometry to interact with each other, and therfore to assemble into a functional spliceosome. If this was true, we would expect to find the spatial relationship between splicing signals to be nonrandom, and, on average, to illustrate the optimal arrangement between any such pair of signals. For example, branchpoints in mammalian introns appear to be a relatively constant distance from the 3' splice site, thus suggesting

an interaction between factros, presumably snRNPs, at each of these sites (Ruskin et al., 1985, Reed and Maniatis, 1985).

In this light we have examined the spacing of the conserved intron sequences found at the 5' splice site, the branch site, and the 3' splice junction in introns from Saccharomyces cerevisiae and other fungi. In that the primary sequences of the 5' junction sequence and the branch point, or TACTAAC box in Saccharomyces cerevisiae, are more strictly conserved in fungal introns (see Guthrie et al., 1986), possible secondary structures or spatial arrangements may also be highly conserved. Based on our analyses, we conclude that there are conserved spatial relationships between both the 5' splice site and the branchpoint, as wells as between the branchpoint and the 3' splice site. We hypothesize that these spacings are optimal to promote interactions of the snRNPs at each of these sites. In cases where the spacing between elements is not optimal, as based on linear distance, we propose that there are additional intron features that restore the optimal geometry. In the case of the branch point to 3' splice site, we suggest this to occur by the interaction of a trans-acting factor with an additional conservered sequence element, a polyuridine tract. Alternatively, as in the 5' junction to branch point case, we can identify cis-acting secondary structures that restore the optimal spacing.

II.BRANCH SITE-3'SPLICE JUNCTION RELATIONSHIP

As our first step in the examination of the relationships of conserved intron sequences, we gathered a data base of a number of

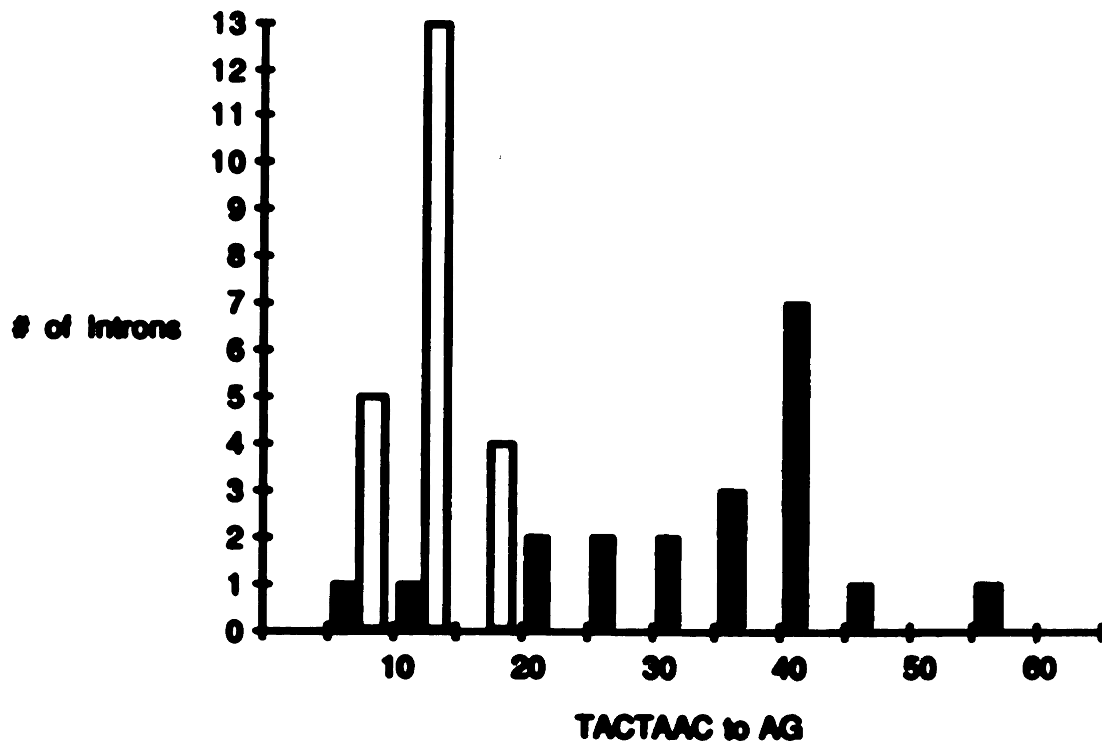
fungal introns. The introns we have used for the analyses in this paper include 21 introns from Saccharomyces cerevisiae, five introns from Schizosaccharomyces pombe, three introns from Schizophyllum commune, five introns from Neurospora crassa, five introns from Asperigillus niger, and four introns from Trichoderma reesei. These introns are listed and the references given in the legend to Fig. 1.

A. Two Types of 3' Splice Sites in Fungi

We have examined the structure of the 3' splice site by comparing the distance between the branch point and the AG at the 3' junction. The results of this analysis, illustrated in the histogram in Fig. 1, indicate that the distance between these two sequence elements is not random, and moreover suggest that there are two types of spacings. In introns from fungi other than S. cerevisiae and in two of the S. cerevisiae introns (found in the MATa1 gene), the branch site is quite close to the AG at the 3' junction, the distance varying between 5 and 15 nucleotides. We refer to introns containing this type of 3' splice site as being type 3'S introns (for short 3' splice site). In contrast, in the reminder of S. cerevisiae introns the spacing between the branch site and the 3' junction is significantly larger (22-137 nucleotides) with a distinct preference for a spacing of approximately 40 nucleotides. We refer to this type of intron as being a type 3'L intron (for long 3' splice site).

Having discerned two different classes of 3' splice site regions in fungal introns, we searched for further similarities within the introns in each class. Type 3'S introns do not appear to possess any

Figure I. Branchpoint to AG spacings in fungal introns. The figure shows the number of nucleotides between the branchpoint and the 3' splice site in a number of fungal introns. The nucleotides are counted from the nucleotide 3' of the branchpoint to the nucleotide immediately preceding the AG. The number of introns within a given spacing interval are represented for *Saccharomyces cerevisiae* (dark bars) and several other fungi (open bars).



13
12
11
10
9
8
7
# of Introns
6
5
4
3
2
1
0
10
20
30
40
50
60
TACTAAC to AG

conserved sequences other than the absolutely conserved AG dinucleotide found at the 3' splice site of all nuclear mRNA introns (data not shown). Even the nucleotide preceding the AG, which is a pyrimidine in all mammalian introns and all fungal 3'L introns, is sometimes a purine in 3'S introns.

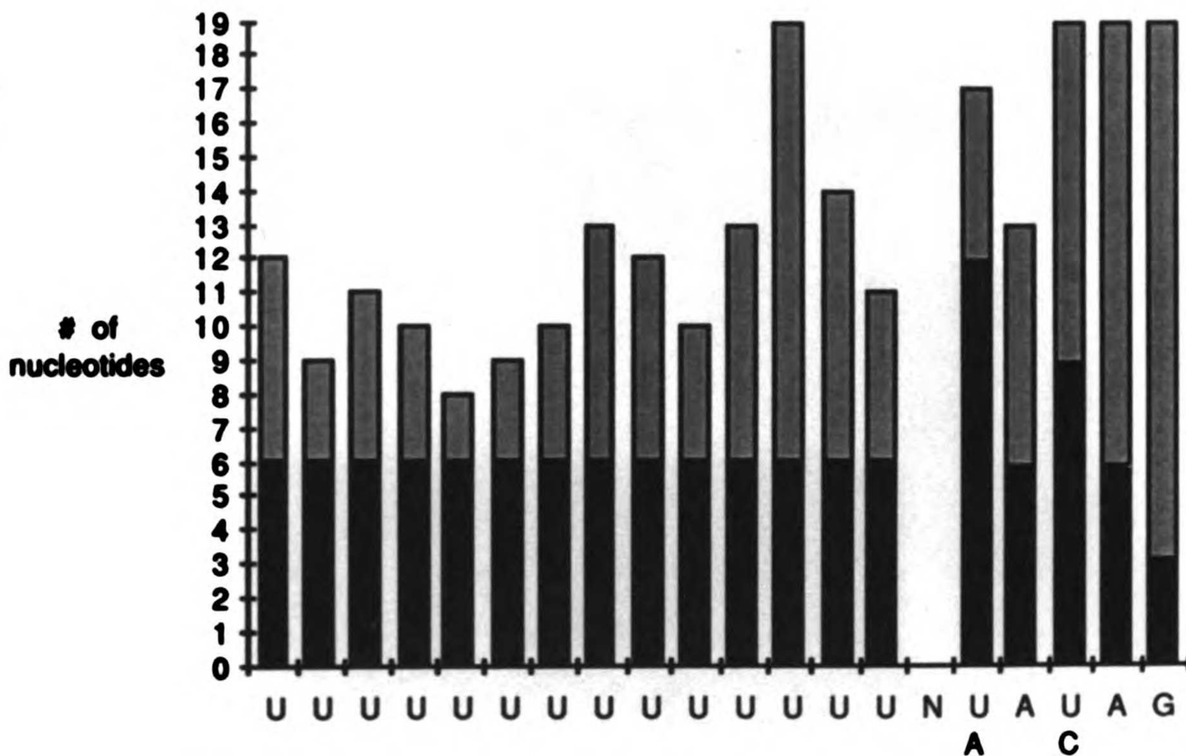
In 3'L introns, on the other hand, we find significant sequence conservation in the region upstream of the AG dinucleotide. As illustrated in Fig. 2, we observe both a 3' splice site sequence of N(U or A)APyAG, as well as a general enrichment of uridines extending at least to position -20 (as measured from the site of 3' cleavage). This enrichment is especially striking from nucleotides -7 to -13, and includes a 100% conserved uridine at position -9. The relative enrichment of uridines is demonstrated in Fig. 2B, which shows the expected (based on the base compostion of the 21 S. cerevisiae introns discussed in this work) and observed frequencies of the consensus nucleotides at each position.

Extensive sequence analyses in both mammals (Mount, 1982) and plants (Brown et al., 1986) have been performed, and it is interesting to compare the consensus sequences derived with the one we suggest for 3'L introns of S. cerevisiae. Mammalian introns terminate in the sequence NPyAG, those from plants in GPyAG, and those in yeast in N(U or A)APyAG (in all cases, we arbitrarily define the 5' end of these subsequences as the first position which is not a pyrimidine). Thus in each case we observe a conserved sequence motif, although the sequences share only the PyAG as a common signal. In all cases, the region preceding this sequence is pyrimidine rich. In mammalian introns, the enrichment is especially striking,

Figure II. Analysis of conserved sequences in 3' splice sites in 3'L introns. (A) Sequence alignment and analysis: the 3' terminal 20 nucleotides (including the AG) of 19 _Saccharomyces_ _cerevisiae_ introns have been aligned. The absolute number of nucleotides observed at each position (Total of 19) are shown, as well as the percentage of each nucleotide at each position. The expected numbers of and percentages for each nucleotide (calculated from the base composition of the _S_. _cerevisiae_ introns in Fig. 1) are shown in parentheses on the right. The consensus sequence we suggest is shown at the bottom. (B) Relative frequencies of consensus nucleotides at each position. Solid bars, the expected number of occurrences of the consensus nucleotide(s) at each position; stippled bars, occurrence of the consensus nucleotide(s) above the level expected.

A

U	G	U	C	U	A	U	A	U	U	A	U	A	U	G	U	U	U	A	G	actin
A	C	A	A	U	A	U	U	U	U	U	U	U	U	G	U	A	C	A	G	cyh2 (L29)
U	G	A	U	U	A	A	U	G	A	U	U	C	A	U	C	G	C	A	G	kin28
U	U	G	A	G	A	U	C	U	U	U	U	U	U	A	A	C	U	A	G	L17A
U	G	A	A	U	U	U	A	C	U	U	U	U	U	G	U	A	U	A	G	L25
U	A	C	C	A	U	U	U	U	A	U	U	U	U	U	A	A	U	A	G	L34
U	U	U	G	A	U	U	U	U	G	U	U	U	U	C	U	A	C	A	G	rp16A
U	U	G	U	G	C	A	U	U	U	U	U	U	C	A	A	U	U	A	G	rp16B
U	U	U	U	C	A	U	U	U	U	U	U	U	U	U	U	U	U	A	G	rp28A
U	U	U	U	C	U	U	G	U	C	U	U	A	A	U	C	A	C	A	G	RP28B
A	A	A	A	A	C	G	U	G	G	A	U	U	A	A	U	A	U	A	G	rp29
U	U	U	U	G	U	A	U	U	G	C	U	U	U	U	A	A	U	A	G	rp51A
G	A	U	U	A	U	U	G	C	U	A	U	U	U	U	U	A	U	A	G	rp51B
G	A	U	U	U	A	C	U	A	U	U	U	C	C	A	U	U	U	A	G	rp59
U	A	U	C	G	U	U	U	A	C	A	U	U	U	C	A	A	C	A	G	rp73
A	A	U	G	G	U	A	U	U	A	U	U	U	A	U	A	A	C	A	G	S10A
C	U	G	U	U	G	A	A	A	A	U	U	U	A	A	A	A	U	A	G	S10B
U	U	U	U	U	G	A	U	U	U	C	U	C	U	U	U	A	C	A	G	TUB1
A	U	U	U	U	U	C	U	U	U	U	U	U	A	C	A	A	C	A	G	TUB3
-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	position
4	6	4	4	4	6	6	3	3	4	4	0	2	6	5	8	13	0	19	0	A (6.2)
2	3	3	2	5	2	1	2	2	3	0	0	0	0	3	0	1	0	0	19	G (3.3)
1	1	1	3	2	2	2	1	2	2	2	0	3	2	3	2	1	8	0	0	C (3.0)
12	9	11	10	8	9	10	13	12	10	13	19	14	11	8	9	4	11	0	0	U (6.4)
21	32	21	21	21	32	32	16	16	21	21	0	11	32	26	42	68	0	100	0	%A (33)
11	16	16	11	26	11	5	11	11	16	0	0	0	0	16	0	5	0	0	100	%G (18)
5	5	5	16	11	11	11	5	11	11	11	0	16	11	16	11	5	42	0	0	%C (16)
63	47	58	53	42	47	53	68	63	53	68	100	74	58	42	47	21	58	0	0	%U (34)
U	U	U	U	U	U	U	U	U	U	U	U	U	U	N	U/A	A	U/C	A	G	Consensus



positions -5 to -15 displaying 70-93% occurrence of pyrimidines, with U being the preferred one (Mount, 1982). Both yeast and plant introns show an even greater preference for uridine; indeed, at least in the case of yeast, the proportion of cytosines is close to that expected if there were no specific enrichment for them. Though the net pyrimidine enrichment in both yeast and plant introns is less than that observed in mammals, the prevalence of uridine residues in yeast 3'L introns, particularly from positions -7 to -13, is striking (each position being from 58 to 100% U for 19 introns.

On the basis of the above observations we conclude that there are two distinct types of 3' splice site geometries in fungal introns. One class, which we designate 3'S, contains introns in which the branch point is quite close to the 3' splice site (5-15 nucleotides) and which display no apparent conserved sequence element 3' of the branch point other than a significantly greater branch point to 3' splice site spacing, clustering about 35-45 nucleotides. In these introns we observe both an enrichment for uridines in positions -13 to -7 and a preferred 3' splice site sequence of (U or A)APyAG. These introns appear to be analogous to introns identified thus far in mammalian and plant systems.

B. Implications for 3' Splice Site Variations

The observation that there are two different types of 3' splice sites in fungal introns raises three important questions. (1) Are these introns spliced by the same splicing machinery? (2) How is the 3' splice site, in particular the AG dinucleotide, recognized in each

case? (3) What is the function of the conserved polyuridine sequence found in 3'L introns? In the following sections we address these questions.

1. Splicing of Both Types of Introns Involves Common Factors

Several observations from S. cerevisiae suggest that, in general, the same splicing machinery interacts with both 3'S and 3'L introns. Several genes that have been identified as being required for the splicing of 3'L introns, are also required for the splicing of 3'S introns. These include both protein components of the spliceosome (the RNA2 gene product)(Miller, 1984; Lustig et al., 1986), and an snRNA (snR7) required for splicing in vivo (Patterson and Guthrie, 1987). In addition, 3'S introns contain all known splicing signals other than the polyuridine tract. Furthermore, a deletion of the actin intron (3'L type) which brings the AG within 7 nucleotides of the TACTAAC box (thus creating a type of 3'S intron) is spliced only marginally less efficiently than wild type (Fouser and Griesen, 1987). From these arguments, we conclude that 3'L introns represent a slightly more complex intron than the 3'S type.

2. Recognition of the 3' Splice Site

How is the 3' splice site recognized in these two classes of introns? To answer this question, it is useful to review the process of 3' splice site recognition in mammalian introns, which are similar to 3'L introns. In this case, the recognition of the 3' splice site

appears to be initiated by the binding of a factor, probably the U5 snRNP, to the polypyrimidine stretch (Frendewey and Keller, 1985; Ruskin and Green, 1985; Chabot et al., 1985). It is important to note that the actual interaction with the substrate appears to be mediated by a protein component of the U5 snRNP (Tazi et al., 1986; Gerke and Steitz, 1986). This is followed by the binding of the U2 snRNP to the branchpoint in a manner that may require both U2-U5 interactions as well as U2-intron interactions. In contrast, in yeast the first detectable intermediate in the assembly pathway involves the binding of the yeast U2 snRNP (Ares, 1986), snR20 (also called LSR1), to the branchpoint in a manner dependent on recognition of the 5' splice junction (Pikielny et al., 1986). This observation suggests that in yeast the TACTAAC box is the primary signal in 3' intron recognition.

Though the order of snRNP addition may be different, we suggest that recognition of the 3' splice site in yeast 3'L introns occurs in a manner analogous to mammalian introns. We have recently identified a yeast snRNA, snR7, with striking homology to mammalian U5 (Patterson and Guthrie, 1987). We suggest that the snR7 snRNP is involved in the recognition of the polyuridine tract and 3' splice junction in yeast in a manner analogous to the interaction of the U5 snRNP with the 3' splice site in mammalian introns. Furthermore, the preferred TACTAAC-to-AG spacing we observe in yeast (about 40 nucleotides) is strikingly reminiscent of that seen for branchpoint-to-AG spacing in mammals, raising the possibility that in yeast the U2 (snR20) and U5 (snR7) snRNPs involved in the recognition

of sequences in the 3' portion of the intron interact as they are thought to do in mammals.

In this light it is interesting to consider the recognition of the 3' splice site in 3'S introns. In these introns, the AG dinucleotide must be recognized in the absence of a polyuridine tract. This is a notable problem given that the U5-associated protein which binds the polypyrimidine tract is dependent on the presence of the AG for efficient binding (Tazi et al, 1986; Gerke and Steitz, 1986). If recognition of the polyuridine-AG sequence in yeast is analogous to that in mammals, then recognition of the AG in 3'S introns presumptively occurs in one of the following ways: (1) The protein that recognizes the AG in 3'L introns also recognizes it in 3'S introns, but in a manner independent of the polyuridine tract; or (2) Alternatively, recognition of the AG in 3'S introns occurs by a mechanism independent of the protein-mediated U5 snRNP-intron interaction identified to date. Note that this recognition could be mediated by an RNA-RNA interaction. While such a recognition process could be unique to 3'S introns, it seems more likely that it would also occur in 3'L intrhelpimabugons, concurrent with or subsequent to, the recognition of the polyuridine tract. In either case, because 3'L introns require additional interactions, we predict that there should be specific mutations that inhibit the splicing of 3'L introns but do not affect the splicing of 3'S introns.

3. Function of the Polyuridine Tract

As we have just argued, it seems likely that the polyuridine tract is recognized by the yeast equivalent of U5, snR7. Is there any evidence which indicates that the polyuridine tract plays a role similar to the mammalian polypyrimidine tract? Several reports have described the deletion of sequences between the TACTAAC box and the AG. In general, the conclusion is that these sequences are not required for the first (Rymond and Rosbash, 1985; Fouser, and Friesen, 1987) or second step (Fouser and Friesen, 1987) in the splicing process. However, in that deletions in this region would, in many cases, convert a 3'L intron into a 3'S intron such experiments do not address the role of these sequences in their original configuration. Similarly in many instances, the deletions reported either fail to remove the conserved sequences, or replace then with sequences similar to the consensus. This leads us to conclude that, while these experiments demonstrate significant differences between splicing in yeast and in mammals [where the deletion of the normally used polypyrimidine stretch completely blocks splicing (Fukamaki et al., 1982; Ruskin and Green, 1985)], the role of the conserved nucleotides near the 3' splice site in 3'L introns has not been tested rigorously and requires further experiments.

Given that 3'S introns are spliced without the contribution of a polyuridine stretch, its presence in 3'L introns must be explained. We suggest that its role is to adapt the greater spacing seen in 3'L introns such that it resembles that seen in 3'S introns. In other words, since 3'S introns are spliced efficiently with a branch point

to AG spacing of about 10 nucleotides, this may represent a preferred or optimal spatial configuration for splicing. Recognition of the polyuridine tract in 3'L introns may allow a trans-acting factor to bind, perhpas snR7, and through interactions with other snRNPs, position the AG dinucleotide into the proper geometry for cleavage and ligation.

Is there any evidence that the polyuridine tract plays a role in the positioning of the AG dinucleotide? The only data we are aware of is indirect. The observation that a mutation of the AG dinucleotide in the actin intron to AC (C303/305, Vijayraghavan et al., 1986), is still spliced, though inefficiently, to the proper site, suggests to us that the position to the 3' splice site is marked in a manner independent of the AG dinucleotide. We believe it is unlikely that the observed splicing represents a low-level ability of the splicing machinery to recognize the dinucleotide AC, because (1) a mutation that alters the AG to GG appears to show the same behavior (Fouser and Friesen, 1987) and (2) in many yeast introns, an AC immediately precedes the AG, thus erroneous splicing to this dinucleotide would result in non-functional proteins.

Further evidence that a polyuridine-AG motif might provide a positioning function comes from the comparison of two introns with similar branchpoint-to=AG spacings but different nucleotide compositions. In the case of the TUB3 intron, the branch point-to-AG spacing is 137 nucleotides and the region upstream of the AG is an excellent match to the consensus shown in Fig. @. By Northern analysis, no precursor or lariat is detected in vivo (Schatz et al., 1986). We compare this to a synthetic intron constructed by the

insertion of pBR322 sequences just 3' of the TACTAAC box in the yeast actin intron, generating a TACTAAC-to-AG spacing of 120 nucleotides. Here, the sequence upstream of the AG is devoid of uridines in the -7 to -13 region. In vivo, the primary product of this gene is the lariat intermediate (Cellini et al., 1986). Although other explanations are also possible, this observation is consistent with the polyuridine tract playing a role in positioning the 3' splice site over large distances.

In summary, we propose that the function of the polyuridine tract is to interact with a component of the splicing machinery, probably the snR7 snRNP. Whether or not this interaction contributes to the efficiency of the assembly of the spliceosome requires further experiments. In addition, we suggest the possibility that the polyuridine tract functions to position the 3' splice junction in the appropriate geometry for cleavage and ligation. A prediction of this model is that mutation of the polyuridine tract in an intron in which the AG has been previously mutated, for example, the C303/305 mutation discussed above, should prevent the low-level accurate splicing which still occurs in this mutant.

III. BRANCHSITE-5' SPLICE JUNCTION RELATIONSHIP

A second feature of the intron that would be anticipated to affect the efficiency of the splicing process is the spatial relationship of the 5' splice junction to the branch site. While an optimal relationship may in itself provide certain aspects of the proper intron geometry for splicing, it is more likely that conserved

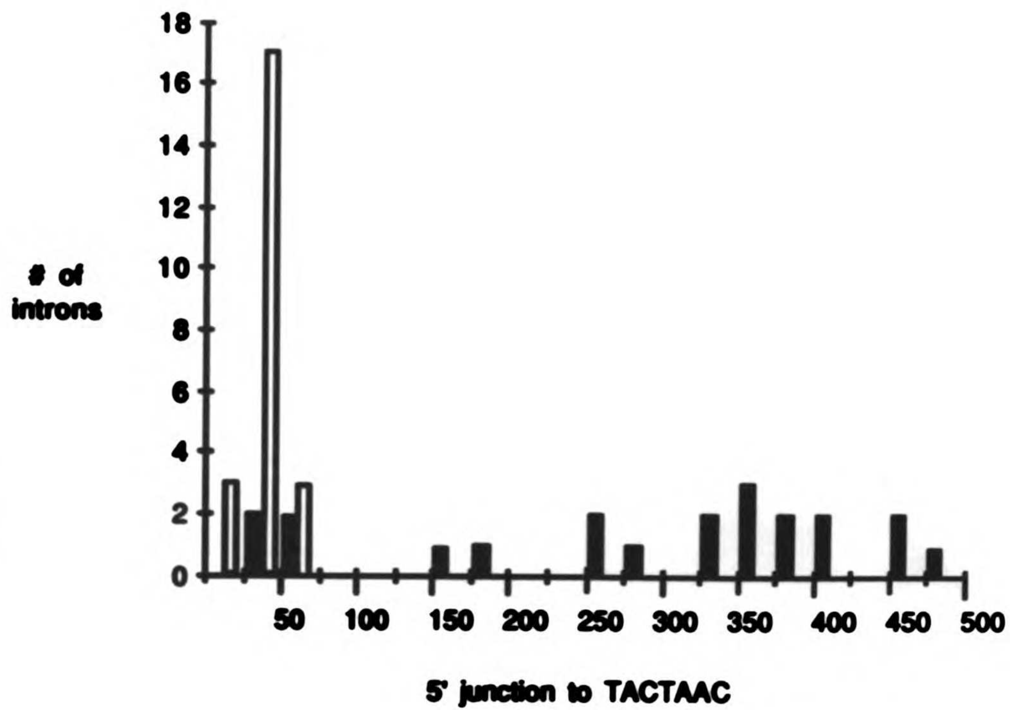
relationships between these intron sequences represent features that promote interactions between the U1 snRNP bound at the 5' splice site and the U2 snRNP bound at the branch site. Interaction between these two snRNPs has been proposed previously (Black et al., 1985), and the coimmunoprecipitation of U1 snRNPs and U2 snRNPs by monospecific antisera directed against either of these snRNPs is suggestive of their direct interactions (Mattaj et al., 1986)k.

As described above, a functional yeast analog of U2, snR20 (also called LSR1), has been identified previously. Pikielny and his co-workers (1986) have observed that the first detectable intermediate in the assembly pathway for a yeast intron involves the binding of this U2 snRNP to the intron in a manner dependent on the presence of the 5' splice junction. This suggests that interactions between the U1 snRNP and the U2 snRNP may initiate, and therefore dominate, the assembly process in yeast. (Although no U1 analog has yet been identified in Saccharomyces cerevisiae, the strong similarity of the splicing process in many aspects, including the presence of a conserved sequence at the 5' junction similar to that which interacts with the U1 snRNP in mammalian introns, argues strongly that one exists and will soon be described). Since the initial events that can be detected in the assembly of mammalian spliceosomes occur at the 3' splice site, interactions between U1 and U2 snRNPs may be more important to the assembly process in yeast and other fungi as compared to mammals. This suggests that features of the intron that influence interactions between these two snRNPs may be more highly conserved than in mammalian introns.

As shown in Fig. 3, we have examined the relationship between the 5' splice junction and branch site and again have found two distinct distance relationships. In all introns from fungi other than S. cerevisiae and four of the S. cerevisiae introns, the spacing between these two elements is quite small (about 40 nucleotides). We refer to introns containing this spacing between the 5' splice site and the branchpoint as being a 5'S intron. Is seems likely that this small spacing could facilitate direct interaction between the U1 and U2 snRNPs. Ths size of the region bound by each of these snRNPs in mammalian extracts (approximately 15 nucleotides for U1 (Mount et al., 1983), about 35 nucleotides for U2 (Black et al., 1985)) is consistent with this notion. In contrast, in the remainder of S. cerevisiae introns the spacing is significantly larger (200-450 nucleotides). We refer to introns containing this spacing between the 5' splice site and the branchpoint as being a 5'L intron.

How do the U1 and U2 snRNPs interact in introns with a large spacing between the 5' junction and branchsite? One possibility is that secondary structure features of the intron could bring the 5' splice junction and the branchsite closer together in three-dimensional space. Interestingly, we have noted that most of the S. cerevisiae introns with a large spacing between the 5' junction and the branch site can be folded into secondary structures which position the 5' junction and the branchsite in a similar linear arrangement to that of the small introns just discussed. Some of these structures are shown in Fig. 4 and the remainder are listed in Table I. A common feature of these structures is that, while the exact position of the helix varies, the number of nucleotides between

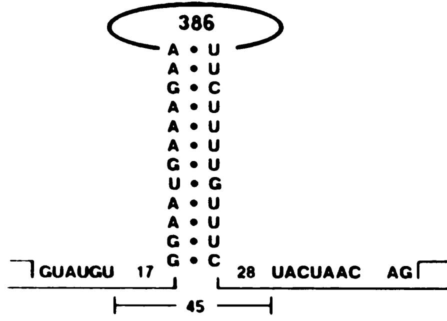
Figure III. Analysis of 5' junction to branch point spacings in fungal introns. Introns are grouped by the number of nucleotides between the branchpoint and 5' junction, excluding the conserved nucleotides that make up each signal (6 nucleotides at the 5' junction and the 7 nucleotides equivalent to the TACTAAC box at the branchpoint). The number of introns within a given spacing interval are shown for _S_. _cerevisiae_ (solid bars) as well as other fungi (open bars).



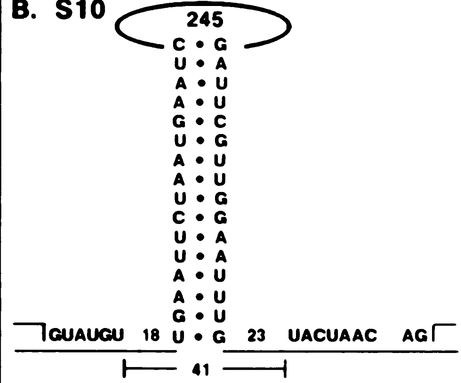
18
16
14
12
10
8
6
4
2
0
of
introns
50
100
150
200
250
300
350
400
450
500
5' junction to TACTAAC

Figure IV. Possible helical structures in Saccharomyces cerevisiae introns. The figure shows possible secondary structures for four introns from S. cerevisiae. Exon sequences are shown as open boxes, intron sequences as a solid line. The conserved sequences found at the 5' splice junction, the branch site, and the AG at the 3' splice junction are also shown. Nucleotide distances between the 5' junction sequence and the base of the helix are shown with the first base counted as the base following the 3' uridine of the 5' consensus. The distance between the helix and the TACTAAC box is counted from the first unpaired nucleotide 3' of the helix to the nucleotide preceding the first U of the UACUAAC sequence. The distance between the 5' junction and the branchpoint across the helix is the sum of these two. The number of nucleotides in the large loop is shown above each helix. The S10 intron shown is from the S10A gene.

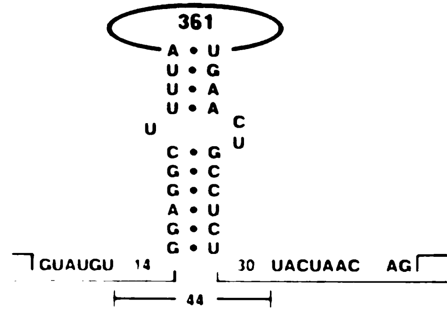
A. cyh2



B. S10



C. rp51A



D. Actin

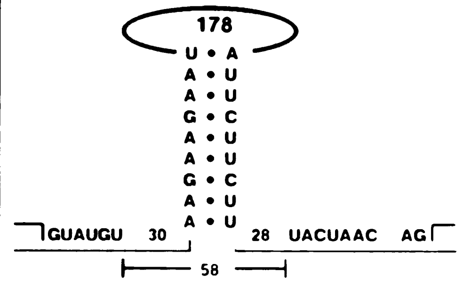


Table I. Possible helical structures in *Saccharomyces cerevisiae* introns. The table shows the 5' and 3' portions of large helical structures formed within some *S. cerevisiae* introns. In each case, the nucleotides forming the stem are shown in a 5'->3' direction. Nucleotides which are bulged are underlined. No strong helical region could be found for the rp28A intron which gave a similar spacing to the others. The introns from the L34, S20B, and rp29 genese were not examined for this structure. The spacing across the helix is calculated as described in the legend to Fig. 4.

TABLE 1 Possible Secondary Structures in *cerevisiae* Introns

Intron	5' Portion	3' Portion	Spacing Across Helix
L17A	UGGUUAAAUUGA	UCAAGAUUAACCA	58
L25	AGCACGAGAAAAUUGAGAGGAAGA	UUUUCUUCUAUUUGGUGCU	44
rp16a	AGUAAUUGGUAAUGCAAAGU	GCUUUGUAUAUCAUUAUU	48
rp16B	UAGAGAAUGAAUGGAU	AUUCCAUAUUCUAUUUUUA	47
rp28A	Not Found (See Footnote to Table)		
rp28B	AAGAGCUAU	AUAGGACUUCUU	46
rp51B	GGAUGAUGAUGGA	UCUAUCUUAUCC	45
rp59	ACAUAGUGAACAUUUUU	AAAGAUACUAUGU	35
rp73	UAUUUCGCUUUAUUCUUUCAUGGGA	UCUCCAUGACUGUGAAAUAAAAGUGAGAUA	40

the 5' splice site and the TACTAAC box is relatively constant (after substracting the nucleotides in the stem and loop), around 45 nucleotides (see Table I), and is approximately the same as in the small introns.

Are there any data that suggest these helices may actually play a role in the splicing of the large yeast introns? First, it is worthwhile pointing out that the trans-splicing experiments of Solnick (1985) and Konarska and her colleagues (1985) demonstrate that the splicing machinery has the ability to splice across such helical structures. In addition, the observation that the conserved sequences required for splicing in yeast can be found in other transcripts which are not spliced (Parker, 1985) suggest that there are additional features to yeast introns that have not yet been identified. Similarly, the observation that yeast does not in general use cryptic 5' splice junctions cannot be due solely to the strong sequence conservation of yeast splicing signals. The yeast splicing machinery can clearly use a mutant 5' splice site in the right place (note the use of many mutant junctions (Parker and Guthrie, 1985; Fouser and Friesen, 1986; Vijayraghavan et al., 1986; Jacquier et al., 1985)), though a perfect 5' consensus sequence is not used when inserted a significant distance into the intron, even in the absence of the normal 5' junction (A. Newman, personal communication). This argues that the optimal position of the 5' junction is marked in some way.

In addition, two experimental observations are consistent with these helices playing a significant, though perhaps nonessential role in splicing. As noted by Pikielny and Rosbash (1985), there is

asmall region near the 5' splice site in the intron of the rp51A gene which is required fro optimal splicing. The region they defined includes tha majority of the helix we predict for the rp51A intron (the bottom 6 base pairs of the helix shown in Fig. 4). Similarly, when small deletions are made which remove the 3' portion of this stem in both the actin and cyh2 introns, splicing is inhibited both in vivo and in vitro (A. Newman, personal communication). Interestingly, at least in some cases, the splicing of these introns can be restored by much larger deletions, consistent with a role of the helix in bringing the 5' junction and the TACTAAC box into close proximity (A. Newman, personal communication).

On the basis of the above observations, we can make a model in which the interaction between U1 at the 5' junction and U2 at the branchsite is optimized in yeast introns by one of two ways. First, the distance between the two signals may be such that it promotes this interaction directly as in the small introns. Alternatively, larger introns may fold into secondary structures which bring these two elements into closer proximity.

IV. PERSPECTIVES

To this point, we have discussed two different types of 3' intron organization as well as two distinct manners in which the 5' splice junction and the branchpoint are related to each other. Interestingly, the majority of introns with a short 5' splice site to branch point distance (5'S) also have a short branch point to 3' splice site spacing (3'S). Similarly, intorns with a large spacing

between the 5' junction and the branch site (5'L) always contain a splice site distinguished by a large branch site to 3' splice site distance and a conserved polyuridine tract (3'L). Though the significance of this nonrandom pairing of organizational features of the intron is not understood, the observation that mixed introns exist either naturally (for example, the KIN28 intron is both a 5'S and a 3'L intron) or can be created (as discussed in Section III.B) suggest that these different types of intron organization are not absolutely required together.

This reases the interesting possibility that the different types of introns we see reflect the evolution of pre-mRNA introns. The simplest, and perhaps progenitor, pre-mRNA intron is represented by the fungal introns distinguished by a short spacing between the 5' junction and the branch point as well as a 3'S splice site. By our models, this type of intron is optimized bof snRNP-snRNP interactions and as such requires nont of the additional features we have discussed. The introns found in the MATa1 gene in S. cerevisiae are a good example of this calss of intron.

If our view of 5'S and 3'S introns as being optimized for spliceosome assembly is correct, why have more complex introns evolved and been maintained? One possibility is that the additional requirements for the splicing of more complex introns represent opportunities for the regulation of gene expression. For example, in 5'L introns factors that destabilize the helix (perhaps by stabilizing a mutually exclusive structure) would be expected to decrease the efficiency of splicing. This notion provides a basis for why mammalizn genomes, with their greater requirement for

regulatory networks, encode, in general, introns of the more complex type.

REFERENCES

Abovich, N., and Rosbash, M. 1984. Mol. Cell. Biol. 4:1871-76.

Ares, M., Jr. 1986. Cellu 47:49-59.

Bertrand, H., Bridge, P., Collings, R.A., Garriga, G., and Lambowitz, A.M. 1982. Cell 29:517-26.

Black, D.L., Chabot, B., and Steitz, J.A. 1985. Cell 42:737-50.

Boel, E., Hansen, M.T., Hjort, I., Hoegh, I., and Fiil, N.P. 1984. EMBO J. 3:1581-85.

Brown, J.W.S., Feix, G., and Frendewey, D. 1986. EMBO J. 5:2749-58.

Cech, T.R. 1986. Cell 44:207-10.

Cellini, A., Felder, E., and Rossi, J.J. 1986. EMBO J. 5:1023-30.

Chabot, B., Black, D.L., Lemaster, D.M., and Steitz, J.A. 1985. Science 230:1344-49.

Dona, J.J.M., Mulder, G.H., Rouwendal, G.J.A., Springer, J., Bremer, W., and Wessels, J.G.H. 1984. EMBO J. 3:2201-06.

Fouser, L.A., and Friesen, J.D. 1986. Cell 45:81-93.

Fouser, L.A., and Friesen, J.D. 1987. Mol. Cell. Biol. 7:225-30.

Frendewey, D., and Keller, W. 1985. Cell 42:355-67.

Fukamaki, Y., Ghosh, P.K., Benz, E.J., Reddy, V.B., Lebowitz, P., Forget, B.G., and Sherman, M.W. 1982. Cell:585-93.

Gerke, V., and Steitz, J.A. 1986. Cell 47:973-84.

Green, M. 1986. Ann. Rev. Genet. 20:671-708.

Guthrie, C., Reidel, N., Parker, R., Swerdlow, H.,and Patterson, B. 1986. UCLA Symp. Mol. Cell. Biol. L33:301-21.

Hiraoka, Y., Toda, T., and Yanagida, M. 1984. Cell 39:349-58.

Jacquier, A., Rodriguez, J.R., and Rosbash, M. 1985. Cell 43:423-30.

Kaufer, N.F., Fried, H.M., Schwindinger, W.R., Jasin, M., and Warner, J.R. 1983. Nucl. Acids Res. 11:3123-35.

Kinnaird, J.H., and Fincham, J.R.S. 1983. Gene26:253-60.

Konarska, M.M., Padgett, R.A., and Sharp, P.A. 1985. Cell 42:165-71.

Langford, C.J., and Gallwitz, D. 1983. Cell 33:519-27.

Leer, R.J., van Raamsdonk-Duin, M.M.C., Molenaar, C.M.T., Cohen, L.H., Mager, W.H., and Planta, R.J. 1982. Nucl. Acids Res. 10:5869-78.

Leer, R.J., van Raamsdonk-Duin, M.M.C., Hagendoorn, M.J.M, Mager, W.H., and Planta, R.J. 1984. Nucl. Acids Res. 12:6685-700.

Leer, R.J., van Raamsdonk-Duin, M.C., Molenaar, C.M.T., Witsenboer, H.M.A., Mager, W.H., and Planta, R.J. 1985. Nucl. Acids Res. 13:5027-39.

Lustig, A.J., Lin, R.-J., and Abelson, J. 1986. Cell 47:953-63.

McGraw, P., and Tzagoloff, A. 1983. J. Biol. Chem. 258:9459-68.

Maniatis, T., and Reed, R. 1987. Nature 325:673-78.

Mattaj, I.W., Habets, W.J., and Venrooij, W.J. 1986. EMBO J. 5:997-1002.

Miller, A.M. 1984. EMBO J. 3:1061-65.

Mitra, G., and Warner, J.A. 1984. J. Biol. Chem. 259:9218-24.

Molenaar, C.M.T., Woudt, L.P., Jansen, A.E.M., Mager, W.H., and Planta, R.J. 1984. Nucl. Acids Res. 12:7345-58.

Mount, S. 1982. Nucl. Acids Res. 10:459-72.

Mount, S.M., Pettersson, I., Hinterberger, M., Kormas, M., and Steitz, J. 1983 Cell 33:509-18.

Ng, R., and Abelson, J. 1980. Proc. Natl. Acad. Sci. USA 77:3912-16.

Parker, R. 1985. Ph.D. Thesis, University of California at San Francisco.

Parker, R., and Guthrie, C. 1985. Cell 41:107-18.

Patterson, B., and Guthrie, C. 1985. Cell 49:613-24.

Peebles, C.L., Perlman, P.S., Kecklenburg, K.L., Petrillo, M.L., Tabor, J.J., Jarrell, K.A., and Cheng, H.-L. 1986. Cell44:213-33.

Pikielny, C.W., and Rosbash, M. 1985. Cell 41:119-26.

Pikielny, C.W., Rymond, B.C., and Rosbash, M. 1986. Nature 324:341-45.

Piller, T., Lang, B.F., Steinburber, I., Vogt, B., and Kaudewitz, F. 1983. J. Biol. Chem. 258:7954-59.

Reed, R., and Maniatis, T. 1985. Cell 41:95-105.

Ruskin, R., and Green, M. 1985. Cell 41:95-105.

Ruskin, B., Greene, J.M., and Green, M.R. 1985. Cell 41:833-44.

Rymond, B.C., and Rosbash, M. 1985. Nature 317:735-37.

Schaap, P.J., Molenaar, M.T., Mager, W.H., and Planta, R.J. 1984. Curr. Genet. 9:47-54.

Schatz, P.Z., Pillus, L., Grisafi, P., Solomon, F., and Botstein, D. 1986. Mol. Cell. Biol. 6:3711-21.

Shoemaker, S., Schweikart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K., and Innis, M. 1983. Bio/Technology 1:691-96.

Simon, M., Faye, G. 1984. Proc. Natl. Acad. Sci. USA 81:8-12.

Simon, M., Seraphin, B., and Faye, G. 1986. EMBO J. 5:2697-704.

Solnick, D. 1985. Cell 42:157-64.

Taxi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. Cell 47:755-66.

Teem, J.L., and Rosbash, M. 1983. Proc. Natl. Acad. Sci. USA 80:4403-07.

Teem, J.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., van Raamsdonk-Duin, M.M., Mager, W.H., Planta, R.J., Schultz, L., Friessen, J.D., and Rosbash, M. 1984. Nucl Acids Res. 12:8295-312.

Toda, T., Adachi, Y., Hiraoka, Y., and Yanagida, M. 1984. Cell 37:233-42.

Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J., and Guthrie, C. 1986. EMBO J. 5:1683-95.

Woudt, L., Pastink, A., Kempers-Veebstra, A.E., Jansen, A.E.M., Mager, W., and Planta, R.J. 1983. Nucl. Acids Res. 11:5347-60.

CHAPTER IV: A ROLE OF POLYURIDINE STRETCH IN YEAST SPLICING IS REVEALED BY EMPLOYING A BRANCHSITE MUTATION

We have investigated the effect of sequences in the 3' splice site of a yeast intron on the efficiency of splicing. By making alterations in the region of the yeast actin intron corresponding to the distinctive polypyrimidine stretch of mammalian introns, we sought to uncover a corresponding element in a yeast intron and to investigate its role in splicing. In order to increase the sensitivity of our assays, we also introduced these changes into an intron bearing a point mutation at the branchsite. We find that by themselves, substitutions in the polypyrimidine region of the yeast actin intron have only a marginal effect on splicing. However, in conjunction with a branchsite mutation, these alterations result in significant impairment of the second step of splicing.

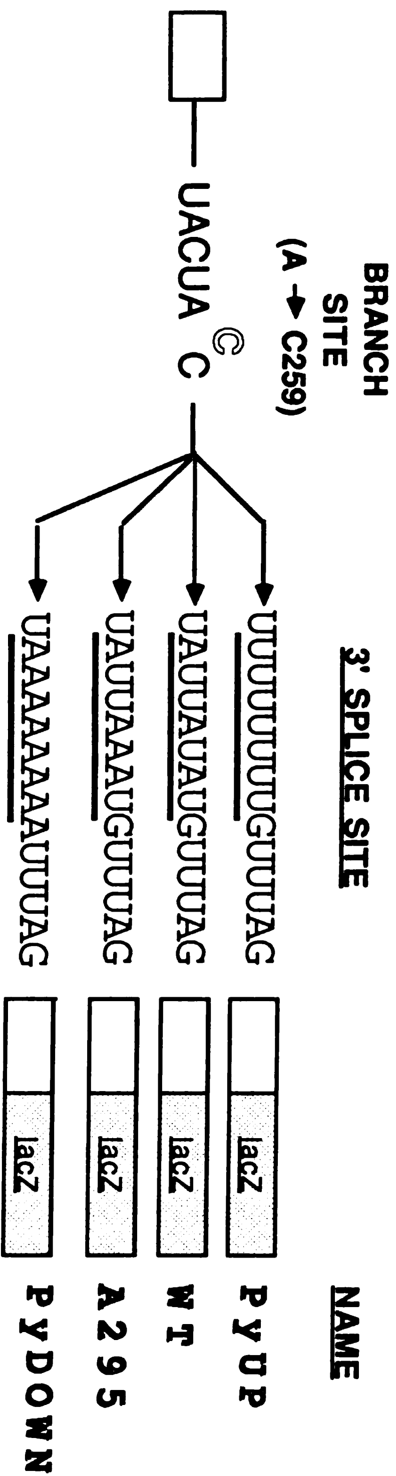
Early experiments designed to test the generality of pre-mRNA splicing mechanisms from yeast to mammals suggested that there were several differences between the two systems operated particularly at the 3' splice site: 1) a yeast intron spliced in mammalian extracts used a branchsite different from that used in yeast extracts (Ruskin et al., 1986), and 2) sequences 3' of the branchpoint are critical for all complex formation and detectable splicing in mammalian in vitro systems (Frendewey and Keller, 1985), but are wholly dispensable or sequence independent in yeast splicing extracts (Rymond et al., 1987). Although these findings were originally thought to reflect profound mechanistic differences, further characterization of the respective machineries (in particular, the

snRNAs) suggests very similar mechanics with variances in timing and stringency of requirements for the separate sequence elements.

Nonetheless, it remains clear that, at least for introns studied to date, the absolute sequence requirements at the 3' splice site are different in yeast and mammalian systems. The requirement of a polypyrimidine stretch for the earliest steps of complex formation on mammalian introns arises from the essential role of U2AF in directing the U2 snRNP to the assembling spliceosome (Ruskin and Green, 1988). Yeast splicing clearly lacks an absolute requirement for this interaction, an observation which can be taken to account for the observed lower degree of sequence conservation in the polypyrimidine region. Nonetheless, yeast introns do display significant enrichment of uridines in the area, suggesting an important role for sequences in the region of pre-mRNA splicing (Parker and Patterson, 1987).

In order to investigate the role of this sequence element in yeast introns, we enriched or depleted uridines from the region from -6 to -13 nucleotides from the 3' splice site of the yeast actin gene. These alterations were made in the context of a wild type intron as well as one bearing a branchsite C -> A transversion (the C259 mutation). This mutation was selected because it shows significant accumulation of both precursor and mature species, and a detectable amount of lariat intermediate (Vijayraghavan et al., 1985), thereby allowing us to readily detect changes in both the first and second step of splicing. Further, in our preliminary studies of several intron mutations, C259 revealed detectable phenotypes associated with for polypyrimidine alterations. The alterations we made are shown in Figure I. We designated these

Figure I. Alterations at the 3' splice site of the C259-*lac*Z fusion. The mutant branchsite is shown shadowed. Altered regions in the polyuridine stretch are underlined. Mutant 3' splice sites are designated as shown. The WT construction is from Vijayrhagavan et al. (1985), all others are the result of oligomutagenesis of this construction to yield the sequences shown. Open boxes represent actin exon sequences.


BRANCH SITE
(A → C259)
UACUA C C
3' SPLICE SITE
UUUUUUUUGUUUAG
UAUUAUAUGUUUAG
UAUUAAAUGUUUAG
UAAAAAAAAUUUAG
lacZ
lacZ
lacZ
lacZ
NAME
PYUP
WT
A295
PYDOWN
Actin Intron

changes PyUP, which alters 3 adenines to uridines in the transcript, A295, which alters the uridine at position -9 to an adenine (this position is a U in 28 of 29 sequenced yeast introns and therefore represents an almost invariant position in the yeast polyuridine stretch), and PyDOWN, which replaces 4 uridines with adenines. All mutations were introduced into an actin-lacZ fusion construct to allow analysis of products both by measuring β-galactosidase activity and by primer extension.

The results of the β-galactosidase assays are shown in Table I. For derivatives bearing the wild type branchsite, none of the alterations significantly affects production of β-galactosidase, although the two mutations which decrease the uridine content of the region under study do reproducibly yield slightly decreased activity. Thus in the context of a wild-type branchsite, the alteration of sequences in the "polyuridine stretch" only marginally affects production of mRNA.

The introduction of the C259 branchsite mutation alters the results dramatically. β-galactosidase activity resulting from the splicing of the intron with no polyuridine alterations is about 16% that of the intron with a wild type branchsite, consistent with previous observations (Vijayraghavan et al., 1985). Enrichment for uridines results in only a slight impairment of activity. However, the A295 and PyDOWN mutations now have a dramatic effect on production of β-galactosidase activity. The single nucleotide alteration results in a 2.5-fold decrease in activity, while the multiple change causes 3-fold decrease. Thus by examining polyuridine alterations in the crippled intron, we have uncovered a

CONSTRUCT	MU	%WT	%C259
WT	163	100	---
PyUP	164	101	---
A295	153	94	---
PyDOWN	157	96	---
C259	25.9	16	100
C259-PyUP	24.6	15	95
C259-A295	11.0	7	42
C259-PyDOWN	8.0	5	31

Table I. β-galactosidase assays of strains containing actin-lacZ fusions with 3' splice site alterations. Nomenclature is as described in the text. β-galactosidase assays were performed as described by Miller (1982). MU is Miller Units of β-galactosidase activity. WT is the unaltered starting fusion from Vijayraghavan et al. (1985). %WT represents the percentage activity of the fusion shown compared with the activity of the wild-type fusion. %C259 represents the percentage activity of the fusion shown of the C259 fusion with an unaltered polyuridine stretch.

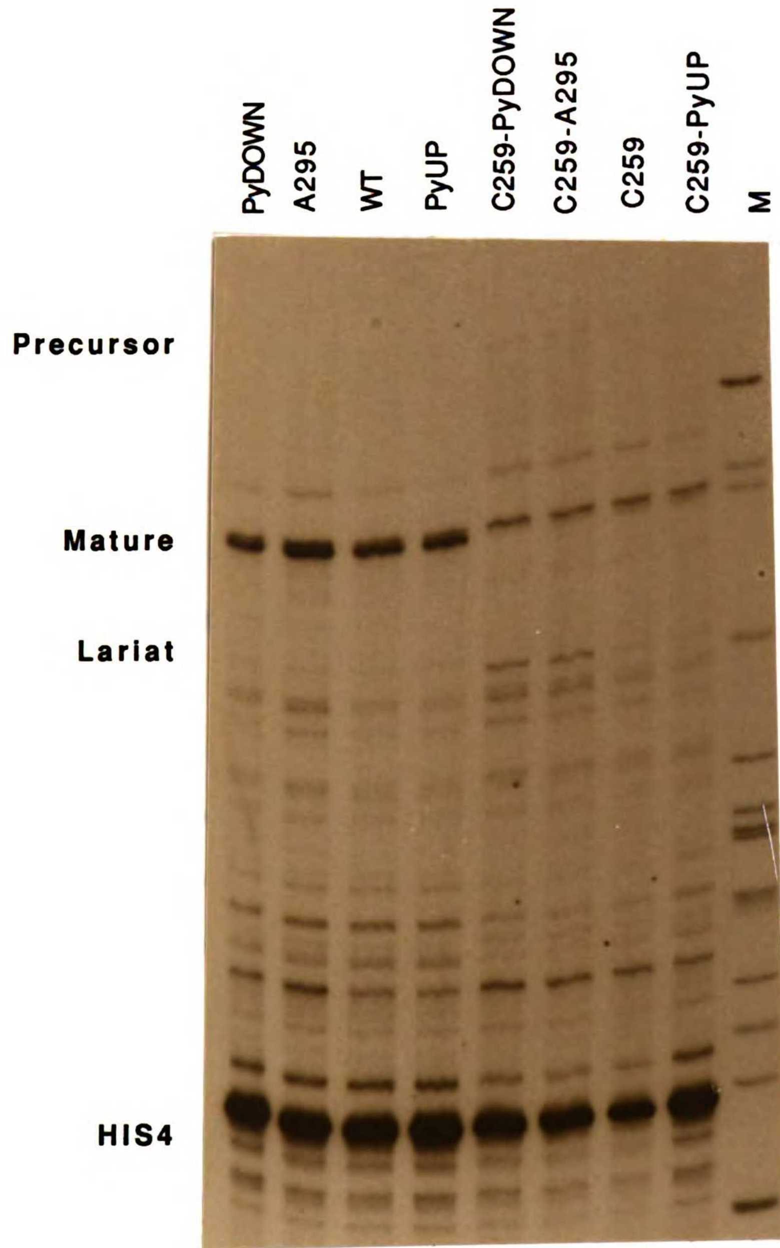
role for sequences in the polyuridine stretch, and shown that changes to adenines are deleterious while alterations to uridines have only minor effects.

We next sought to determine the specific defects caused by the depletion of uridines from the polyuridine stretch. We isolated RNA from the eight strains discussed above and analyzed it by primer extension. The results of this experiment are shown in Figure II. In the case of introns with a wild-type branchsite, we detect only mature mRNA. However, in the case of the introns bearing the C259 branchsite mutation, alterations in the polyuridine stretch dramatically alter the spectrum of products and intermediates. The C259 mutation alone causes accumulation primarily of unspliced precursor (Vijayrhagavan et al., 1985), but also produces a significant amount of mRNA. The PyDOWN and A295 mutations are sharply depleted for mRNA formation compared to wild type and instead accumulate large amounts of lariat intermediate. The PyUP mutation closely resembles wild type.

These results represent the first experimental demonstration of the functional existence of such an element in _Saccharomyces cerevisiae_. Further, we present the first evidence that the polypyrimidine stretch is required for the second step of the splicing reaction. In the context of the C259 mutation, the enrichment of uridines in the region from -6 to -12 relative to the 3' splice site contributes to the efficiency of conversion of the lariat intermediate to mRNA. The importance of the polyuridine stretch in the late steps of the splicing reaction accounts for the conservation of this element in introns in which it is not required

for complex formation and the first step of splicing. The ability to perform the early steps of the splicing reaction in introns with 3' splice site mutations affords a unique opportunity to dissect the roles of this element in late splicing events.

Figure II. Splicing phenotypes of actin-*lacZ* fusions with 3' splice site alterations. All constructions are as shown in Figure I. Note that the mature product is the band just below a non-specific primer extension product (and is not the primary band in the region in the C259-containing fusions). Positions of precursor, lariat intermediate and mature RNA are as shown. The primer anneals to *HIS4* sequences in the fusions as well as to endogenous *HIS4* transcripts (labelled as *HIS4*). RNA extraction and primer extension analysis are as described in Chapter IV.


PyDOWN
A295
WT
PyUP
C259-PyDOWN
C259-A295
C259
C259-PyUP
M
Precursor
Mature
Lariat
HIS4

REFERENCES

Frendewey, D. and Keller, W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42:355-67.

Parker, R. and Patterson, B. 1987. Architecture of fungal introns: implications for spliceosome assembly. In Molecular Biology of RNA: New Perspectives, M. Inouye and B.S. Dudock, Eds. Academic Press, New York. pp.133-49.

Ruskin, B., Zamore, P.D. and Green, M.R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52:207-19.

Ruskin, B., Pikielny, C.W., Rosbash, M. and Green, M.R. 1986. Alternative branch points are selected during splicing of a yeast pre-mRNA in mammalian and yeast extracts. Proc. Natl. Acad. Sci. USA 83:2022-26.

Rymond, B., Torrey, D. and Rosbash, M. 1987. A novel role for the 3' region of introns in pre-mRNA splicing of Saccharomyces cerevisiae. Genes and Develop. 1:238-46.

Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J. and Guthrie, C. 1985. EMBO J. 5:1683-95.

CHAPTER V: 3' SPLICE SITE REQUIREMENTS IN YEAST

The first clue to the mechanism of intron recognition and removal during nuclear pre-mRNA splicing came from the discovery of consensus signals at the 5' and 3' ends of introns (Breathnach et al., 1978). Subsequently, some of the roles of these sequences have begun to be illuminated: the 5' splice site is recognized by the U1 snRNP (Mount et al., 1983) and the branchpoint by the U2 snRNP (Parker, Siliciano and Guthrie, 1987). The role(s) of the third conserved signal, the polypyrimidine stretch, has been more difficult to unambiguously assign.

In mammalian systems, the polypyrimidine stretch is an easily discerned enrichment of pyrimidines in the 15 nucleotides preceding the PyAG which demarcates the 3' splice site (Mount, 1982). This element is essential for even the earliest detectable complex formation (Frendewey and Keller, 1985; Ruskin and Green, 1985; Bindereif and Green, 1987). One basis for this requirement has been elucidated by Ruskin et al. (1988), who found that the mammalian polypyrimidine stretch is the binding site for a protein factor, U2AF, which is required for binding of the U2 snRNP to the substrate.

U2AF is not, however, the only mammalian splicing factor which interacts with the polypyrimidine stretch. A second protein, Intron Binding Protein or IBP, is also capable of specifically binding the polypyrimidine stretch (Tazi et al., 1986, Gerke and Steitz, 1986). IBP has been shown to be a component of the U5 snRNP, thus implicating U5 in the recognition of the polypyrimidine stretch. Unlike the case with U2AF, the timing and importance of the

U5-polypyrimidine stretch interaction is completely obscure. Clearly, the possibility that the IBP-mediated recognition of the polypyrimidine stretch is related to late steps in the splicing reaction must be entertained.

A second observation which has been used to argue that yeast introns do not have anelement corresponding to the polypyrimidine stretch lies in the structure of the two introns of the MATa1 transcript (Miller, 1984). These 3'S introns clearly lack a direct analog of the polypyrimidine stretch in that the branchpoint-to-3'-splice-site spacing is 10 nucleotides in one and 11 in the other. Not only is this spacing inconsistent with the encoding of a polypyrimidine element, there is no significant enrichment for uridines. There also exist several synthetic yeast introns in which a similar branchpoint to 3' splice site spacing has been created by recombinant techniques (Fouser and Friesen, 1987; Langford and Gallwitz, 1985). In all cases the AG dinucleotides are utilized (albeit with varying efficiencies). This observation has to the so-called "first AG" rule: that in yeast introns a 3' splice site is selected by scanning from the branchpoint to the first AG dinucleotide, which is then utilized by the splicing machinery.

We have instead proposed that short branchpoint to 3' splice site motif (which predominates in other fungi) actually distinguishes a second class of introns, which we designate 3'S introns (for Short 3' splice site spacing. These are contrasted with 3'L introns (Long 3' splice site spacing) which contain polyuridine elements preceding the 3' splice site and have distant (averaging 40 nucleotides from branchpoint to PyAG) 3' splice sites. We suggest that the ability of

yeast to utilize a closely spaced AG arises from a machinery predisposed use such an AG rather than sole reliance upon a simple scan for PyAG sequences (Parker and Patterson, 1987).

In Saccharomyces cerevisiae, however, our knowledge concerning a polypyrimidine-like element in 3'L intronsis largely anecdotal. The lack of a clear enrichment of pyrimidines of the degree seen in mammalian introns has lead to the general impression that there is no corresponding element in yeast introns. Nonetheless, a careful examination of sequenced yeast introns reveals significant conservation of pyrimidines, specifically of uridines, in the region from 5 to 18 nucleotides 5' of the 3' splice site Parker and Patterson, 1987). The -9 position is most striking, where 24 of 25 yeast intron sequences encode a uridine (Parker and Patterson, 1987; Schneider and Guarente, 1987; Magdolen et al., 1988) (the single exception is the COX5 gene (Cumsky et al., 1987)). Thus there is a clear preference for specific sequences in the region of yeast introns corresponding to the mammalian polypyrimidine stretch.

The final aspect of yeast introns which serves to distinguish them from their mammalian counterparts is that they do not require sequences 3' of the branchsite for initiation of spliceosome formation. Although in mammalian systems the polypyrimidine stretch is absolutely required for any stable complex formation (Frendewey and Keller, 1985; Ruskin and Green, 1985; Bindereif and Green, 1987), in yeast the first step of splicing (formation of the lariat intermediate) can be performed in vitro on transcripts which contain any of the four ribohomopolymers downstream of the branchsite (Rymond et al., 1987). Indeed, stable complex formation is observed in vitro

on transcripts that terminate 5 nucleotides 3' of the branchpoint (Rymond et al., 1987). These experiments demonstrate that there is no absolute requirement for specific sequences in the polypyrimidine region for the first step of splicing, but leave open the question of why yeast introns show sequence specificity in this region. Of particular interest is the possible contribution of these sequences to late steps of the splicing reaction, a possibility which has not been addressed to date.

We reasoned that the above mentioned aspects of yeast 3' sequence requirements represent a unique opportunity to study the role of the polypyrimidine stretch in late steps of the splicing reaction. Since sequences 3' of the branchsite are not required for initiation of the splicing reaction, we can observe their importance for subsequent events. Given the lack of strong conservation of the polypyrimidine element in yeast, we chose to establish a competitive system in which two alternative 3' splice sites (PyAG elements) are present. By varying the sequences which precede these signals, we sought to demonstrate the importance of an element analogous to the mammalian polypyrimidine stretch in yeast, and to gain insights into its importance relative to other intron/exon signals at the 3' splice site.

MATERIALS AND METHODS

Plasmid Construction: We started with the plasmid B102 (pBR322 with the SalI-AvaI fragment replaced with the *CYC1* terminator, regenerating the SalI site (B. Osborne, pers. comm.)), which was

opened at the PvuII site, into which we inserted a blunt-ended 1400 nte. fragment encoding the yeast TRP1 gene and ARS1. The resulting plasmid was opened at the unique ClaI site, which was filled in and re-ligated, resulting in disruption of the site. This plasmid was cleaved at the unique BamHI and SalI sites, and BamHI-SalI fragments bearing the actin intron constructions (see Oligo Mutagenesis) were inserted. In cases where oligonucleotides were to be inserted, the two complementary oligos were annealed by heating to 90° for 3 minutes followed by slow cooling to 0°. The relevant plasmids (from the BamHI-SalI insertion step above) were opened at the ClaI site (introduced in the actin fragment) and the annealed oligos were ligated in. After heat inactivating the ligase, the ligation mixtures were digested with ClaI to prevent recovery of the reclosed vector. Orientation of the inserted material was determined by Taql digestion, which allowed us to determine which of the two Taql fragments meeting at the former ClaI site had been regenerated, and which had been extended by the length of the oligonucleotide. These plasmids were then cut with SmaI and SalI, and the SmaI-SalI fragment of pMC1871 (Casadaban et al., 1983), which contains the lacZ coding region, was inserted. Nomenclature of the final products is as follows: p3SS designates the series of final products, the sequence at the branchpoint-proximal splice site is indicated by +0 for no insert at the Cla site, +A or +T for the 12 nte insertions, and +AA or +TT for insertion of the 22 nte. insertions. The sequence at the distal site is designated PyDOWN, A295, PyUP, WT or C303/305 in keeping with the nomenclature we employ for the sequence changes at

these sites (see below). Thus the plasmid encoding no insert at the proximal site and PyUP at the distal site is designated p3SS+0-PyUP.

Oligo Mutagenesis

The variations in the distal (wild type) 3' splice site were introduced by oligomutagenesis of mACT, an M13 mp18 derivative carrying a 1419 nte. BamHI-SalI fragment which contains the actin upstream control region, 5' exon, intron, and part of the 3' exon (Vijayrhagavan et al., 1986). In all cases the dut- ung- method as described in Siliciano and Guthrie (1988) was employed to obtain mutants. The original substrate was first altered by introducing an adenine 9 ntes. 3' of the branchsite using the ACT3'S oligo to create m3'S. This was followed by the introduction of PvuII and SmaI restriction sites 20 ntes. into the actin second exon to produce m3'S-A2L. Finally, we created the changes in the distal 3' splice site by using oligos that annealed to the polyuridine region and which created the variants we designate PyDOWN, A295, PyUP and C303/305 (see text). This series of phage were designated m3'S-A2L-PyDOWN etc. The oligos we employed for these changes are shown below.

Oligonucleotides used in this study:

ACT3'S: GAATGAAGCTAATCGATGT

A2L: ATACCAGAACCCGGGACAGCTGATAACCAAAG

PyDOWN: ACCTCTAAATTTTTTTTATAGCAACA

A295:   TAAACATTTAATATAGC

PyUP: CCTCTAAACAAAAAAAATAGCAACAA

C303/305: GCAGCAACGTGTAAACATATAA

Act+A: CGAAACAACAAA

Act+T: CGTTTGTTGTTT

Act+AA: CGAAAACAAAACAACAAACAAA

Act+TT: CGTTTGTTTGTTGTTTTGTTTT

pPY.pri2: ACCCGGGACAGCTGATAACCAAAG

RNA Preparation and Analysis

RNA was prepared essentially as described in Wise et al. (1983) except that two cold phenol/chloroform extractions were performed after the hot phenol extraction and a second ethanol precipitation was performed prior to using the RNA.

Primer extensions were performed as follows: For the annealing reaction, 50 μgs of RNA and 2 ng of kinased oligo (pPY.pri2) in a total volume of 15 μl. were added to 3μl of annealing buffer (1.5M. KCl, 0.12M. Tris-Hcl pH 8.0, 1.2mM EDTA). The samples were heated at 90° C. for 3 minutes followed by incubation at 60° C. for 45 minutes. The samples were then cooled by spinning 10 seconds in a microfuge, and 50 μl of a stock reverse transcriptase solution (70 mM Tris-Hcl pH 8.3, 14 mM $MgCl_2$, 0.7 mM EDTA, 0.7 mM each dNTP, and 7mM EDTA) added followed by addition of 0.5 ul of reverse transcriptase. The reaction mixture was incubated at 45° for 45 minutes. Reactions were stopped by addition of 2 ul of (2.5 M NaOH, 25mM EDTA) and heating to 90° C. for 3 minutes. Stopped reactions were then spun through 1ml

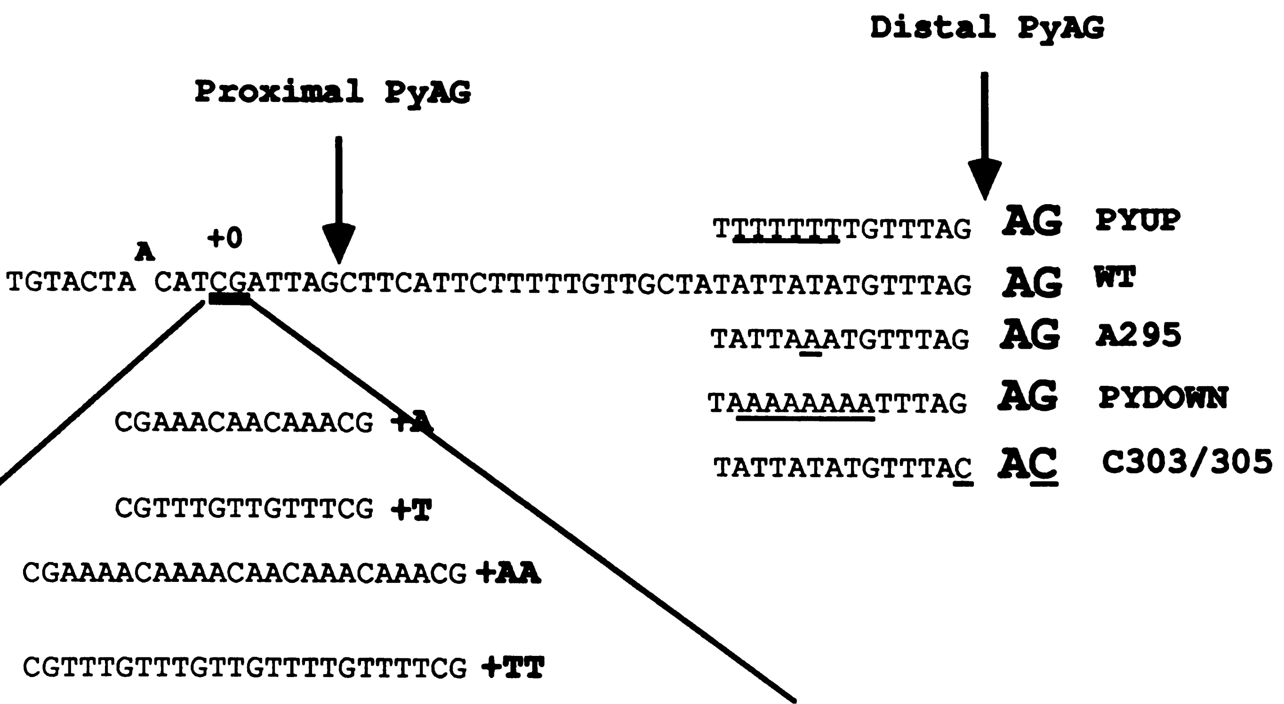
G50 spin columns, dried and resuspended in 50% formamide. Samples were heat denatured at 90° C. for 3 minutes, chilled on ice and loaded onto a 6%, 7M urea polyacrylamide gel.

RESULTS

CONSTRUCTION of 3' SPLICE SITE COMPETITION SYSTEM

We selected the yeast actin intron as the starting substrate for our analysis for several reasons. First, it is the best characterized intron in *Saccharomyces cerevisiae*. Further, deletion analysis by Fouser and Friesen has demonstrated that splicing of this intron is not impaired by alterations in the intron sequences preceding the 3' splice site, and that this intron can function when the wild-type 3' splice site is positioned only 9 ntes. from the branchpoint (Fouser and Friesen, 1987). We created a similarly spaced 3' splice site by introducing a single nucleotide, an A, at a position 9 ntes. 3' of the branchpoint (see Figure I). This novel sequence is 34 nucleotides 5' of the wild type actin 3' splice site. We chose this position for several reasons: (1) the resulting 3' splice site mimics the branchpoint-to-AG distance found in the *MAT*a1 introns (Miller, 1984), (2) this site lies just distal to a ClaI restriction site, allowing us to conveniently introduce sequences just upstream of the novel AG in order to examine the effect of sequence and spacing alterations on use of this site, and (3) splicing to this site results in an alteration of reading frame compared to the use of wild type 3' splice site.

Figure I. Alterations in the 3' splice site of the yeast actin gene. The central transcript is the +0-WT construction described in the text. The branchsite is the reaised A in the TACTAAC sequence. The sequences of the insertions made at the proximal 3' splice site are as shown (+A, +T, +AA, and +TT). The dark bar shows the sequences replaced by the insertions. Alterations introduced at the 3' splice site are also shown (PyDOWN, A295, PyUP and C303/305). Regions changed at the distal 3' splice site are underlined. Arrows mark the alternate 3' splice junctions. Exon sequences are shown in larger type. Other details of the plasmids are described in the Materials and Methods.


Proximal PyAG
Distal PyAG
A
+0
TGTACTA CATCGATTAGCTTCATTCTTTTTGTTGCTATATTATATGTTTAG AG WT
TTTTTTTTGTTTAG AG PYUP
TATTAAATGTTTAG AG A295
TAAAAAAAATTTAG AG PYDOWN
TATTATATGTTTAC AC C303/305
CGAAACAACAAACG +A
CGTTTGTTGTTTCG +T
CGAAAACAAAACAACAAACAAACG +AA
CGTTTGTTTGTTGTTTTGTTTTCG +TT

We also made several alterations in the distal (wild type) 3' splice site in order to vary its pyrimidine content: we changed T295 to A295, thus altering the most highly conserved position (9 nucleotides 5' of the 3' splice junction) of the yeast polypyrimidine stretch. We also changed 5 positions in the region from 6 to 12 ntes. 5' of the 3' splice site from thymidines to adenines in order to create a pyrimidine-poor element which we call PyDOWN. Finally, we changed three positions from adenines to thymidines to create a pyrimidine rich element which we call PyUP. We also inactivated the wild type 3' splice site by incorporating the C303/305 mutation, which converts the wild-type (branchpoint-distal) AG dinucleotide to and AC. This mutation blocks the second step of splicing of the actin intron both in vivo and in vitro (Vijayrhagavan et al., 1986). These splice sites generated are shown in Figure I.

In order to conveniently assay the products of the competition, we introduced several changes into exon II of the actin gene. These created two restriction sites which would allow us to make in-frame fusions to the E. coli lacZ gene. Initially, we made fusions in frame with both 3' splice sites, but technical problems with the fusion in frame with the novel 3' splice site prevented them from providing useful information.

We also wanted to investigate the ability of various sequences and spacings to influence the utilization of the proximal site. This was done by synthesizing complementary oligos which introduced 12 or 22 nucleotide inserts just 5' of the novel (branchpoint-proximal) 3' splice site. Depending on orientation, these introduced either uridine-rich (+T, +TT) or adenine-rich (+A, +AA) tracts upstream of

the novel 3' splice site. In sum, we produced 5 variants at the novel 3' splice site (no insert, +A, +T, +AA and +TT). These constructs are shown in Figure I. There are 25 combinations of the proximal and distal splice site variants. All 25 possible combinations were analyzed by primer extension using a primer (pPY.pri2) which anneals to the 3' exon detects all products and intermediates of the splicing of these transcripts.

SEQUENCE DETERMINANTS OF 3' SPLICE SITE SELECTION IN YEAST

In all cases, except the +TT series of constructs, we observe some utilization of each of the two alternative 3' splice sites as well as variable amounts of precursor and lariat intermediate. We will discuss our results in three groups, based on the alterations made at the novel site: no insert, +A and +T, and +AA and +TT.

No Insert

The creation of an alternative 3' splice site at a position just distal to the branchsite in the actin intron poses several interesting questions. First, what is the preference of the yeast splicing machinery when confronted with both a 3'S and 3'L splice site in the same intron? The results of Fouser and Friesen demonstrate that the actin intron is efficiently spliced when the wild type 3' splice site is brought into close juxtaposition with the branchsite (Fouser and Friesen, 1987). Therefore the general structure of the actin intron is compatible with the utilization of a 3' splice site of this type. Second, is the utilization of a

proximal site of this type influenced by the polypyrimidine element at the distal site or is the competition taking place on some other basis (i.e. spacing or exon sequences)? Finally, is this site used exclusively when the distal 3' splice site is inactivated by an AG to AC alteration?

Our results are as shown in Figure II and can be summarized as follows: (1) the branchpoint-proximal site competes successfully, albeit at a low level, with the branchpoint-distal site. (2) the competition is influenced only marginally by sequences in the polypyrimidine region of the distal site, and (3) the C303/305 alteration causes exclusive use of the proximal site. This latter result is important in that it demonstrates that the proximal site is capable of functioning efficiently as a 3' splice site, even though it is a poor competitor.

+A, +T

The failure of the proximal 3' splice site to compete efficiently with the distal site led us to ask whether changing the spacing branchpoint-to-3'-splice-site of this splice site alone could influence its utilization and whether spacing and provision of a polypyrimidine stretch could be influential. To ask this question, we introduced either an adenine-rich or a uridine-rich tract to precede the novel PyAG. The resulting branchpoint-to-3'-splice-site spacing for the proximal element is identical to that observed in the yeast S10A intron (Leer et al., 1982).

Our results are shown in Figure III and can be summarized as follows: (1) Improving the branchpoint to AG distance dramatically

Figure II. Splicing of the +0 series. All transcripts have no insertions at the proximal 3' splice site (see Figure I). The sequences at the 3' splice site are indicated over the lanes. Mature-P represents mRNA produced by splicing to the branchpoint-proximal 3' splice site, and Mature-D indicates mRNA produced by splicing to the branchpoint-distal 3' splice site. Positions of the lariat intermediate (Lariat) and precursor (Precursor) species are marked. M labels the lane containing markers produced by cleaving pBR325 cleaved HpaII.

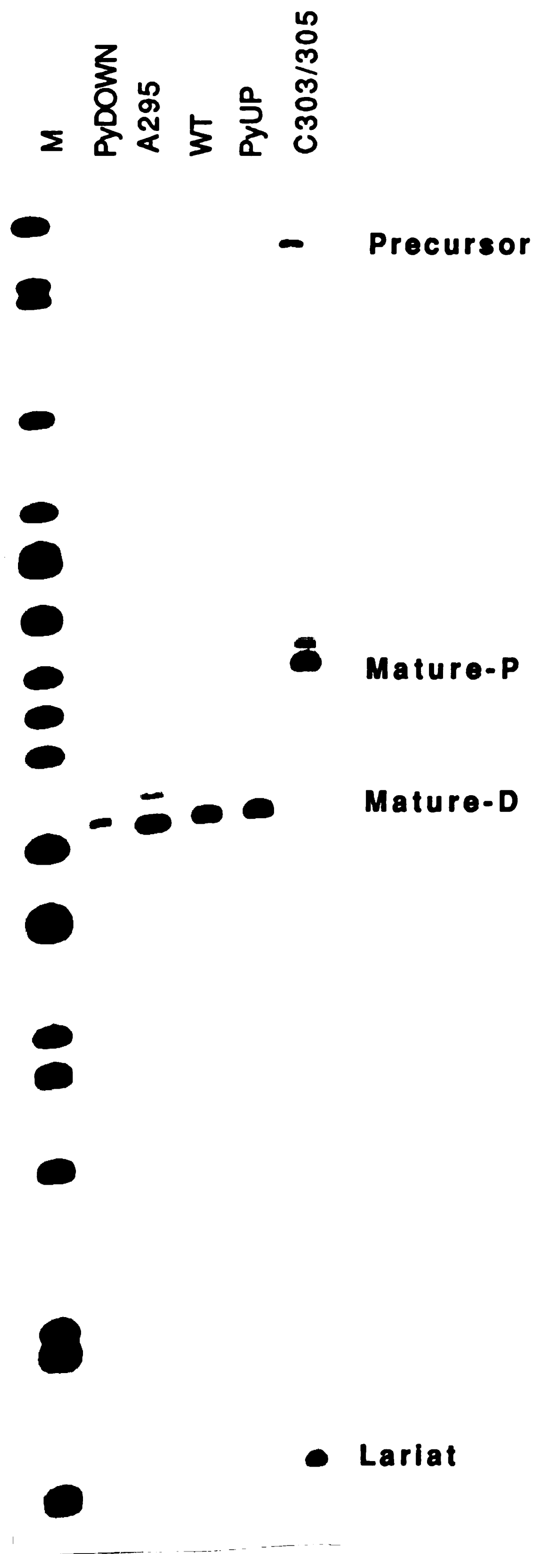
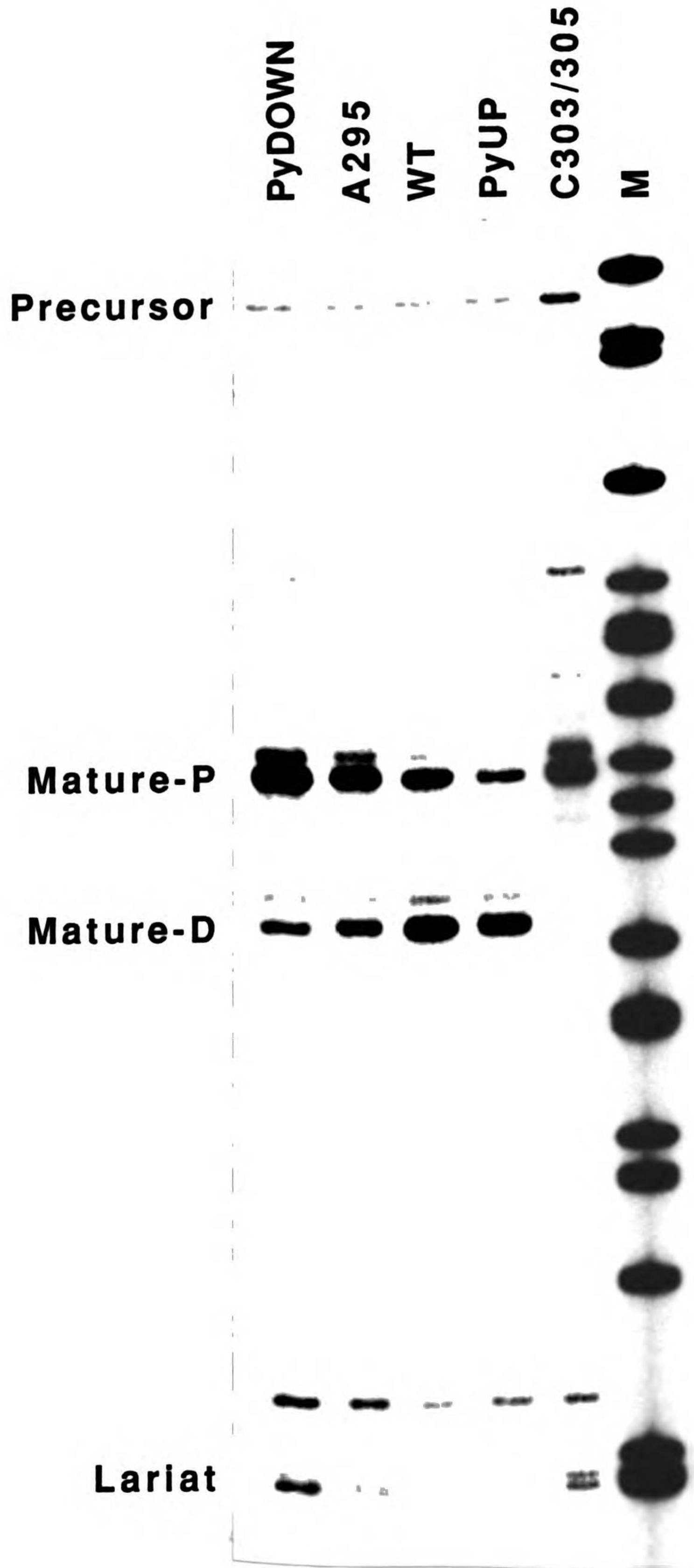
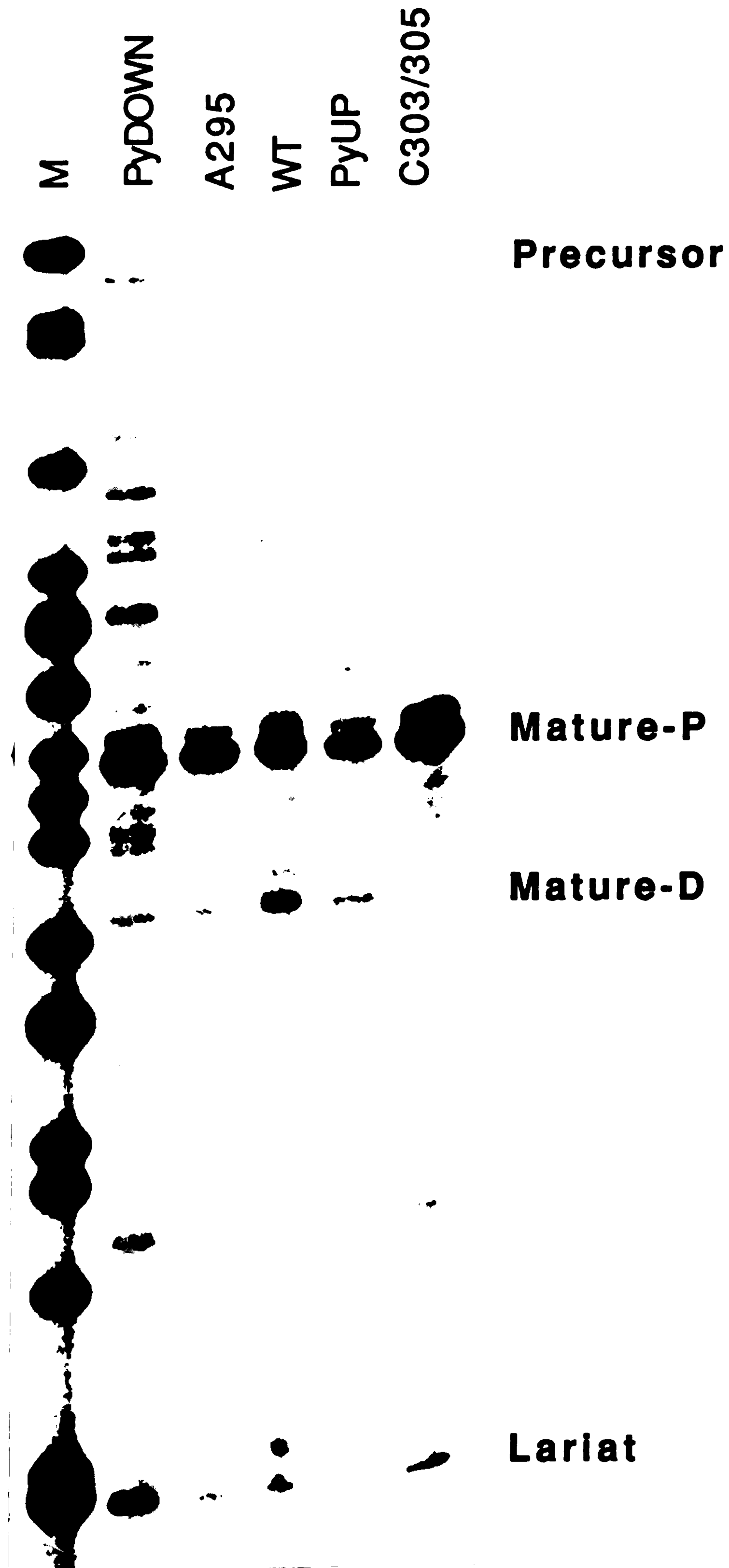

M
PyDOWN
A295
WT
PyUP
C303/305
Precursor
Mature-P
Mature-D
Lariat

Figure III. Splicing of the +A and +T series. Transcripts have insertions preceding the branchpoint-proximal 3' splice site. A) insertion of a 12 nucleotide adenine-rich sequence. B) insertion of a 12 nucleotide uridine-rich sequence. All other conventions are described in Figure II.


A.
PyDOWN
A295
WT
PyUP
C303/305
M
Precursor
Mature-P
Mature-D
Lariat


B.
M
PyDOWN
A295
WT
PyUP
C303/305
Precursor
Mature-P
Mature-D
Lariat

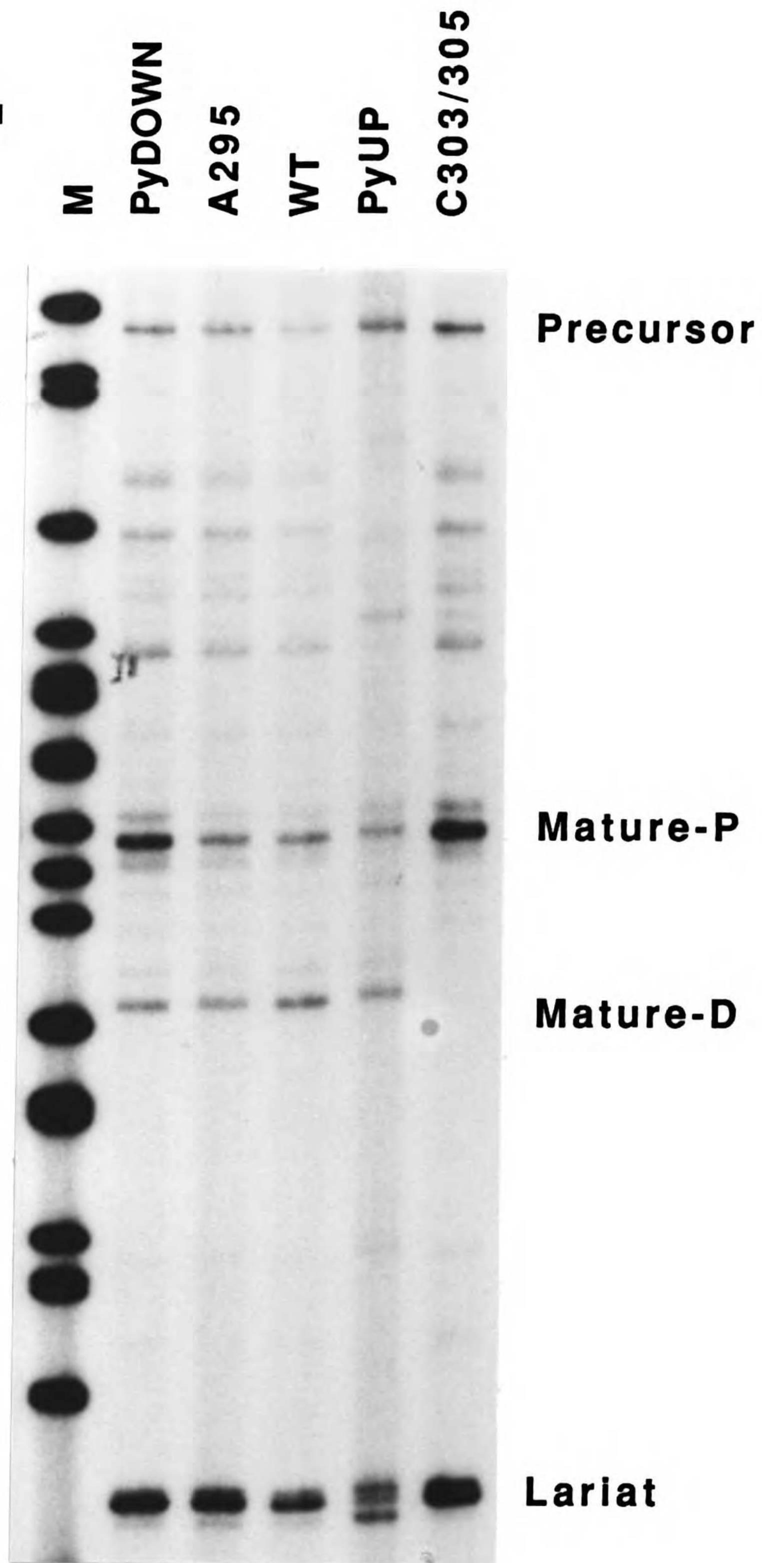
improves the utilization of the proximal site, (2) The ability of the distal sequence to compete is strongly influenced by its pyrimidine content. This is particularly clear in the +A series, where we observe roughly constant amounts of precursor and lariat intermediate, while the ratio of the two mature messages changes to reflect the pyrimidine richness of the distal site. (3) The proximal site is a markedly better competitor when preceded by a pyrimidine rich rather than a purine rich element, and (4) The proximal element is exclusively used when the distal site is inactivated, regardless of the sequence preceding it.

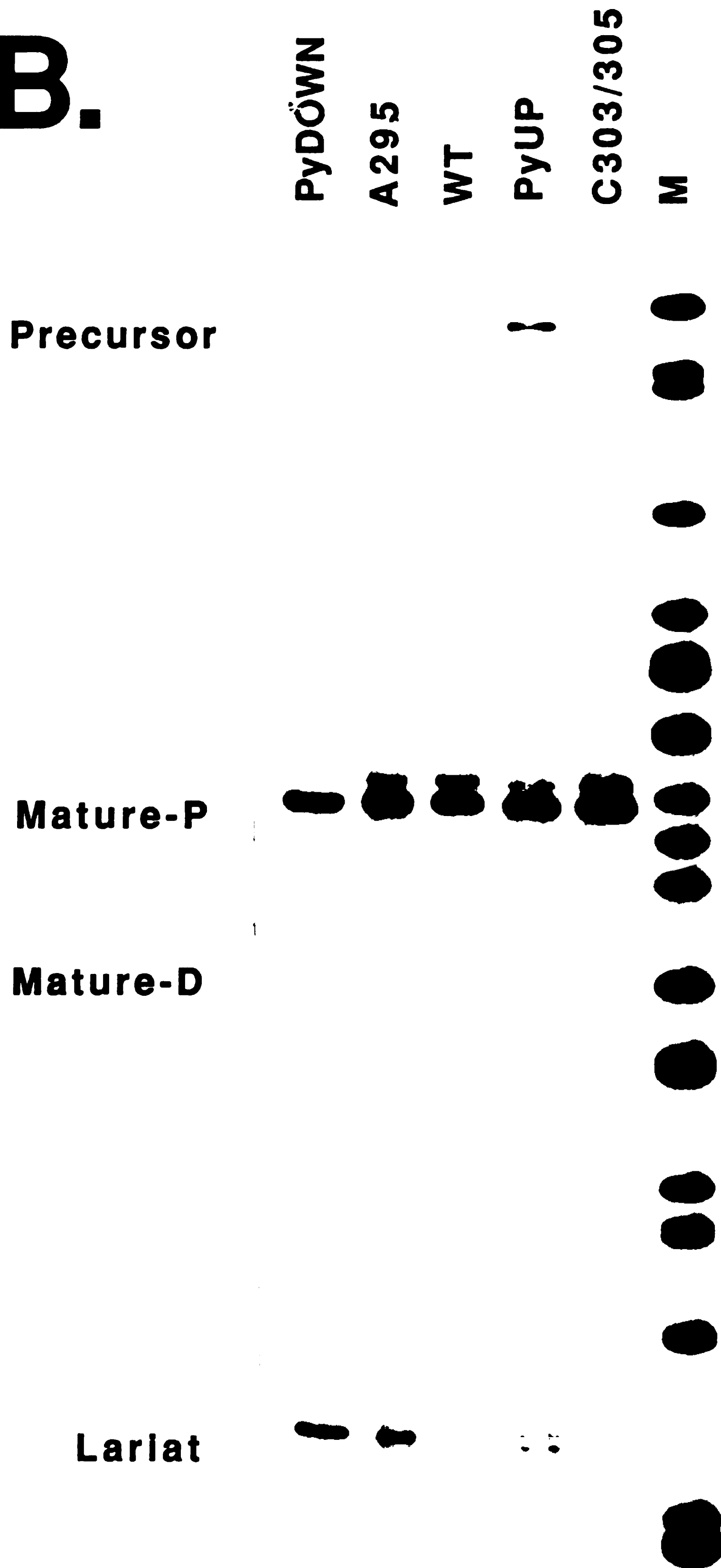
+AA, +TT

The fact that the distal site was still being utilized in the presence of the pyrimidine-enhanced proximal site led us to examine the effect of even larger insertions. In this case, we introduced 22 nucleotides between the branchpoint and the proximal AG, resulting in a spacing resembling that seen in the majority of yeast introns. The branchpoint to 3' splice site spacing for the distal site becomes 66 nucleotides, slightly larger than that seen in the yeast rp51A intron (58 ntes.) but below that of the PFY1 gene (72 ntes.)(Magdolen et al., 1988) and still well below that observed in the exceptional TUB3 gene (143 nucleotides) (Schatz et al., 1986).

Our results are shown in Figure IV and can be summarized as follows: (1) Increased spacing alone (i.e., introduction of the +AA sequence) does not markedly improve the ability of the proximal site to compete (as compared with the A series), (2) The predominant species in all of the +AA series insertions is the lariat

Figure IV. Splicing of the +AA and +TT series. Transcripts have insertions preceding the branchpoint-proximal 3' splice site. A) insertion of a 22 nucleotide adenine-rich sequence. B) insertion of a 22 nucleotide uridine-rich sequence. All other conventions are described in Figure II.


A.
M
PyDOWN
A295
WT
PyUP
C303/305
Precursor
Mature-P
Mature-D
Lariat


B.
PyDOWN
A295
WT
PyUP
C303/305
M
Precursor
Mature-P
Mature-D
Lariat

intermediate, (3) Insertion of the +TT series completely blocks utilization of the distal splice site, regardless of the sequence at the distal site, and (4) In the context of the large, uridine rich proximal insertions, increasing the strength of the distal site increases the relative proportion of precursor (compare the +TT-PyUP construct with the rest of the +TT serires) suggesting that two complete 3' splice sites in the vicinity of the branchpoint can interfere with the early steps of splicing.

DISCUSSION

We have established a 3' splice site cis competition assay in the yeast actin intron. By introducing a single nucleotide insertion in the actin intron, we have created the minimally defined 3' splice site in yeast, a PyAG motif. This 3' splice site is closely juxtaposed to the branchpoint, a configuration which is found in the yeast MATa1 introns (Miller, 1984) as well as a variety of synthetic introns (Fouser and Friesen, 1987; Langford and Gallwitz, 1985). The "first AG rule" predicts that this novel AG should be the primary 3' splice site in this intron on the basis of its proximity to the branchpoint (Langford and Gallwitz, 1985).

In sharp contrast to this prediction, a PyAG motif introduced just distal to the branchsite in the actin intron is utilized only at a low level. The predominant 3' splice site selected in this construction is the wild type site 34 nucleotides downstream (+0-WT construct). We note that this is the first instance in which the context of the distal 3' splice site is that of an actual yeast

splice junction rather than a random or exonic sequence. Nevertheless, the branchpoint-proximal site represents an actual 3' splice site because when the distal (wild type) 3' splice site is inactivated (+0-C303/305 construct), this proxiaml site is efficiently utilized (although there is a significant accumulation of precursor and lariat intermediate).

In order to investigate the parameters of 3' splice site selection in yeast, we studied variants of the wild type 3' splice site in the presence of this novel upstream site. In general, sequence alterations preceding the wild type AG do not markedly increase the ability of the proximal site to compete, although a drastic change of uridines in the region 6 to 12 ntes. 5' of the 3' splice junction to adenines (PyDOWN mutant) does result in slightly increased use of the proximal site. Thus we conclude that the distal site is selected by some active mechanism, and that the "first AG rule" is an artifact arising from the nature of the sequences which have been examined to date.

In order to further characterize the determinants of 3' splice site selection, we introduced alterations in the spacing and sequence of the proximal element to determine what changes were required to activate it. The first change we introduced was simply to move the proximal 3' splice site further 3' from the branchpoint, to determine the spacing constraints on its usage. We introduced an A-rich sequence because yeast intron sequences are typically adenine- and uridine-rich (Parker and Patterson, 1987), and we considered adenines to be a "neutral" choice given the uridine-rich nature of yeast 3' splice sites. By introducing this sequence, we found that the close

juxtaposition of the novel site with the branchpoint is a major a factor in its underutilization. Equally striking was the finding that alterations in the uridine content of the distal site strongly influenced the competition. When the distal (wild-type) site contained primarily uridines in the -5 to -15 region (PyUP mutation) it was the favored site, whereas when it consisted primarily of adenines (PyDOWN), it was a poor competitor. These results allowed us to identify two determinants of 3' splice site choice: spacing (distances greater than 10 ntes. from the branchpoint to the 3' splice site are preferred) and sequence (the region preceding the 3' splice site should be uridine rather than adenine rich). The latter finding was extended when we introduced uridines rather than adenines upstream of the novel 3' splice site. This 3' splice site configuration is the preferred site regardless of the distal sequence. Thus independent of location, a uridine rich element confers a significant advantage on a 3' splice site.

These results demonstrate that the element affecting the competition can indeed be thought of as a "poluyuridine stretch". While making alterations in an RNA molecules inherently carries the risk that alternate structures may be introduced which have an influence on the phenomenon under study, we can to some extent rule out this possibility. By examining four sequence variants at the distal (wild type) site, two of which vary by only a single nucleotide change (+A-WT and+A-A295) and observing a correlation between uridine richness and site utilization the data strongly suggest that it is the sequence of the region which causes the alteration. The fact that the same qualitative differences are seen

when a complementary proximal sequence is studied (+T series) further bolsters this argument.

Since none of the constructs mentioned above resulted in exclusive use of the proximal site, we investigated the effects of even longer insertions. Again, we introduced either uridine-encoding or adenine-encoding oligonucleotides. Insertion of the 22 nte. adenine rich sequence (+AA) did not noticably change splice site preference compared to the 12 nte. adenine-rich sequence (+A). From this we conclude that alleviating the steric constraint on the proximal site only partially alleviates its competitive disadvantage. The fact that the distal site, even when significantly uridine-poor, still competes well when distant from the branchpoint (66 ntes.) can be interpreted to mean that it retains sequences or a three dimensional configuration which allows it to be used even when another 3' splice site is more closely located to the branchsite. The involvement of exon sequences in splice site selection has been demonstrated by the experiments of Reed and Maniatis (1987), and could well play a role here.

Insertion of a 22 nte. stretch encoding primarily uridines led to exclusive use of the proximal site. Our experiments do not distinguish the contribution of three possible factors which would allow the 22 nte. U-rich sequence (+TT) to improve proximal splice site performance compared to the 12 nte. sequence (+T): 1) the increased length of the uridine stretch preceding the site, 2) the simple increase in branchpoint to AG distance of the proximal site, which now resembles the average for all yeast introns (about 40 ntes.), and 3) the movement of the distal site to a spacing

significantly beyond the average seen in yeast introns. We suggest that the result arises from near-optimal spacing of a U-rich sequence because 1) the +AA series demonstrates that spacing alone is insufficient, and 2) a 22 nte. polyuridine stretch is significantly longer than the size of the element suggested by sequence comparison and thus would not be anticipated to confer an overwhelming advantage. For these reasons, we conclude that the presence of a uridine rich site, a PyAG motif, and a location 30-50 ntes. 3' of the branchpoint represent the optimal intron 3' spice site arrangement in yeast.

The effects of the C303/305 mutation on the different proximal site configurations is particularly instructive with regard to the role of the polyuridine stretch in splicing. In the +T-C303/305 and +TT-C303/305 constructions, the proximal site is efficiently used and there is little or no accumulation of precursor or lariat intermediate. Thus these represent legitimate 3' splice sites and there is no overriding spacing constraint. However, in the corresponding +A-C303/305 and +AA-C303/305 constructions, significant amounts of lariat intermediate accumulate--indeed, in the +AA-C303/305 construction, it is the primary product. The block at lariat cannot be attributed to competition with the distal site because it has been inactivated by the point mutations, and shows no sign of competing with the +T and +TT constructions. For these reasons, we conclude that the ineffficient performance of the second step of splicing in these constructions is caused by an inability of the splicing machinery to efficiently "find" or utilize the proximal 3' splice site. Therefore, as we have suggested previously (Parker

and Patterson, 1987), one probable role for the polyuridine stretch in yeast splicing is to locate or activate the 3' splice site for the second step of the splicing reaction.

We do not observe significant effects of our alterations on the first step of splicing. While all of our constructions accumulate a low level of precursor, only two seem to enhance significantly the levels of precursor. These are the +0-C303/305 construction and the +TT-PyUP construction. The splcing of the +0-C303/305 construction is very reminiscent of that seen for the #34 mutation of Fouser and Friesen, which juxtaposes the 3' splice site of actin closely with its branchsite. In the +TT-PyUP construction, it is the presence of two strong potential 3' splice sites which appears to cause precursor accumulation. In both cases our results can be interpreted as meaning that alterations in the 3' splice site can inhibit the first step of splicing (in the former case by placing a 3' splice site adjacent to the branchsite, in the latter by assembling two complete candidate 3' splice sites), but that there is no apparent positive contribution of these sequences to the first step of splicing in vivo (hence other alterations in this region do not increase the relative levels of precursor RNA).

The demonstration that yeast expresses a preference for polyuridine sequences during 3' splice site selection suggests that there exists a machinery which recognizes this element. Since the strength of the site influences the competition, we are now in a position to identify components of this machinery by screening for mutations which enhance or interfere with the ability of a given site to function as a competitor with a second, dissimilar site.

REFERENCES

Bindereif, A. and Green, M. R. 1987. An ordered pathway of snRNP binding during mammalian splicing complex assembly. EMBO J. 6:2415-24.

Black, D. L., Chabot, B., and Steitz, J. A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell 42:737-50.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. 1980. Proc. Natl. Acad. Sci. USA 75:4853-57.

Casadaban, M., Martinez-Arias, A., Shapira, S.K. and Chou, J. 1983. B-galactosidase gene fusions for analyzing gene expression in E. coli and Yeast. In Recombinant DNA Part B. R. Wu, L. Grossman and K. Moldeur, Eds. Academic Press, New York. pp.293-308.

Cumsky, M.G., Trueblood, C.K, and Poyton, R.O. 1987. Structural analysis of two genes encoding divergent forms of yeast cytochrome c oxidase subunit V. Mol. Cell. Biol. 7:3511-19.

Fouser, L.A. and Friesen, J.D. 1987. Effects on mRNA splicing of mutations in the 3' region of the Saccharomyces cerevisiae actin intron. Mol. Cell. Biol. 7:225-30.

Frendewey, D., and Keller, W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42:355-67.

Gerke, V. and Steitz, J. A. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47:973-84.

Langford, C.J. and Gallwitz, D. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA Polymerase II transcripts. Cell 33:519-27.

Leer, R.J., van Raamsdonk-Duin, M.M.C., Molenaar, C.M.Th., Cohen, L.H., Mager, W.H. and Planta, R.J. 1985. The structure of the gene coding for phosphorylated ribosomal protein S10 in yeast. Nucl. Acids Res. 10:5869.

Magdolen V., Oechsner, U., Muller, G., and Bandlow, W. 1988. The intron-containing gene for yeast profilin (PFY) encodes a vital function. Mol. Cell. Biol. 8:5108-15.

Miller, A.M. 1984. The yeast MATa1 gene contains two introns. EMBO J. 3:1061-65.

Mount, S.M. 1982. A Catalogue of splice junction sequences. Nucl. Acids Res. 10:459-72.

Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J. A. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell 33:509-18.

Ozkaynak, E., Finlay, D., Solomon, M. and Varshavsky, A. 1987. The yeast ubiquitin genes: a faminly of natural gene fusions. EMBO J. 6:1429-39.

Parker, R., and Patterson, B. 1987. Architecture of fungal introns: implications for spliceosome assembly. In Molecular Biology of RNA: New Perspectives. M.Inouye and B. Dudock, Eds. Academic Press, New York. pp.133-49.

Reed, R. and Maniatis, T. 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell 46:681-90.

Ruskin, B., and Green, M. R. 1985. Specific and stable intron-factor interactions are established early during in vitro pre-mRNA splicing. Cell 43:131-42.

Ruskin, B., Zamore, P. D., and Green, M. R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52:207-19.

Rymond, B., Torrey, D. and Rosbash, M. 1987. A novel role for the 3' region of introns in pre-mRNA splicing of Saccharomyces cerevisiae. Genes and Develop. 1:238-46.

Schatz, P.J., Pillus, L., Grisafi, P., Solomon, F. and Botstein, D. 1986. Two functional a-tubulin genes of the yeast Saccharomyces cerevisiae encod divergent proteins. Mol. Cell. Biol 6:3711-21.

Schneider, J.C. and Guarente, L. 1987. The untranslated leader of nuclear COX4 gene of Saccharomyces cerevisiae contains an intron. Nucl. Acids Res. 15:3515-29.

Siliciano, P.G. and Guthrie, C. 1988. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes and Development 2:1258-67.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. A Protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47:755-66.

Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., et al. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J. 5:1683-95.

Wise, J. A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E. J., and Guthrie, C. 1983. Yeast contains small nuclear RNAs encoded by single copy genes. Cell 35:743-51.

CHAPTER VI: SPLICEOSOMAL snRNAs

INTRODUCTION

The process by which introns are removed from nuclear mRNA precursors shares certain fundamental properties with group II splicing reactions (14). Since the latter can occur autocatalytically in vitro, catalysis must be achieved by the formation of active sites comprised solely of RNA. Highly conserved structural elements which reside primarily within the intron have been shown to underly the geometry required for function. In contrast, the only conserved features of nuclear pre-mRNA introns are restricted to short regions at or near the splice junctions, and splicing takes place in association with an extensive trans-acting machinery dubbed the spliceosome (25,51). Work during the last several years has revealed that prominent trans-acting factors are U1, U2, U4, U5, and U6, five small nuclear RNAs (snRNAs) which exist as ribonucleoprotein complexes termed snRNPs. Presumably the low information content in cis within this class of introns is compensated by the participation of the snRNPs, which act in trans to impart the appropriate architecture for catalysis.

In principle, both RNA- and protein-based interactions contribute to communications between the snRNPs and the intron and the snRNPs with one another. In cognizance of group II reactions, however, it seems evident that nuclear pre-mRNA splicing evolved from an RNA machine. Thus, as with the recent revolution in our understanding of the ribosome, the mechanism may be most directly

appreciated by seeking to establish the RNA-mediated aspects of the reaction pathway. The focus of this review will thus be on the snRNA components of the spliceosomal snRNPs. In addition to these intellectual biases, this choice is supported by the fact that there is currently far less information available about the protein constituents of these particles. Conversely, the comparative ease with which the snRNAs can be isolated and sequenced has led to a virtual explosion of data on this topic.

While earlier studies concentrated on the analysis of evolutionarily related metazoan RNAs, recent progress has witnessed the identification of snRNAs from a far broader phylogenetic range. This has included fungal and plant representatives, as well as a sampling of snRNAs from the evolutionarily distant trypanosomes, organisms which splice by ligating exons encoded in trans to one another. The power of comparative sequence analysis to identify functionally important features of sequence and structure has been clearly demonstrated in the tour-de-force studies of rRNA pioneered by Woese, Noller and their colleagues (50). Given the current availability of snRNA sequences from diverse organisms, we have thus undertaken a consideration of primary sequence alignments and the structural information derived therefrom.

Finally, we will place considerable emphasis on the integration of data from Saccharomyces cerevisiae. Knowledge of the complete set of spliceosomal RNAs from this organism has provided several unexpected insights into the evolutionary constraints operating on the five snRNA species. Moreover, the apparent similarities and differences in splicing mechanisms between this yeast and mammals are

likely to be reflected in corresponding aspects of snRNA structure. Lastly, it now seems clear that the facile genetic techniques available in this organism will be ever more widely exploited for their uniquely informative contributions to our understanding of splicing.

PERSPECTIVES

Background

Historically, studies of snRNAs have focused on the six abundant species (U1-U6) first identified in mammalian nuclei (reviewed in 13; 60); five of these are found in the nucleoplasm, while the sixth (U3) is nucleolar. In addition to their subcellular localization, U1-U5 RNAs have in common a trimethylguanosine (TMG) cap. The cap structure of U6 is unknown; the purified RNA is not precipitable by anti-TMG antibodies (9). Rapid progress in the study of snRNAs was made possible upon the discovery that autoantibodies of the so-called Sm specificity immunoprecipitated RNP particles containing U1, U2, U4, U5 and U6 snRNAs from mammals to insects (40).

Early efforts to identify U snRNAs in S. cerevisiae resulted in several perplexing findings (reviewed in 26). First, there appeared to be many more yeast snRNA species (~ two dozen), yet even the most abundant was present at about 1/1000 the levels of metazoan snRNAs (82,62). Moreover, the RNAs were reported not to be immunoprecipitable by anti-Sm antibodies (40). This observation, together with the demonstration that five of the first six yeast snRNAs analyzed were dispensable for viability (78,76,54), led some

to conclude that there were fundamental differences between S. cerevisiae and metazoan splicing.

An alternative point of view first emerged from the demonstration that two other yeast snRNA genes (SNR7 (55) and SNR14 (71)) were essential for growth, and encoded RNAs containing a consensus sequence for the binding site of the Sm antigen (see below) which could assemble with Sm antigens following microinjection into Xenopus eggs or oocytes (61,77). Indeed, it was found subsequently that, under the appropriate conditions, these snRNAs, as well as two others, could be immunoprecipitated directly from S. cerevisiae extracts (77,62). Gene disruption experiments revealed that the latter snRNAs are also encoded by single-copy essential genes (SNR20 [LSR1] (1) and SNR19 (72,35). Furthermore, strong presumptive evidence that the yeast Sm snRNAs were involved in pre-mRNA processing was provided by the demonstration that this same subset of RNAs co-fractionated with in vitro splicing intermediates when assayed by several independent approaches (58,59,35,17).

It then remained to determine the specific relationship between the yeast Sm snRNAs and the spliceosomal snRNAs in mammals. The first unambiguous homology identified was to the 5' terminal 100 nucleotides of U2 (1). Surprisingly, however, this domain was located at the 5' end of a yeast snRNA (lsrl(1); snR20 (62) 1175 ntes. in length. To account for this six-fold size difference, and in the absence of other information to the contrary, it was suggested that the yeast molecule was a "poly-snRNA" in which the functional equivalents of U2, U4, U5, and U6 were covalently joined (1). As we will describe below, other work has since identified regions of

homologous sequence and/or structure in the three other Sm-associated yeast RNAs which allow their unambiguous assignment as the analogues of U1 (35,72), U4 (71), and U5 (55). More recently, the yeast equivalent of U6, which, like its metazoan counterpart lacks an Sm site, was identified and found to exhibit remarkable conservation in size and sequence (11).

Thus, intriguing differences in overall length and organization notwithstanding, it is now clear that S. cerevisiae has specific analogues for each of the five major spliceosomal snRNAs. These relationships are not summarized. In view of these conclusions, we suggest that the nomenclature of U1, U2, U4-6 be employed exclusively in favor of the original S. cerevisiae designations (of snR19, etc. (62). To preserve historical continuity, however, this nomenclature should be retained when referring to the yeast snRNA genes (SNR19, etc.).

snRNP Structure

While the emphasis of this article is on the RNA components of the spliceosomal snRNPs, it is important to consider what is known about the snRNP proteins insofar as they can inform us about organizing principles of these particles.

To date, each of the non-nucleolar snRNAs has been found to be associated with a common subset of seven polypeptides referred to as the Sm proteins, due to the presence of one or more antigenic epitopes of this specificity; see (43) for a recent review. The most abundant mammalian Sm snRNPs contain U1, U2, U4, U5 and U6 snRNAs. The U1, U2 and U5 RNP particles consist of a single species of RNA

complexed with the Sm antigens and, in addition, one or more proteins unique to the particular snRNP (Table 1). In contrast, the U4 and U6 snRNAs are found together in the same snRNP (28,8), and no proteins unique to this particle have been identified thus far. Moreover, as we will describe below, the U6 snRNA lacks an Sm binding site and only becomes immunoprecipitable by virtue of its association with U4.

Based on several lines of evidence, the binding site for the Sm antigen conforms to the consensus sequence AU(4-6)G (see 7,45 for review). Substitution of this sequence in human U1 (27) or U2 (46) with random sequences prevents stable Sm antigen association, as assayed by immunoprecipitability following microinjection into *Xenopus* oocytes. In contrast, deletion of stem/loop sequences 5' or 3' of this region had no apparent inhibitory effect on Sm association (see however 56). Indeed, the short consensus sequence, when embedded in a T7 construct, directs detectable assembly with Sm antigens (44). Thus the available data suggest that binding of the Sm proteins is specified by a surprisingly simple motif, a short, purine flanked pyrimidine-rich tract in a single-stranded region of RNA.

Pulse-chase experiments have suggested that the Sm proteins D, E, F, and G form a cytoplasmic 6S protein complex to which newly transported snRNA precursors initially bind (18) (however, see 21). These same four proteins remain complexed with the RNAs to yield a salt-resistant "core" structure (39) which protects an RNA fragment 15-25 nucleotides in length from nuclease digestion (41). This core structure may play an imortant role in the assembly of additional, snRNP-specific proteins since U1- and U2-specific snRNP proteins fail

to assemble in vivo on snRNAs lacking an Sm binding site (27,46,47). In vitro reconstitution studies suggest that binding of the U1-specific proteins U1-A and U1-70k requires only the first 50 nucleotides of the RNA (27). However, more sensitive assays suggest that this association is significantly stabilized by the presence of other domains, including the Sm site (27), and mapping of anit-70k antibody-protected regions of U1 in vitro also reveals interactions between the 5' stem/loop and the Sm binding domain (57). Interestingly, the binding of snRNP-specific proteins appears to be significantly weaker than that of the core proteins, in that they can be displaced by high salt (in the absence of Mg^{++}) (41) and fail to protect the RNA from digestion by various nucleases (reviewed in 12,57).

The Sm core proteins appear to be necessary, but not sufficient, for nuclear localization (44). Given the comparative simplicity of other nuclear targeting signals, however, it seems likely that these highly conserved antigens perform additional functions. If so, a key question is whether these roles are primarily structural, e.g. involving the assembly and/or stability of the snRNPs, or alternatively, are more intimately related with functions directly required for splicing. At present, there is no information as to the functions of the snRNP-specific proteins. In this regard, it will be particularly informative to establish the phylogenetic conservation of these protein binding domains in the snRNAs. As we will describe below, the preliminary evidence suggests that these regions are generally highly conserved.

snRNP Function

In both Group II and nuclear pre-RNAs, splicing occurs in a two-step reaction involving successive trans-esterifications (14,25,51). First, cleavage at the 5' splice site (/GT) is accompanied by formation of an intramolecular branch (5'pG-2'A) to the 2'-hydroxyl of an adenosine residue upstream of the 3' splice site; this generates intron-3' exon in a so-called lariat structure and free 5' exon. The second step is 3' cleavage, which is concomitant with exon ligation and intron release. As suggested in the Introduction, the execution of nuclear pre-mRNA splicing appears to be dominated by the requirement for trans-acting machinery whose major role is to fold an intron into a catalytically active configuration.

This process can be broken down into several types of recognition events: 1) interactions between signals in the substrate and trans-acting factors, and 2) the coalescence of these factor-site assemblies into a higher order structure, i.e., the mature spliceosome. In addition, there are likely to be different conformational requirments for catalysis of the two covalent steps of splicing. Finally, other processes must be necessary for the re-cycling of the particles. The message from this view of splicing is that the interactions of snRNPs with substrate and with one other comprise the essence of the splicing process. Thus the goal of this review is to draw attention to regions of the snRNAs likely to be most important for these processes based on their phylogenetic conservation. It is instructive to first briefly review what is known about such interactions.

The archetype of snRNA-mediated recognition events is the interaction between U1 and the 5' splice site. The original proposal that U1 recognizes splice junctions by direct Watson-Crick base pairing was based solely on the observed complementarity between sequences at the 5' terminus of U1 and the consensus sequence at intron/exon boundaries (64,40). This model (in a modified form) has subsequently been verified biochemically and genetically, and has served as a paradigm for the roles of snRNAs in other RNA processing reactions (for recent review, see 74). While U1 recognition of 5' splice site sequences has been demonstrated to occur independently of ATP or other snRNPs (48), it has not yet been determined whether the basepairing interaction is sufficient to direct 5' splice site identification by the U1 snRNP.

Using the strategy of compensatory nucleotide changes, base-pairing with U1 has been clearly demonstrated at intron position 1 (in yeast) (Siliciano and Guthrie, submitted for publication) and position 5 (in yeast and mammals) (83, Siliciano and Guthrie, for publication); evidence for pairing involving intron position 3 is weak, at best (in mammals) (83) and, in yeast, intron position 4 is not complementary to U1 (72,35). Thus not all positions in the 5' splice site appear to be essential for base-pairing interactions with U1. This suggests that the conservation of nucleotides in this region of the intron and the snRNA reflects roles in addition to the requirement to form an intermolecular helix.

In yeast, an snRNA:intron base pairing interaction has also been demonstrated in the recognition of branchpoint sequences by the U2 snRNP (53). In this case, the proposed interaction leaves the

branchpoint nucleotide unpaired, in accordance with the situation known to occur in self-splicing group II introns (70). In each of the two cases tested, the splicing defect caused by mutation of the invariant branchpoint signal (UACUAAC to UAAUAAC or UACAAAC) was suppressed by the predicted compensatory change within the complementary sequence in the yeast U2 snRNA (ntes. 34-39, GUAGUA to GUAUUA or GUUGUA, respectively)(53). The possibility that this interaction may be sufficient to direct 3' splice site selection in yeast is suggested by the observation that sequences 3' of the branchpoint are not required for yeast spliceosome assembly (69). However, it should be pointed out that no stable complex assembly is observed in the absence of an intact 5' splice site (10,81,58,17). Conversely, protection of the 5' junction from RNAse H-mediated cleavage requires the presence of wild-type branchpoint sequences (68). Thus there is reason to think that U1 and U2 snRNP:intron interactions are mutually interdependent even at the earliest stages in yeast spliceosome assembly.

Based on nuclease protection assays, the mammalian U2 snRNP also appears to recognize the region of the intron surrounding the branchpoint (4,16). However, the steps leading up to this event appear to be significantly different in the two systems. In mammalian introns, the branchpoint sequence is poorly conserved and does not appear to possess sufficient information to direct its own selection (25,51). On the other hand, the polypyrimidine stretch is a highly conserved splicing signal, and its presence is absolutely required for spliceosome assembly and branchpoint protection (19, 65). Furthermore, it appears to be the rule that deletion of a

mammalian branchpoint does not block splicing but rather causes utilization of a nearby adenosine residue; cryptic branch sites exhibit little apparent sequence specificity (52,66). The most convincing argument that mammalian U2 snRNP binding cannot depend solely (if at all) upon basepairing between U2 and the branchsite is the observation that RNAse H mediated cleavage of the branchpoint-pairing region of the snRNA does not prevent the assembly of the U2 snRNP onto substrate (34,20). However, it should be noted that complexes formed with this cleaved U2 molecule do not undergo branch formation or binding of the U4/U6 and U5 snRNPs, demonstrating the importance of this region of U2 at some step prior to branchpoint utilization.

A parsimonious explanation of at least some of these findings has come from the characterization of a protein factor known as U2 Auxiliary Factor (U2AF)(67). This factor is required for binding of U2 to substrate, whereas U2AF can bind independently of U2. The binding site for U2AF appears to be the polypyrimidine stretch/AG, and neither U2AF nor U2AF + U2 binding appears to require the presence of the 5' splice site. Thus for mammalian introns, branchpoint recognition appears to proceed as follows: U2AF recognizes the 3' splice site of an intron (polypyrimidine/AG) and on this basis positions U2 on substrate in such a way that U2 is now able to select a branchsite (presumptively by intermolecular base pairing interactions). It is of interest to note that stable U2 binding to substrate requires the 5' 15 nucleotides of U2 (as assayed by both U2-specific complex formation and by protection of the intron from RNAse A digestion after specific cleavage of the U2

nucleotides)(34,20). This requirement might reflect the need for the 5' cap of U2, snRNA sequences per se, or a protein factor which binds U2 in this region.

There is conflicting information as to whether the U5 snRNP plays a direct role in substrate recognition. Initial experiments suggesting that U5 snRNA is associated specifically with 3' splice site sequences (15) were followed by the finding that this interaction (assayed by RNase T1 protection) is mediated by a protein associated with the U5 snRNP (22,75). This Intron Binding Protein (IBP) can specifically bind the 3' splice site even in the absence of the U5 snRNP. This result of itself is interesting in light of the binding specificity of U2AF. The observation that the U5 snRNP joins the spliceosome after the binding of U1 and U2 (24,32,3) suggests that IBP might replace U2AF at the 3' splice site. In any case, identification of the 3' junction in mammalian introns appears to be a protein-mediated recognition event, albeit by a protein component of a conserved snRNP. By this view, one role of the U5 particle is to deliver IBP to the spliceosome. The interactions by which this protein is bound to the U5 snRNP are not known, but the dependence of the association on the magnesium concentration (22,75) suggests the possibility that the conformation of the U5 snRNA is important.

Our understanding of the snRNP-snRNP interactions which mediate the formation of higher order structure is currently quite limited. Studies of complex assembly in both yeast and mammalian systems have led to a scenario in which recognition of the 5' splice site by U1, in conjunction with recognition of the 3' splice site by U2 and

associated factors, is a prerequisite for assembly of the other snRNPs on the substrate (24,32,3,38,37,20). As measured by protection from nuclease digestion in Hela in vitro systems, this step correlates with increased strength of factor binding at the 5' splice site, and expanded protection of the intron at both the 5' and 3' splice sites (34,67,4,16). How the successful recognition of 5' and 3' splice sites is signalled is unknown, but it is obviously appealing to contemplate a model in which U1 bound to the 5' splice site and U2 bound to the branchsite communicate directly. The possibility of a physical interaction between U1 and U2 is suggested by the observation that anti-U1 antibodies also precipitate U2, albeit at a low level (47).

By far the best understood interaction between snRNP components is the association between U4 and U6. As indicated above, both snRNAs are ordinarily isolated together in a single snRNP particle (28,8). The notion that the snRNAs interact by base pairing was first supported by the demonstration that, in the absence of associated proteins, the two RNAs co-migrate on native gels as a complex with a T_m of 37^0C (28). Subsequently, photochemical cross-linking was used to establish a specific intermolecular interaction between two 8-nucleotide complementary regions (63). More recently, phylogenetic data have allowed the identification of a second intermolecular helix, 16-18 base pairs in length (11). This extensive base pairing interaction is of particular interest in light of unexpected results from both Hela and yeast in vitro systems, suggesting that spliceosomal complexes about to undergo the first nucleolytic cleavage event have lost U4, but retain U6 (17,37). It is important

to stress that the absence of U4 from these complexes may well be an artifact of the gel methodologies used for the analysis. Nonetheless, these results clearly establish that the association between U4 and U6 is a dynamic one, entailing a substantial conformational change during the course of the splicing reaction.

A final observation pertaining to snRNP-snRNP interactions is the ubiquitous observation of complexes containing U4/U6 and U5. While nothing is known about the structural requirements of formation of this multi-snRNP particle or the interactions involved, it is important to note that in both yeast and mammalian systems, the distribution of complexes containing different combinations of these three snRNAs is strikingly altered by the availability of ATP (37,32,17, see also 42).

In summary, a variety of strategies of snRNP-mediated recognition events in splicing have already been identified. These range from direct base pairing interactions involving the snRNA components of snRNPs, on the one hand, to interactions apparently orchestrated solely by protein factors, on the other. One striking theme is the widespread effect of ATP on snRNP structures and associations; this is illustrated both by the influence of ATP on the association of the U4/U6 and U5 snRNPs (see above), and by the ATP-induced increased sensitivity of U2, U4 and U6 snRNAs to oligo-directed RNAse H cleavage (4,33,5,2). These findings, emphasized by the multiple requirements for ATP throughout the process (see, for example, 17), suggests the likelihood that individual snRNPs participate in multiple, dynamic interactions during the course of the splicing reaction. The identification of

conserved sequence and structural elements of the snRNPs provides a focus for pinpointing these interactions.

U4/U6

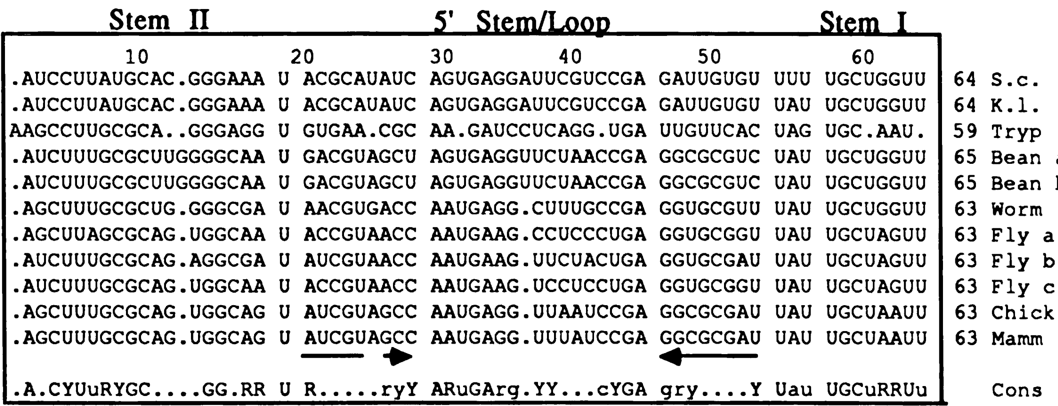
Unlike the case with other yeast spliceosomal snRNAs, the U4 (71) and U6 (11) analogues are virtually identical in size to their mammalian counterparts. Nonetheless, the U4 RNAs share only limited regions of primary sequence homology; as shown in Figure 1, with the exception of the Sm binding site, the homology appears to be largely constrained to the 5' portion of the molecule. In contrast, the U6 RNAs (Figure 2) can be aligned almost perfectly throughout their length, except for a short, variable region at the very 5' terminus. Indeed, there is 60% identity over the full length of yeast and mammalian U6, increasing to more than 80% identity in the middle third of the RNAs. This feature makes U6 by far the most highly conserved snRNA; in fact, it vies only with 5S rRNA as the most conserved RNA in biology. The apparent constraints on both the size and sequence of U6 argue for a central role in the splicing process and suggest that this snRNA is in close contact with several components of the machinery (11).

We have divided U6 into four domains (Figure 2, Figure 6). The 5' terminal region contains a stem/loop structure which is highly

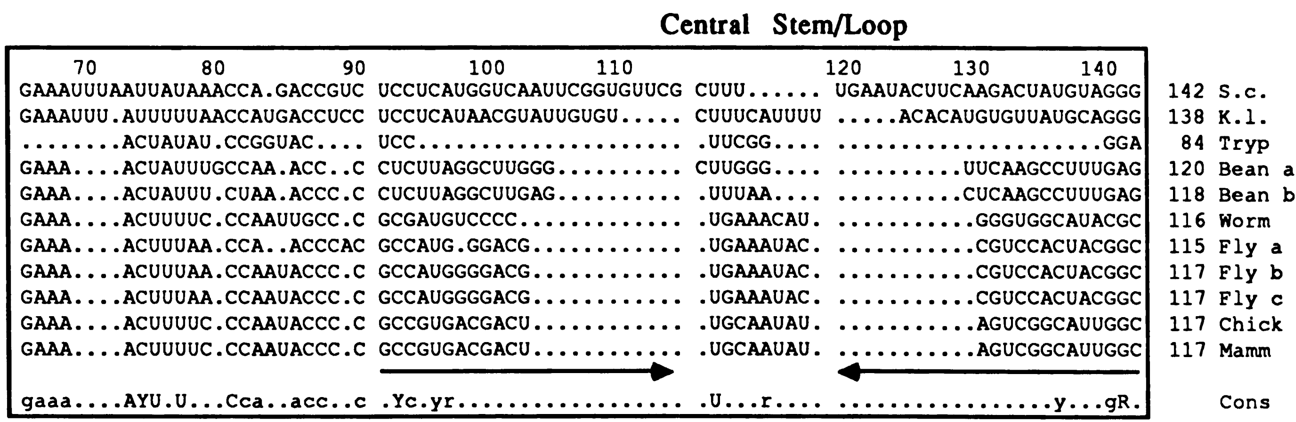
Figure 1: U4 alignment

The sequences for Bean (Vicia faba), Fly (Drosophila melanogaster), Chick (chicken), and Mamm (mouse, rat and human U4 sequences show only minor variations at two postions and do not alter the consensus line, so only the human sequence is shown) are taken from (60); for S.c. (Saccharomyces cerevisiae) from (71); for K.l. from (Kluyveromyces lactis) (H. Roiha and C. Guthrie, in preparation); for Tryp from (Trypanosoma brucei) (Mottram et al., submitted); and for Worm (Caenorhabditis elegans) from (J. Thomas, E. Zuker-Aprison, and T. Blumenthal, submitted for publication). The nucleotide position number refers to the S. cerevisiae sequence. The last digit of the nucleotide position number is above the nucleotide to which it refers. The RNA has been arbitrarily divided into three domains. Stems I and II refer to intermolecular helices formed with U6. Sm refers to the consensus Sm protein binding site. Dots have been introduced to show where spacing has been changed to improve the alignment. Dashes have been employed where an snRNA lacks an analogue of a given domain. Inverted arrows indicate the presence of an intramolecular helix. Small unpaired regions or bulged nucleotides within a helix have been included within the helix. Helices are extended as far as base pairing is possible except in those cases where they would include other recognizable domains (e.g. Sm sites). The consensus line (Cons.) is derived by the following rules: i) invariant residues are indicated as a capital A, G, C, U; ii) Y and R indicate an invariant pyrimidine or purine, respectively; iii) a residue which is identical in all available sequences but one is indicated by a lower case letter; iv) two or more differences

(including deletion or insertion of a residue) are indicated by a dot; v) regions containing dashes (or blanks, see below) are not considered in deriving the consensus line.



5' Terminal Domain



Central Domain

```
Sm Site         3'  Stem/Loop

   150                                            160
AAUUUUUGG ---- ------- ---- ------- AA..UACCUUU  160 S.c.
AAUUUUUGG ---- ------- ---- ------- AACCUACCU..  156 K.l.
AAGUU.UGC UA.. .CCCACC .ACG GGUGGG. A..........  110 Tryp
AAUUUCUGG AAGG ..ACUCC CUUU GGGGU.. AAAGUCUA...  155 Bean a
AAUUUCUGG AAGG ..ACUCC .UUC GGGGU.. AAAGUCUA...  152 Bean b
AAUUUUUGA AC.. ..GCCUC UAG. GAGGC.. AGGA.......  144 Worm
AAUUUUUGG A... .AGCCCU UACG AGGGCU. AA.........  143 Fly a
AAUUUUUGG AA.. ...GCCC GAGA GGGC... CA.........  142 Fly b
AAUUUUUGG AA.. ...GCCC GAGA GGGU... CAA........  143 Fly c
AAUUUUUGA .... CAGUCUC UACG GAGACUG G..........  145 Chick
AAUUUUUGA .... CAGUCUC UACG GAGACUG G..........  145 Mamm
               ------>      <------

AAuUUyUGr .... ....yYc .... gRr.... ...........      Cons
```

3' Terminal Domain

Figure 2: U6 alignment

The sequences for Fly (Drosophila melanogaster), and Mamm (mammals; the sequences for mouse, rat and human are identical) are taken from (60); Bean (Vicia faba) is from (T. Kiss, M. Antal, and F. Solymosy, submitted for publication), for S.c. (Saccharomyces cerevisiae from (11); for K.l. (Kluyveromyces lactis) from (H. Roiha and C.Guthrie, in preparation); for Tryp (Trypanosoma brucei) from (J. Mottram, K. Perry, P. Lizardi et al., submitted for publication); and for Worm (Caenorhabditis elegans) from (J. Thomas, E. Zuker-Aprison, and T. Blumenthal, submitted for publication). The nucleotide position number refers to the S. cerevisiae sequence. The division of U6 into four domains is as described previously (11); stems I and II refer to intermolecular helices formed with U4. Sequences may be missing from the 5' end of the bean RNA (F. Solymosy, personal communication). Other conventions are as in the legend to Figure 1.

```
          5' Stem/Loop                                                          Stem I     Stem II

             10                   20     | 30        40        50           |    60         70         80    |         90            100           110      |
...GUUCGCGAAG..  UAACC.  ..CUUCGUGGACAU  | UUGGUCAAUUUGAAACAAUACAGAGAU      | GAUCAGCA  GUUCCCCUGCAUAAGGAU | GAACCGUUUUACAAA....GAGAUUUAUUUCG..UUUU   | 112 S.c.
..AGUCCAUAUUACC  CUCCGU  GGUUUUAUGGACAU  | UUGGUCAAUUUGAAACAAUACAGAGAU      | GAUCAGCA  GUUCCCCUGCAUAAGGAU | GAACCGUUUUACAAA....GAGAUUUAAGA....UUUU   | 116 K.l.
GGAGCCC........  .UUCG.  .......GGGACAU  | CCA.CAAACUGGAAAUUCAACAGAGAA      | GAUUAGCA  CUCUCCCUGCGCAAGGCU | GAUGU......CAAUCUUCGAGAGAUAUAGC...UUUU   | 100 Tryp
      C........  .UUCG.  .......GGGACAU  | CCGAUAAAAUUGGAACGACACAGAGAA      | GAUUAGCA  UGGCCCCUGCGCAAGGAU | GACACG...CACAAA..UCGAGAAAUGGUCCAAAUUUU   |  98 Bean
...GUUCU.......  .UCCG.  .......AGAACAU  | AUACUAAAAUUGGAACAAUACAGAGAA      | GAUUAGCA  UGGCCCCUGCGCAAGGAU | GACACG.....CAAA.UUCGUGAAGCGUUCCAAAUUUU   | 101 Worm
...GUUCUUGC....  .UUCG.  .....GCAGAACAU  | AUACUAAAAUUGGAACGAUACAGAGAA      | GAUUAGCA  UGGCCCCUGCGCAAGGAU | GACACG.....CAAA.AUCGUGAAGCGUUCCACAUUUU   | 106 Fly
...GUGCUCGC....  .UUCG.  .....GCAGCACAU  | AUACUAAAAUUGGAACGAUACAGAGAA      | GAUUAGCA  UGGCCCCUGCGCAAGGAU | GACACG.....CAAA.UUCGUGAAGCGUUCCAUAUUUU   | 106 Mamm
   ---------->           <-----------    |                                  |                              |                                          |
...GuyC........  .uyCg.  ........GrACAU  | .YR.u.AA.UuGRAAcra.ACAGAGA.      | GAUYAGCA  ...cCCCUGCRYAAGGaU | GA..cg.....CAAa....G.GA..YR...c...UUUU   |     Cons
```

5' Terminal Domain Central Domain U4 Interaction Domain 3' Terminal Domain

variable in length (14 to 31 ntes.) and is responsible for essentially all the difference in size of the U6 RNAs (98-116 ntes.). Interestingly, however, the first, second and fourth base pairs are invariant, as is the dinucleotide 3' of the stem (AU), which delimits the 5' terminal domain. The response of this region in RNase H-directed cleavage experiments (5) suggests that it may serve as a binding site for a U6-specific protein. We are unable to draw conserved structures for the central or 3' terminal domains. It seems likely that portions of these regions are single-stranded, due to their RNase H susceptibility (5); moreover, the particularly striking conservation of the 3' half of the central domain suggests that these parts of U6 are involved in important interactions.

The longest contiguous stretch of conserved sequence in U6 has been proposed to participate in an extensive intermolecular base-pairing interaction with U4 (11). The existence of stem I was initially demonstrated by chemical cross-linking of mammalian snRNAs (63). Compensatory changes of two A-U base pairs to G-C and G-U have since been found in yeast, bean, and worm (J. Thomas, E. Zucker-Aprison, and T. Blumenthal, submitted for publication); the geometry of stem I in trypanosomes is slightly altered, but the same total number of base-pairs (eight) is maintained. Stem II (first proposed in (28) is supported by numerous compensatory changes; while most organisms have one or more mismatches (or A-G oppositions) within this region, the trypanosome stem contains 16 perfect base-pairs!

In U4, the two intermolecular helices are separated by an intramolecular helix. This stem is perfectly conserved between yeast

and mammals with respect to its length, the position of a bulged nucleotide and the identity of its terminal base pairs, yet, remarkably, all 13 nucleotides comprising the stem's central portion change identity between yeast and mammals (11). The 17-nucleotide loop shows strong conservation at a number of positions. Its sensitivity to RNase H cleavage only after phenol extraction makes it a good candidate for a U4-specific protein binding site (5).

We have divided the remainder of U4 into two domains (Fig. 1). The 5' region of the central domain contains a region of high sequence conservation (except for short insertions or deletions in yeast and trypanosomes, respectively) and is available for targeted RNAse H cleavage in extracts (5,2), suggesting it is exposed in the snRNP. The 3' portion of this domain forms a stem/loop structure which exhibits extreme phylogenetic variability in size and sequence. In the two yeasts, the helix contains 18 base-pairs (as well as several bulged nucleotides), while in trypanosomes it appears to be amputated to only three base-pairs. The 3' terminal domain contains the Sm binding site. In all organisms except _S. cerevisiae_ and _K. lactis_, this is followed by a 3' terminal stem.

In sum, the most remarkable aspect of the U4/U6 snRNP is the interaction domain. The Tm of the immunoaffinity-purified human U4/U6 snRNP was found to be $37^{o}C$, either as the intact particle or after treatment with SDS and Proteinase K (28). Similar measurements of the yeast particle gave a Tm about 15 degrees higher, though the value was again identical with and without deproteinazation; in this work, both $37^{o}C$ and $55^{o}C$ Tm forms of the U4/U6 complex were identified in Hela extracts (11). These apparent differences in stability

notwithstanding, the strength of the interaction is intriguing in light of the recent suggestion that U4 and U6 undergo cyclic dissociation during the course of the splicing reaction (17,37). As discussed above, even if the observed loss of U4 is more accurately interpreted as a destabilization, it seems clear that the two RNAs must undergo a significant conformational rearrangement, presumably involving dissociation of the two intermolecular helices. Interestingly, portions of U4 and U6 involved in the interaction domain become sensitive to RNase H targeting only in the presence of ATP (5,2), suggesting the possible role of an energy-dependent helicase (17,37). In any event, it seems important to bear in mind that the U6 sequences bound by U4 show markedly higher conservation that does the comparable region of U4. One exciting explanation is that it may be the role of U4 to sequester a catalytically active domain of U6 until other conditions of spliceosome assembly have been satisfied. According to this view, U6 plays a central role in the spliceosomal apparatus, while U4 functions primarily as a negative effector. Mutational analysis of these molecules should be particularly illuminating in this regard.

U5

U5 snRNAs vary in size over a two-fold range, from 107 ntes. in dinoflagellates to 214 ntes. in _S. cerevisiae_ (for the longest of two forms; see below) (Figure 3). The molecule appears to be organized into two domains: 1) a 5' terminal compound helix, and 2) a 3' terminal domain containing a single-stranded region, which includes the Sm binding site, and a stem/loop. The 3' terminal stem/loop

varies considerably in size and sequence and, in fact, half of the transcripts from the single copy yeast gene (SNR7) terminate just prior to this structure (55). The dispensability of this terminal stem in the long form of the yeast RNA has now been demonstrated genetically (B. Patterson and C. Guthrie, unpublished observations). Thus this domain is phylogenetically variable and functionally non-essential (in at least one organism).

As shown in Figure 6, the structure of U5 is dominated by the long compound helical element designated stem I. In the consensus diagram this stem is pictured as divided into several substructures. The terminal loop (Loop I) is supported by two major stems (Ic, Ib) which are separated by an internal loop (IL2); a second internal loop (IL1) separates stem Ib from a shorter stem, Ia. The maximum conservation is found in the top portion of this structure. Loop I is 11 nucleotides in length, 9 of which are invariant in sequence; the length of stem Ic is also tightly constrained and most positions are either perfectly conserved or vary by transitions. IL2 contains elements of absolute sequence conservation on both the 5' and 3' sides of the stem. From here to the base of the compound element, however, the degree of conservation declines markedly.

How can the "extra" residues in yeast U5 be accounted for? Some appear to simply extend the length of single-stranded regions already present in the consensus structure (e.g., the insertion of 12 and 8 ntes. 5' and 3', respectively, to the Sm binding site). In contrast, 34 ntes. can be folded to form a novel stem-loop structure (I.1), located between IL2 and the top of stem Ib. The existence of this domain is supported by sequence analysis of U5-like RNAs from

Figure 3: U5 alignment

The sequences for Human, Rat, Mouse, Chick (chicken), Dino (dinoflagellate, *Crythecodinium cohnii*) and Tet (*Tetrahymena*) are from (60), Pea sequences are from (36), Fly (*Drosophila melanogaster*) is from (49) and S.c. (*Saccharomyces cerevisiae*) is from (55). The nucleotide position numbers refer to the yeast sequence. The yeast U5 molecule is found in two forms, the shorter of which ends at nte. 179 and thus lacks Stem/loop II. We have shown the long form. The RNA has been arbitrarily divided into two domains: a 5' terminal domain consisting of sub-stems Ia, Ib and Ic, as well as terminal loop Loop I and internal loops IL1 and IL2, (and in S. cerevisiae, Stem/Loop I.1), and a 3' terminal domain, consisting of a single-stranded region containing the Sm site, and a terminal stem loop structure (Stem/Loop II). Stem I.1 is not considered in the consensus line since it appears only in *S. cerevisiae* (and in other related yeasts (H. Roiha and C. Guthrie, unpublished observations). Other conventions are as in the legend to Figure 1.

```
         Ia             IL1             I b                    Stem/Loop  I.1              IL2        Ic        Loop I      Ic'    IL2'
    ----------------------------------------------------------------------------------------------------------------------------------
            10            20           30        40         50         60         70          80         90         100        110
AA. GCAGCUUUAC AGAUCAAUGGCGGAG GGAGGUCAACAUC. AAGAACUGUGGGC CUUUUAUU GCCUAUAGAACUU AUAACGAAC AUGGUUCU UGCCUUUUACC AGAACCAU CCG  113 S.c.
GA. GCCG...... ............... .UGUGAUGAUGACA ------------- -------- ------------- .UAGCGAAC UA.UCUUU CGCCUUUUACU AAAGAAUA CUG   56 Pea a
GGA GCCA...... ............... .UGUGAUAAGUGCA ------------- -------- ------------- .AAGCGAAC UA.UCUUU CGCCUUUUACU AAAGAAUA CUG   57 Pea b
AU. .CAC...... A.............. ..GUGUUCACUUCA ------------- -------- ------------- ..ACCGAAU CAAUCUUU CGCCUUUUACU AAAGGUUG CCG   55 Dino
AU. .CAC...... ............... .AGAACUCAGCUCA ------------- -------- ------------- .AUACGCUU UAAUUUUU CGCCUUUUACU AAAGAUUA CCG   56 Tet
AU. ACUC...... UG............. ...GUUUCUCUUCA ------------- -------- ------------- AUGUCGAAU AAAUCUUU CGCCUUUUACU AAAGAUUU CCG   58 Fly
AU. ACUC...... UG............. ...GUUUCUCUUCA ------------- -------- ------------- .GAUCGUAU AAAUCUUU CGCCUUUUACU AAAGAUUU CCG   57 Chick
AU. ACUC...... UG............. ...GUUUCUCUUCA ------------- -------- ------------- .GAUCGUAU AAAUCUUU CGCCUUUUACU AAAGAUUU CCG   57 Mouse
AU. ACUC...... UG............. ...GUUUCUCUUCA ------------- -------- ------------- .GAUCGUAU AAAUCUUU CGCCUUUUACU AAAGAUUU CCG   57 Rat
AU. ACUC...... UG............. ...GUUUCUCUUCA ------------- -------- ------------- .GAUCGCAU AAAUCUUU CGCCUUUUACU AAAGAUUU CCG   57 Human
    --------->                 ------------->                                                ------->             <-------
R.. .C........ ............... ......u.....Ca                                      ..r.CG.aY .a.uYUuU cGCCUUUUACu AaAgr.u. CYG      Cons.
    ----------------------------------------------------------------------------------------------------------------------------------
```

5' Terminal Domain

```
Ib'            IL1'        Ia'                              Sm Site                             Stem/Loop II
 120            130         140        150        160        170          180        190           200       210
GGUGUUGUCUCC.  AUAGAAACAG  GUAAAGCUGU CCGUUACUGUGGGCUUGCCA UAUUUUUUGG  AACUUUUC  UGCCCUUUUUCUC  AAU..  GAGUAAGGAGGGCG  U......  214 S.c.
UGUCAGCGUCACA  AUUAG.....  CGGC...... AUACGCU............. AGU.UUUUGG  A.......  AGAGUUCUCAA..  GUU..  ...UUGAGGGCUCU  G......  121 Pea a
UGUACGUGUCACA  AAGCGG....  UGGC...... AUACGAGU............ AAU.UUUUGA  AU......  GAGUUCUCU....  UGUU.  .....AGAGAACUC  UG.....  123 Pea b
UGAAUGGGACAC.  AUCAAU....  GUGA...... AUCUCUC............. AAU.UUUUGA  ........  GGGC.........  UCU..  ..........GCCC  CAC....  107 Dino
UGGGCUGGGUUCU  ACAAU.....  GUG....... AAUUAUUAA........... AAU.UUUUGA  ........  GGAUU........  GUGUG  .........AAUCC  UA.....  112 Tet
UGGAGAGGAAC..  ACUCUAA...  GAGU...... CUAAAACU............ AAUUUUUUAG  UC......  AGUCUUG......  UCG..  .......CAAGACU  GGGGCCA  124 Fly
UGGAGAGGAAC..  AACUCU....  GAGU...... CUUAAACC............ AAUUUUUUGA  ........  GCCUUGUU.....  CCGGC  .........AAGGC  UA.....  116 Chick
UGGAGAGGAAC..  AACUCU....  GAGU...... CUUAAACC............ AAUUUUUUGA  ........  GGCCUUGUC....  UUGGC  ........AAGGCU  A......  117 Mouse
UGGAGAGGAAC..  AACUCU....  GAGU...... CUUAAACC............ AAUUUUUUGA  ........  GCCUUGUU.....  CCGGC  ........AAGGCU  A......  116 Rat
UGGAGAGGAAC..  AACUCU....  GAGU...... CUUAACCC............ AAUUUUUUGA  ........  GCCUUGC......  CUUG.  .......GCAAGGC  UA.....  116 Human
<----------               <------                                                  ---------->          <--------
uG.r...g.....  A.........  ..g....... .y.................. aaU.UUUUgR  ........  r..yy........  .....  .........rr..y  .......      Cons.
```

5' Terminal Domain

3' Terminal Domain

several fungi phylogenetically closely related to S. cerevisiae (H. Roiha, E. Shuster, and C. Guthrie, unpublished observations), as well as by the results of chemical modification studies (D. Frank and C. Guthrie, unpublished observations). (It should be noted that the proposed structure is different from that originally suggested for S. cerevisiae) (55). In sum, the pattern of accomodating additional sequences by a combination of the expansion of existing domains as well as the creation of phylogenetically unique structures is reminiscent of observations from rRNAs (50).

U5 appears to be the most highly structured of the spliceosomal snRNAs, which in turn is likely to account for its extraordinarily low degree of primary sequence conservation. Moreover, it is known that mammalian U5 is surprisingly resistant to microccocal nuclease and RNAse H cleavage, (15,4) suggesting that even the single-stranded RNA sequences are largely inaccessible. Since, with the exception of the Sm site, primary sequence conservation is mainly restricted to Loop I and I12, these regions provide attractive candidates for specific contacts with IBP, the U5-associated protein which appears to interact with the 3' splice site in a sequence-dependent manner (22,75). In this regard, it is interesting to note that the S. cerevisiae U5 snRNP is specifically precipitated with antisera directed against a 260 kD protein known to be required for splicing (see Table 1) (42); this protein is the product of the RNA8 gene (29). Suggestive evidence that this protein may be directly contacting the conserved loop domain comes from genetic experiments performed in yeast. Strains have been constructed in which a temperature-sensitive allele of rna8 is combined with a

growth-impairing Loop I point mutation of snr7; the resulting double mutant is inviable (B. Patterson and C. Guthrie, unpublished observations). The significance of this synergistic defect is suggested by its specificity: only a subset of Loop I mutations exhibit this behavior, and only in conjuction with the rna8 mutation (as opposed to several other ts alleles of yeast splicing gene products).

U1

To date, in all organisms besides Saccharomyces cerevisiae (72,35), the size and general organization of U1 is highly conserved (see (60) and references therein) (Figure 6): 1) a 5' single-stranded region, which contains the sequences complementary to the 5' splice site, is followed by 2) a long-range interaction of 4-5 base pairs, which closes off 3) a central domain comprising three contiguous stem-loop structures designated I, II, and III; 4) the 3' terminal domain consists of the single-stranded Sm binding site, followed by 4) a terminal stem-loop structure, designated IV. The length of each domain is also generally conserved in these cases, the total size of U1 varying only between 165 (mouse) and 147 (Schizosaccharomyces pombe) ntes.

The situation appears to be surprisingly different in S. cerevisiae, however, which contains a U1 analogue of 568 nucleotides. As shown in Figure 4, there is strong sequence and structural homology to all U1 species within the 5' terminal 50 nucleotides of the yeast RNA, including a 9/10 match to humans in the single-stranded loop (I). Moreover, it is possible to form a 5

Figure 4 U1 alignment

The sequences for C. Bean (common bean, Phaseolus vulgaris), Fly (Drosophila melanogaster), Frog (Xenopus laevis), Chick (chicken), Mouse, and Human are taken from (60); for S.c. (Saccharomyces cerevisiae) from (72,35); for S.p. (Schizosaccharomyces pombe) from (G. Porter and J. Wise, unpublished observations); for Alga (Chlorella saccarophila) from (31); and for Soy (soybean U1 a and b) from (79). The RNA has been divided into three domains, as described in the text. "5' SS Rec" is the region which recognizes the 5' splice site by base-pairing and "Long Range" refers to a helix formed by long-range intramolecular base pairing; this helix is frequently closed with an A-G base pair, a familair motif in rRNA (50). The nucleotide position number refers to the human sequence. Note that residues 40-544 have been omitted from the S. cerevisiae sequence because we are unable to identify homologies in this region. Other conventions are as in the legend to Figure 1.

5' SS Rec.	Long Range		Stem/Loop I			Stem/Loop II				
AUACUUACCUU	AAGAU	AUCAG..AGGA.	GAUCAAGAAG	.UCCUACUGAU				40	S.c.
..ACUUACCUG	GCAUG	AGUUU.CUGCAG	CA.CAAGAA.	UUGUGGAGACU	.CAGUUAUUUGUCUUGGC	AUUGCACUGA.	GCCCUGACGAAUAACUG	88	S.p.
AUACUUACCUG	UCCGG	.CCUGCGACCUC	GAGCAAGAAG	GGGGUCUAGGU	AGUGCUUGU..ACCUCGC	CUUGUACUAU.	GCUUGGGGU...AGCGC	88	Alga
AUACUUACCUG	GACGG	GGUCAA.UGGAU	GAUCUAUAAG	GUCCA.UGGCC	.UAGGG.AAGUGACCUUC	AUUGCACUCA.	GAAGGGGUGCUACUCUA	90	C.bean
AUACUUACCUG	GACGG	GGUCAA.UGGAU	GAUCAAGAAG	GUCCA.UGGCC	.UAGGG.AAGUAACCUCC	AUUGCACUGA.	GGAGGGGUGCCUUUCUA	90	Soy a
AUACUUACCUG	GACGG	GGUCAA.UGGAU	GAUCAAUAAG	GUCCA.UGGCC	.UAGGG.AAGUAACCUCC	AUUGCACUUA.	GGAGGGGUGCUUUCCUA	90	Soy b
AUACUUACCUG	GCGUA	GAGGUUAACCGU	GAUCACGAAG	GCGGU.UCCUC	.CGGAGUGAGGCUUGGCC	AUUGCACCUC.	GGCUGAGUUGACCUCUG	92	Fly
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UCUCC	.CAGGGCGAGGCUCAGCC	AUUGCACUCC.	GGCUGUGCUGACCCCUG	91	Frog a
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UCUCC	.CAGGGCGAGGCUCAGCC	AUUGCACUCC.	GGCCGUGCUGACCCCUG	91	Frog b
AUACUUACCUG	GCAGG	GGAGAC.ACCAU	GAUCAGGCAG	GUGGU.UUUCC	.CAGGGCGAGGCUCAUCC	CCUGCACUCC.	GGGUGUGCUGACCCCUG	91	Chick
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UUUCC	.CAGGGCGAGGCUUAUCC	AUUGCACUCC.	GGAUGUGCUGACCCCUG	91	Mouse a1
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UUUCC	.CAGGGCGAGGCUCACCC	AUUGCACUUUG	GGGUGUGCUGACCCCUG	92	Mouse b1
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UUCCC	.CAGGGCGAGGCUCACCC	AUUGCACUUUG	GGGUGUGCUGACCCCUG	92	Mouse b4
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCGCGAAG	GUGGU.UUUCC	.CAGGGCGAGGCUCACCC	AUUGCACUUUG	GGGUGUGCUGACCCCUG	92	Mouse b5
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UUUCC	.CAGGGCGAGGCUUAUCC	AUUGCACUCC.	GGAUGUGCUGACCCCUG	91	Rat
AUACUUACCUG	GCAGG	GGAGAU.ACCAU	GAUCACGAAG	GUGGU.UUUCC	.CAGGGCGAGGCUUAUCC	AUUGCACUCC.	GGAUGUGCUGACCCCUG	91	Human
		⟶		⟵	⟶		⟵		
auACUUACCUg	r...r	r.........r.	gA.Cr..aAg	.y....y...Y	.yrG...r.....Y...C	.YUGcACu...	G...g.gy......Cur		Cons.
10		20	30	40	50 60	70	80 90		

5' Terminal Domain

Central Domain

```
                                          Long
     Stem/Loop  III                       Range   Sm Site                    Stem/Loop  IV
                                          AUCUU | AU AAUUUUUGA UUUA ------------- ------ ------------ UUUU | 568 S.c.
UGGACUGGC UAAG...  GUCAGCUCCG GAUGC | AU CAUUUUUGA GUUC ........GUCCC UCAUUU GGGGC....... .... | 147 S.p.
UGUGUGGGG CAAGU..  CCUCGUUACA ACGGA | AU AAUUUCUGG CAG. GCCGUUGCACGCG CUUG.. CGCGUCCUCGGC AA.. | 162 Alga
AGGUCUGUC CAAGU..  GAUGGAGCCU ACGUC | AU AAUUUGUGG UAGU GGGGGCCU..GCG UUCG.. CGC.GGCCCCU. UAC. | 162 C.bean
AGGUCUGUC CAAGU..  GACAGAGCCU ACGUC | AU AAUUUGUGG UAGU GGGGGCCU..GCG UUCG.. CGC.GGCCCCU. UUC. | 162 Soy a
AGGUCUGCC CAAGU..  GGCAGAGCCU ACGUC | AU AAUUUGUGG UAGU GGGGGCUU..GCG UUCG.. CGC.AGCCCCU. UC.. | 161 Soy b
CGAUUAUUC CUAAUGU  GAAUAACUCG UGCGU | GU AAUUUUUGG UAGC .CGGGAAU.GGCG UUCG.. CGCCGUCCCG.. A... | 164 Fly
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUCUGG UAGU GGGGGACU..GCG UUCG.. CGCUUUCCCCU. G... | 164 Frog a
CGAUUUCCC CAAAUGU  GGGAAAGUCG ACUGC | AU AAUUUCUGG UAGU GGGGGACU..GCG UUCG.. CGCUUUCCCCU. G... | 164 Frog b
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 164 Chick
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 164 Mouse a1
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 165 Mouse b1
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 165 Mouse b4
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 165 Mouse b5
CGAUUUCCC CAAAUGC  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUCUCCCCU. G... | 164 Rat
CGAUUUCCC CAAAUGU  GGGAAACUCG ACUGC | AU AAUUUGUGG UAGU GGGGGACU..GCG UUCG.. CGCUUUCCCCU. G... | 164 Human
 ------->           <-------                                    ------------->        <---------
.Gr.Y...c caARu..  g...R..yC. r...y | aU aAUUU.UGR y... ...g...y..gCg uuyg.. cGc...cyc... .... |     Cons.
       100            110       120 |       130      140               150       160
```

Central Domain — 3' Terminal Domain

base pair long-range interaction, the 3' end of which would directly abut an Sm binding site embedded in a 10/12 nte. homologous stretch. The 3' end of yeast U1 terminates shortly thereafter; thus, as with the case of U4 and U5, S. cerevisiae lacks a 3' terminal stem/loop element. Unexpectedly, however, we have been unable to identify any unambiguous counterparts of stem/loop elements II and III. Moreover, genetic analyses show that at least 196 ntes. can be deleted between the 3' end of stem/loop I and the putative Sm site without significantly impairing viability (P. Siliciano and C. Guthrie, unpublished observations). While further studies are needed to define the maximum deletion endpoints, it is conceivable that, as with U2 (see below), U1 contains internal structural domains which are phylogenetically variable and functionally non-essential. To resolve this question, sequence information is also being sought from fungi related to Saccharomyces.

The absolute conservation of the 5'-most stem/loop element (I) fits nicely with recent data showing that this region provides the major recognition determinants for the snRNP-specific proteins U1-70k (27,57) and U1-A (27) (the unavailability of antibodies unique to U1-C preclude its analysis). While other domains of the snRNA can make stabilizing contributions to this RNA:protein association, as indicated by certain assays, other results suggest that Loops II and III are at least partially accessible to nuclease digestion (i.e., under conditions when stem/loop I is protected) (27,57,60). More detailed experiments are needed to clarify precisely which regions of the snRNA are bound by protein.

It is, of course, of obvious and considerable importance to determine which sequences of U1 are available for RNA base-pairing interactions during splicing. We discussed earlier the arguments for a direct interaction between the U1 and U2 snRNPs. Several base-pairing schemes have been proposed previously (see, for example, 5). One suggestion (U1 ntes. 65-72:U2 ntes. 30-37) requires the participation of the loop sequences of helix II, which do not appear to be retained in yeast U1 (even the closest homology would confer a central mismatch). A second hypothesis (U1 11-20:U2 5-14) is not phylogenetically supported (see, e.g., S. cerevisiae and S. pombe U1 sequences) and, moreover, is at variance with the current consensus structures in Figure 6 in that it would require opening both the long-range and helix I stems in U1, and a third stem (I) in U2. The failure to identify long conserved regions of potential base pairing suggests that the U1/U2 interaction may occur via multiple, shorter interactions. This model is consistent with the observation that the low-level cross-immunoprecipitation of U1 and U2 snRNPs is sensitive to perturbation of any of several U2 structural domains tested (47). Alternatively, though in our view less likely, the association between U1 and U2 may be indirect, or mediated exclusively by protein: protein interactions.

U2

The known size variation of U2 is by far the greatest for the spliceosomal snRNAs, ranging from 148 (trypanosomes) to 1175 (S. cerevisiae) ntes. (Figure 5). This exceeds the differential found for extremes of rRNAs by at least a factor of two. In the case of U2, _

Figure 5: U2 alignment

The sequences for Human, Rat, Mouse, Chick (chicken), Frog (Xenopus laevis), Fly (Drosophila melanogaster), Bean (Vicia faba), Tryp (Trypanosoma brucei), and S.c. (Saccharomyces cerevisiae) are from (60). Worm (Caenorhabditis elegans) is from (J. Thomas, E. Zuker-Aprison, and T. Blumenthal, submitted for publication). submitted), Arab (Arabidopsis thaliana) is from (80), S.p. (Schizosaccharomyces pombe) is from (P. Brenwald, G. Porter, and J. Wise, submitted for publication). The molecule has been arbitrarily divided into three domains: 5' and 3' terminal and central. The sequences with the potential to form a pseudoknot are designated (psi1 and psi2). Note that residues 116 to 1082 have been omitted from the S. cerevisiae sequence because we are unable to identify clear homologies in this region. Other conventions are as in the legend to Figure 1.

```
                          Stem/Loop I                  SS Region              Ψ1      IIa       Ψ2       IIb     IIb'                                  IIa'
ACGAAUC UCU.UUGCC UUUU GGCUUAGA | UCA.AGUGUAGUAUCUGUUCU UUUCAG UGUAACAA CUGAAA.  .UGACCUC .AAU.  GAGGCUCA..  UUACCUUUUAAU. UUGUUACA    105 S.c.
AUUC.UC UCU.UUGCC UUUU GGCUUAGA | UCA.AGUGUAGUAUCUGUUCU UUUCAG UUUAAUCG CUGAAAU  ..CACCUC .ACU.  GAGGUG....  UUC.......... CGAUUAA.     85 S.p.
AUAU.CU UCU.CGGCU AUUU AGCUAAGA | UCA.AGUUAUUAAACUGUUCU UAUCAG AGUAACUC CUGAUAC  ...GGGCC .UUU.  GGCCC.....  AAGGAUCAAAACU ..GUUGC.     85 Tryp
AUAC.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUUAG UUUAAUAU CUGAUA.  UGUGGGCC .AAU.  GGCCCACA..  CG........... AUAUUAAA    87 Bean
AUACCUU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUA.  UGUGGGCC .AUC.  GGCCCACA..  CG........... AUAUUAA.    88 Arab
AUCG.CU UCUUCGGCU UAUU AGCUAAGA | UCAAAGUGUAGUAUCUGUUCU UAUCGU AUUAACCU ACGGUA.  .UACACUC GAAU.  GAGUGUA...  AUAA......... AGGUUA..    90 Worm
AUCG.CU UCU.CGGCC UUAU GGCUAAGA | UCAAAGUGUAGUAUCUGUUCU UAUCAG CUUAACAU CUGAUA.  GUUCCUCC .AUU.  GGAGGACAAC  AA........... AUGUUAAA    90 Fly
AUCG.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUAC  ..GUCCCC UAUCU  GGGGAC....  CAU.......... AUAUUAAA    87 Frog
AUCG.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUAC  ..GUCCUC GAUGA  GAGGAC....  UUU.......... AUAUUAAA    87 Chick
AUCG.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUAC  ..GUCCUC UAUCC  GAGGAC....  AAU.......... AUAUUAAA    87 Mouse
AUCG.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUAC  ..GUCCUC UAUCC  GAGGAC....  AAU.......... AUAUUAAA    87 Rat
AUCG.CU UCU.CGGCC UUUU GGCUAAGA | UCA.AGUGUAGUAUCUGUUCU UAUCAG UUUAAUAU CUGAUAC  ..GUCCUC UAUCC  GAGGAC....  AAU.......... AUAUUAAA    95 Human

Au...YY UCU.Y.GCY uuuU RGCU.AGA | UCA.AGUguaguAuCUGUUCU U.Ucag ..UAAY.. cuGa.A.  ......YC .a.y.  GR........  ............. ..RUUa..       Cons.
             10          20     |  30          40        50         60                 70           80                         90
```

5' Terminal Domain

Central Domain

```
        Sm Site                            Stem/Loop III
AUACA CAUUUUUUGG CACCC                                                        1079 S.c.
UCU.. UGUUU.UUGG UUUGA.  GUUGGAAAGCUC....UGG CUUG CUAUCUUUCCGAC.  AC           143 S.p.
----- ----- ---- ------ ------------------- ---- --------------   --            98 Tryp
UU... UAUUUCUUGA GGGG..  AAGAGGCC.ACCACAGUAG CUUG CUAUUGGGUCUCUU  AC           148 Bean
CUC.. UAUUUUUUAG GGA...  GAAAGCCC.ACUAAGAUAG CUUG CUAUCUGGGCUUUC  AC           149 Arab
UAU.. GAUUU.UUGG AACCUA  GGGAAG...ACU..CGGGG CUUG CUCCGACUUCCC..  AA           146 Worm
CU... GAUUU.UUGG AAUCA.  GACGGAG..UGCU.AGGAG CUUG CUCCACCUCUGUC.  GC           149 Fly
UG... GAUUU.UUGG AACAG.  GGAGAUGGAAGA...AGAG CUUG CUCUGUCCACUCC.  AC           146 Frog
CG... GAUUU.UUGG GCGCG.  GGAGUUGGACCC...GGAG CUUG CUCCCUCCGCUCC.  GC           146 Chick
UG... GAUUU.UUGG AAGUA.  GGAGUUGGAAUA...GGAG CUUG CUCCGUCCACUCC.  AC           146 Mouse
UG... GAUUU.UUGG AACUA.  GGAGUUGGAAUA...GGAG CUUG CUCCGUCCACUCC.  AC           146 Rat
UG... GAUUU.UUGG AGCAG.  GGAGAUGGAAUA...GGAG CUUG CUCCGUCCACUCC.  AC           146 Human

y.... .aUUU.UUgg ......  gr.R...........r.RG CUUG CU.Y.....Y.yy.  Rc               Cons.
        100        110        120          130       140
```

3' Terminal Domain

Stem/Loop IV

```
AAAGUUAGAGACGUCGCGACCC ..UCGCACUUGU.. .GGAGUCGUUCUUGACUUU UACUUUGGUCGCUUGAUGUUUCUCUCGUCUUCCCGUUCGCUCUUUUU 1175 S.c.
UGGUGUUCUUGCUA........ ..UUGCACUUAC.. .....UGGCAAGCGACGCC GAA............................................  184 S.p.
CUGUCCCGCGUUCUUCCGGG.. .GUUCCACUUG... UCCGGACGGAGCGCGACGG U..............................................  148 Tryp
GUGUCGCUCUUGC......... .GUUGCACU.AUA. .....GCAAUUGCUGGCGC ACCCCA.........................................  193 Bean
GAGUCGCCCAUGCG........ ..UUGCACU.AC.. ..........UGCACGGGC CUGGCUCAACCCGC.................................  195 Arab
GGGUCGUCCUGGCG........ ..UUGCACU.GC.. ....UGCCGGGCUCGGCCC AGU............................................  187 Worm
GGGUUGGCCCGGU......... .AUUGCAGU.ACC. ...GCCGGGAUUUUGGCCC AAC............................................  192 Fly
GCAUCGACCUGG.......... UAUUGCAGU.ACCU ......CCAGGACCGGUGC ACU............................................  187 Frog
GCAUCGUCCCGG.......... UAUGGCAGU.ACCU ......CCGGGCACGGUGC ACC............................................  187 Chick
GCAUCGACCUGG.......... UAUUGCAGU.ACCU ......CCAGGAACGGUGC ACC............................................  187 Mouse
GCAUCGACCUGG.......... UAUUGCAGU.ACCU ......CCAGGAACGGUGC ACC............................................  187 Rat
GCAUCGACCUGG.......... UAUUGCAGU.ACCU ......CCAGGAACGGUGC ACC............................................  187 Human

..Ruy...y............. ..UygCA.U.Ry.. ..............r...y ...............................................      Cons
 150                160        170       180
```

3' Terminal Domain

sequence and structural conservation of this domain is of particular interest in light of the finding that RNAse H-mediated cleavage of the first 15 nucleotides completely inactivates the snRNP (20,4,33,34), as judged by substrate protection or complex assembly assays (20,34). Thus this region is exposed and apparently available for an intermolecular interaction that is an early, essential step in mammalian splicing. It is curious that the extent of sequence conservation is greater for the stem and loop than it is for the 5' single-stranded tail.

Stem I is followed by a long, potentially single-stranded region with several interesting characteristics. Nucleotides 33-39 contain the sequence proposed to base-pair with the TACTAAC box in yeast introns (53). As described above, Watson-Crick interactions have been tested genetically at two positions (ntes. 36 and 37) to date. In mammalian systems, RNAse-H mediated degradation of this region does not block U2 snRNP binding to substrate (as assayed by its ability to protect substrate from RNAse cleavage) (34), but lariat formation (i.e. branchpoint utilization) is not observed (34,20). This suggests that the base-pairing interaction may also play a role in mammalian splicing, but at a later step in the pathway.

Immediately following the GUAGUA sequence is an absolutely invariant pyrimidine-rich heptanucleotide; to our knowledge, no function has been proposed for these residues. Interestingly, the proposed (30) single-stranded region immediately adjacent to this conserved sequence is capable of forming a pseudoknot with sequences drawn as part of the internal loop of stem II (M. Ares, personal communication). ("Pseudoknot" is the term used to describe the

structure which results when a single helix is formed by continuous base pairing of one region of an RNA with two other, non-contiguous regions of the same RNA.) As shown in the alignment, the capacity to form the pseudoknot is supported by compensatory changes in the U2 RNAs from two yeasts and from worm. Nonetheless, the phylogenetic data do not address the critical question of whether the two halves of the pseudoknot helix form simultaneously, or are present in two mutually exclusive structures. While no data yet distinguish between these possibilities, it is interesting to consider the notion of a conformational "switch" . The presence of a number of absolutely conserved base pairs in the lower half of stem II (IIa) are consistent with the notion that the sequences in this region play roles beyond stem formation per se.

All U2 snRNAs terminate in a stem/loop structure (IV). The helix is variable in length (10 - 16 base pairs) and is generally interrupted midway by a single internal loop, the size of which is not constant. By far the most conserved aspect of this domain is the terminal loop (Loop IV). A 10 nte. consensus can be derived, from which no species varies at more than two positions. It is interesting

revealed the curious finding that strains lacking all or part of this stem give rise to very slow-growing colonies containing U2 RNAs with improperly terminated 3' ends (E. Shuster and C. Guthrie, submitted). One provocative interpretation of these results is that U2-specific proteins are required for a 3' processing event which is itself not absolutely essential for snRNP function.

We will close with a consideration of the trypanosome data, the interpretation of which presents a special challenge. On the one hand, the phylogenetic depth of these organisms provides an invaluable benchmark for identifying absolutely conserved features of snRNA sequence and structure. On the other hand, important variations are expected to account for mechanistic differences which distinguish cis from trans splicing. Thus we should be particularly alert to deviations from the database consensus which are unique to the trypanosome snRNAs, as these may provide valuable clues to this special mode of processing. In view of the yeast data just cited, it would appear that the absence of domain III is not restricted to trypanosomes. However, two other differences deserve comment.

The first relates to the role of the Sm protein in splicing. While it has been known for some time that trypanosome snRNAs fail to be immunoprecipitated by anti-Sm antibodies (S. Michaeli and N. Agabian, unpublished observations), this has left untested the possibility that they might be bound to a functionally equivalent but structurally diverged antigen. This hypothesis is supported by the U4 alignment (see Fig. 1), which neatly places a sequence with a potential match to the consensus (AAG_UUUG) in the predicted position. In contrast, trypanosome U2 does not contain an identifiable homology to the Sm binding site, and furthermore can be drawn in a configuration where stem II is directly followed by stem IV, with no single-stranded region where the Sm proteins would be expected to bind. We are left to conclude either that the U4 homology is fortuitous or, alternatively, that trypanosome U2 specifically lacks this binding site and perhaps, like U6 in other organisms,

depends upon association with another snRNA (such as U4) which does bind the Sm-equivalent protein.

The second difference addresses the mechanism of branchpoint selection in cis vs. trans splicing. The GUAGUA sequence (ntes. 34-39) is phylogenetically invariant with the single and striking exception of T. brucei, which contains a 5 out of 6 nucleotide substitution (UAUUAA). If the yeast example is of general relevance, this finding predicts that branchpoints recognized by a trans-splicing machinery will conform to a different consensus than do cis sites. In the absence of mapped branchpoints, this hypothesis cannot be directly tested; however, sequence analyses of 3' splice site regions do not reveal obvious candidates (see, for example, 23 and references therein). An alternative explanation is that the mechanism of branchpoint selection is uniquely different in trans, and that the corresponding region of U2 plays an unrelated role. The analysis of the snRNA populations in nematodes provides the opportunity to distinguish these alternatives, since this organism carries out splicing in cis and in trans. Using a probe complementary to the GUAGUA region, a number of nematode U2 species have recently been cloned (J. Thomas , E. Zuker-Aprison and T. Blumenthal, submitted for publication). The question is whether there remains to be identified a distinct class of U2 molecules, perhaps trypanosome-like in sequence, which function exclusively in trans splicing. A virtue of this model is that it offers a pleasing partial solution to the puzzle of how these two types of splicing, which operate on common transcripts, are kept separate.

CONCLUDING REMARKS

It is our intent in this section to briefly review the kind of information generated by the alignments themselves and to consider several applications for these data.

Comparative analysis of this set of RNAs demonstrates not only the degree of change tolerated within a given snRNA, but also the relative conservation among the five spliceosomal snRNAs. For example, the observation that large regions of U2 and U6 snRNAs are absolutely invariant at 60-80% of positions (1,11) suggests that these domains may be more evolutionarily constrained by their participation in multiple interactions in the splicing process. In the case of U6, we have speculated that the extreme conservation of size, sequence, and structure reflects an essential and inflexible role, perhaps one at the heart of trans-mediated catalysis of splicing. Conversely, with the exception of U6, entire domains in each of the spliceosomal snRNAs appear to be phylogenetically hyper-variable. The most extreme examples include the addition or deletion of large regions in *S. cerevisiae* U2, U1, and U5, and the deletion (or severe amputation) of stem/loop domains in U2 and U4 RNAs from the trypanosome *T. brucei*. Deletion experiments in yeast demonstrate that, in at least some cases, phylogenetically variable domains are also biologically dispensable (E. Shuster and C. Guthrie, in preparation, P. Siliciano, B. Patterson and C. Guthrie, unpublished observations). More work is needed to determine whether more subtle phenotypes result from such manipulations. In any event, the absence (or dispensability) of a 3' terminal stem/loop from from yeast snRNAs appears to be the rule rather than the exception.

Another striking observation is the notable difference of trypanosome U2 from other U2 snRNAs in two regions which would otherwise be considered absolutely invariant -- the GUAGUA region (corresponding to the branchsite recognition domain in yeast) and the Sm binding site. The potential for these differences to be obligate correlates of trans splicing will obviously bear close watching. Moreover, C. elegans, which performs both cis and trans splicing, should provide a proving ground for any models invoking specialized snRNA sequences/structures in trans splicing. The available sequences for worm U2, U4 and U6 (J. Thomas, E. Zuker-Aprison, and T. Blumenthal, submitted for publication) provide strong phylogenetic confirmation for invariant features of metazoan snRNAs, but it remains to be established whether these particular species are involved in cis- or trans-splicing.

A particularly important use of these alignments is to allow "gedanken" testing of base pairing models for snRNA-substrate and snRNA-snRNA interactions. A case in point is the previously proposed U1-U2 interactions, which are clearly not supported by phylogenetic comparison (see U1). In contrast, the base pairing interaction between U4 and U6 is considered proven by the existence of so many phylogenetic variants. By comparison of highly conserved regions between snRNAs, and the search for co-variants, the alignments should now provide the opportunity to generate new models for snRNA-RNA interactions.

Figure 6: Consensus structures of the U snRNAs

Consensus structures for the snRNAs are generated in the following fashion: starting with the most recent compatible structure known (for U1, (6), for U2, (Ares), for U4/U6, (11), and for U5, (55)) the sequence presented in the consensus line of each snRNA is folded into the consensus form. The following conventions are used: 1) all nucleotides from the consensus lines are presented with the same conventions used to generate them: 2) variable positions are designated by N; 3) base pairing is signified with a dashed line: 4) positions containing from zero to any number of non-conserved bases are indicated by (N): 5) variable-length regions of stems are signified by (N-N), where the number of base pairs represented is from 0 to any number: 6) in some cases, if a single nucleotide in a highly conserved region appears only in some sequences, it is shown in parentheses, and 7) the highly variable stem/loop domain in U5 is represented by a stick drawing to indicate this aspect. The Sm binding site consensus is indicated by a bold underline. Some consensus nucleotides from stems II and IV of the U1 alignment are not shown because they do not fall in uniform positions structurally.

U5

Loop I

Stem Ic

IL2

Stem/Loop I.1

Stem Ib

Stem/Loop II

IL1

Stem Ia

GpppR (N) N-NNy (N) aaU.UUUUgR (N) r-y (N)$_{OH}$

U1

Stem/Loop II

Stem/Loop IV

Stem/Loop I

Gppp auACUUACCUgr-ya U a A U U U . U G R y N N N (N-N) (N)$_{OH}$

5' SS Rec

U2

Loop II

IIb

Ψ2

IIa

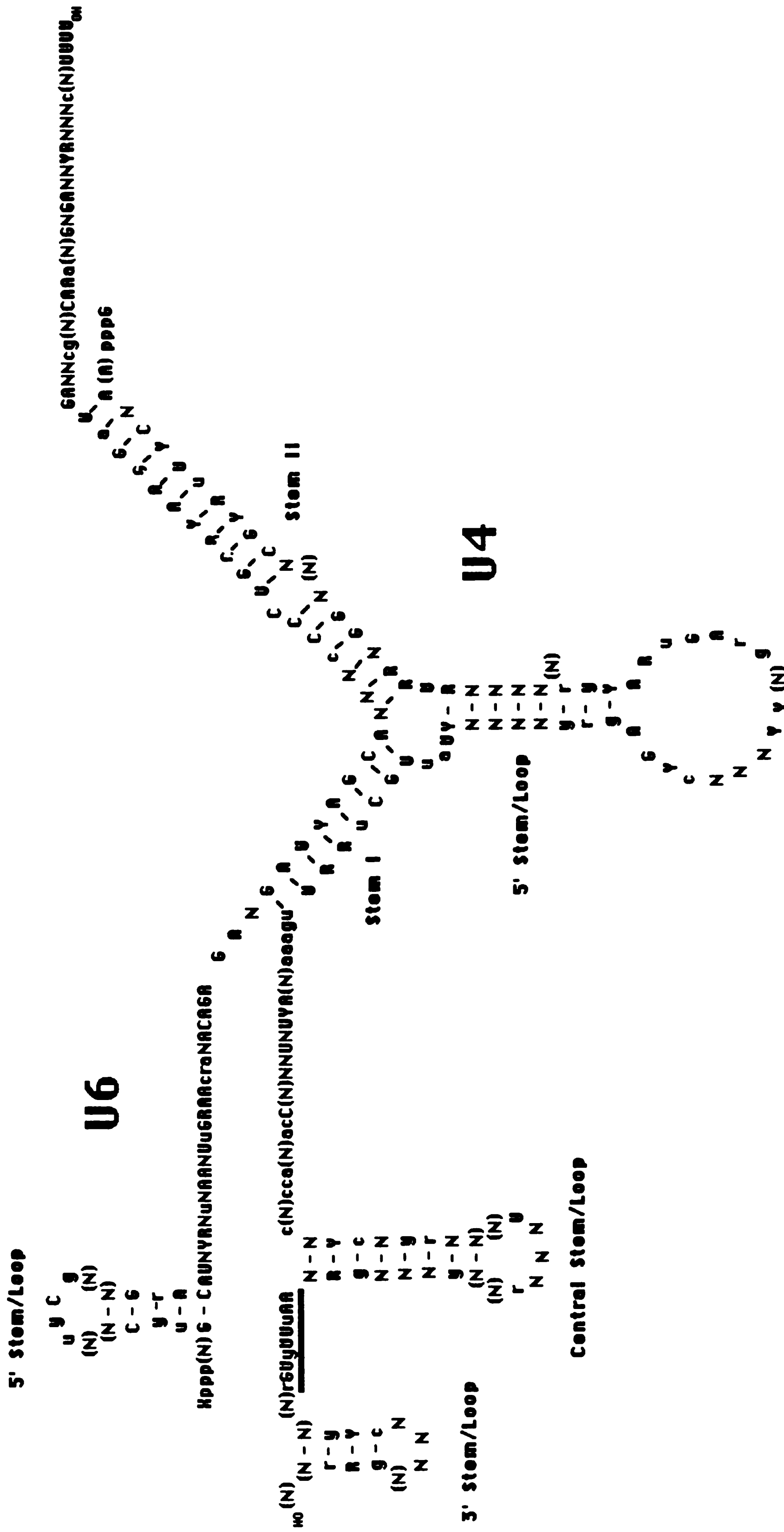
Stem/Loop III

Stem/Loop I

Stem/Loop IV

GpppAuNN (N) YYU-AUCA (A) AGUguaguAuCUGUUCUUNUcag (N) (N-N) y (N) NaUUU.UUgg (N) g-y N Rc NNRuyN (N) NyNNN (N)

Ψ1


U6
U4
5' Stem/Loop
3' Stem/Loop
Central Stem/Loop
Stem I
Stem II
5' Stem/Loop

Another interesting use of the consensus sequences is their utility in defining potential protein binding sites. The recent development of in vitro reconstitution systems (27,56) should foster a growing body of deletion and protection data to delimit the detailed structural requirements for protein binding. Phylogenetic conservation within these regions should identify good candidates for sequence-specific protein contacts. Further, the observation that at least two of the snRNP specific proteins (U1-A and U2-B") (73) are strikingly similar suggests that their binding sites may also be related, a prediction which can be tested a priori by examination of the U1 and U2 consensus lines for homologies.

Finally, the delineation of highly conserved regions of the snRNA provides a focus for future studies on their function. By restricting attention to these regions, the tasks of directed mutagenesis or RNase H degradation is reduced from monumental to approachable. The information gleaned from targeted degradation of different U2 sequences provides an excellent retrospective example of the power of this type of approach (20). The prospect of targeted mutagenesis in the yeast system has just begun to be exploited (54, P. Siliciano and C. Guthrie submitted; B. Patterson and C. Guthrie unpublished observations). In a general sense, the ultimate power of this approach lies in its potential to uncover roles for highly conserved regions for which no functions have yet been hypothesized.

ACKNOWLEDGEMENTS

We are indebted to our colleagues for generously sharing results prior to publication. We are particularly grateful to N. Agabian, T. Blumentahl, W. Filipowicz, R. Nelson, H. Roiha, F. Solymosy, J. Thomas, and J.A. Wise for allowing us to incorporate submitted sequences into the database. We gratefully acknowledge Dan Frank and Dave Brow for innumerable hours spent generating alignments and consensus structures for U1 and U4/U6, respectively. We appreciate Judy Piccini's participation in the preparation of the manuscript. We thank members of the lab for useful discussions. Cited work from this laboratory was supported by grants from NIH and NSF to CG and NIH Training Grant GM 08120 for BP.

REFERENCES

1. Ares, M. 1986. U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. Cell 47:49-59.

2. Berget, S. M., and Robberson, B. L. 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. Cell 46:691-96.

3. Bindereif, A. and Green, M. R. 1987. An ordered pathway of snRNP binding during mammalian splicing complex assembly. EMBO J. 6:2415-24.

4. Black, D. L., Chabot, B., and Steitz, J. A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell 42:737-50.

5. Black, D. L., and Steitz, J. A. 1986. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. Cell 46:697-704.

6. Branlant, C., Krol, A., and Ebel, J.-P. 1981. The conformation of chicken, rat and human U1A RNAs in solution. Nucl. Acids Res. 9:841-58.

7. Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B., and Jacob, M. 1982. U2 RNA shares a structural domain with U1, U4 and U5 RNAs. EMBO J. 1:1259-65.

8. Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H. et al. 1984. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. EMBO J. 3:1357-63.

9. Bringmann, P., Reuter, R., Rinke, R., Appel, B., Bald, R. et al. 1983. 5'-terminal caps of snRNAs are accessible for reaction with 2,2,7-trimethylguanosine-specific antibody in intact snRNPs. J. Biol. Chem. 258:2745-47.

10. Brody, E., and Abelson, J. 1985. The "spliceosome": yeast premessenger RNA associeates with a 40S complex in a splicing-dependent reaction. Science 228:1344-49.

11. Brow, D., and Guthrie, C. 1988. Yeast U6 snRNA is remarkably conserved. Nature in press.

12. Brunel, C., Sri-Widada, J., and Jeanteur, P. 1985. SnRNPs and ScRNPs in eukaryotic cells. Mol. Subcell. Biol. 9:1-52.

13. Busch, H., Reddy, R., Rothblum, L., Choi, Y. C. 1982. SnRNAs, SnRNPs, and RNA processing. Ann. Rev. Biochem. 51:617-54.

14. Cech, T. R., and Bass, B. L. 1986. Biological catalysis by RNA. Annu. Rev. Biochem. 55:599-629.

15. Chabot, B., Black, D. L., LeMaster, D. M. and Steitz, J. A. 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science 230:1344-49.

16. Chabot, B., and Steitz, J. A. 1987. Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. Mol. Cell. Biol. 7:281-93.

17. Cheng, S.-C. and Abelson, J. 1987. Spliceosome assembly in yeast. Genes and Develop. 1:1014-27.

18. Fisher, D. E., Conner, G. E., Reeves, N. H. Wisniewolski, R. and Blobel, G. 1985. Small nuclear ribonucleoprotein particle assembly in vivo: Demonstration of a 6S RNA-free core precursor and posttranslational modification. Cell 42:751-58.

19. Frendewey, D., and Keller, W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42:355-67.

20. Frendewey, D., Kramer, A. and Keller, W. 1988. Different small nuclear ribonucleoprotein particles are involved in different steps of splicing complex formation. Cold Spring Harbor Symp. Quant. Biol. vol. 53, in press.

21. Fresco, L. D., Kurilla, M. G. and Keene, J. D. 1987. Rapid inhibition of processing and assmbly of small nuclear ribonucleoproteins after infection with Vesicular Stomatitis Virus. Mol. Cell. Biol. 7:1148-55.

22. Gerke, V. and Steitz, J. A. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47:973-84.

23. Glass, D. J., Polvere, R. I. and van der Ploeg, L. H. T. 1986. Conserved sequences in transcription of the hsp70 gene family in Trypanosoma brucei. Mol. Cell. Biol. 6:4657-66.

24. Grabowski, P. J., and Sharp, P. A. 1986. Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonuclear protein particles in the nucleosome. Science 233:1294-99.

25. Green, M. R. 1986. Pre-mRNA splicing. Ann. Rev. Genet. 20:671-708.

26. Guthrie, C. 1986. Finding functions for small nuclear RNAs in yeast. Trends in Biochem. Sci. 11:430-34.

27. Hamm, J., Kazmaier, M. and Mattaj, I. W. 1987. In vitro assembly of U1 snRNPs. EMBO J. 6:3479-85.

28. Hashimoto, C., and Steitz, J. A. 1984. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nucl. Acids Res. 12:3283-93.

29. Jackson, S. P., Lossky, M. and Beggs, J. D. 1988. Cloning of the RNA8 gene of Saccharomyces cerevisiae, detection of the RNA8 protein and demonstration that it is essenctial for nuclear pre-mRNA splicing. Mol. Cell. Biol. 8:1067-75.

30. Keller, E. and Noon, W. 1985. Intron splicing: a conserved internal signal in introns of Drosophila pre-mRNAs. Nucl. Acids Res. 13:4971-81.

31. Kiss, T., Antal, M., Hegyi, H. and Solymosy, F. 1988. Plant small nuclear RNAs. IV. The structure of U1 RNA from Chlorella saccharophila: a phylogenetic support, in terms of RNA structure, for the probable interaction between U1 and U2 snRNPs during the splicing of pre-mRNA. Nucl. Acids Res. 16:2734-?.

32. Konarska, M. M. and Sharp, P. A. 1987. Interaction between snRNP particles in formation of spliceosomes. Cell 49:763-74.

33. Krainer, A. R. and Maniatis, T. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell 42:725-36.

34. Kramer, A. 1987. Analysis of RNAse-A-resistant regions of Adenovirus 2 Major Late precursor mRNA in splicing extracts reveals an ordered interaction of nuclear components with the substrate RNA. J. Mol. Biol. 196:559-73.

35. Kretzner, L., Rymond, B. C., and Rosbash, M. 1987. S. cerevisiae U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. Cell 50:593-602.

36. Krol, A., and Ebel, J.-P. 1983. U1, U2, and U5 small nuclear RNAs are found in plant cells. Complete nucleotide sequence of the U5 snRNA family from pea nuclei. Nucl. Acids Res. 11:8583-94.

37. Lamond, A. I., Konarska, M. M., Grabowski, P. J. and Sharp, P. A. 1988. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc. Natl. Acad. Sci USA 85:411-15.

38. Lamond, A. I., Konarska, M. M. and Sharp, P. A. 1987. A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. Genes and Develop. 1:532-43.

39. Lelay-Taha, M.-N., Reveillaud, I., Sri-Widada, J., Brunel, C. and Jeanteur, P. 1986. RNA-protein organization of U1, U5 and

U4-U6 small nuclear ribonucleoproteins in HeLa cells. J. Mol. Biol. 189:519-32.

40. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. 1980. Are snRNPs involved in splicing? Nature 283: 220-24.

41. Liautard, J. P., Sri-Widada, J., Brunel, C. and Jeanteur, P. 1982. Structural organization of ribonucleoproteins containing small nuclear RNA from HeLa cells. J. Mol. Biol. 162:623-43.

42. Lossky, M., Anderson, G. J., Jackson, S. P., and Beggs, J. 1987. Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. Cell 51:1019-26.

43. Luhrmann, R. 1988. snRNP proteins. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Max L. Birnsteil, Ed. Springer-Verlag. Berlin, Heidelberg, New York, London, Paris, Tokyo. pp 71-99.

44. Mattaj, I. W. 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell 46:905-11.

45. Mattaj, I. W. 1988. UsnRNP assembly and transport. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Max L. Birnstiel, Ed. Springer-Verlag. Berlin, Heidelberg, New York, London, Paris, Tokyo. pp 100-14.

46. Mattaj, I. W., and De Robertis, E. M. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNA proteins. Cell 40:111-18.

47. Mattaj, I. W., Habets, W. J., and van Venrooij, W. J. 1986. Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. EMBO J. 5:997-1002.

48. Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J. A. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell 33:509-18.

49. Myslinski, E., Branlant, C., Wieben, E. D. and Pederson, T. 1984. The small nuclear RNAs of Drosophila. J. Mol. Biol. 180:927-45.

50. Noller, H. F. 1984. Structure of ribosomal RNA. Ann Rev Biochem. 53:119-62.

51. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P.A. 1986. Splicing of messenger RNA precursors. Ann. Rev. Biochem. 55:1119-50.

52. Padgett, R. A., Konarska, M. M., Aebi, M., Hornig, J., Weissmann, C., et al. 1985. Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit B-globin genes. Proc. Natl. Acad. Sci. USA 82:8349-53.

53. Parker, R., Siliciano, P., and Guthrie, C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell 49:220-39.

54. Parker, R., Simmons, T., Shuster, E. O., Siliciano, P. G., and Guthrie, C. 1988. Genetic analysis of snRNAs in yeast: A sextuple mutant is viable. Mol. Cell. Biol., in press.

55. Patterson, B., and Guthrie, C. 1987. An essential yeast snRNA with a U5-like domain is required for splicing in vivo. Cell 49:613-24.

56. Patton, J. R., Patterson, R. J. and Pederson, T. 1987. Reconstitution of the U1 small nuclear ribonucleoprotein particle. Mol. Cell. Biol. 7:4030-37.

57. Patton, J. R., and Pederson, T. 1988. The M_r 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. Proc. Natl. Acad. Sci. USA 85:747-51.

58. Pikielny, C. W., and Rosbash, M., 1986. Specific snRNAs are associated with yeast spliceosomes. Cell 45:869-77.

59. Pikielny, C. W., Rymond, B. C., and Rosbash, M. 1986. Electrophoresis of ribonucleoproteins reveals an ordered assemly pathway of yeast splicing complexes. Nature 324:341-45.

60. Reddy, R. and Busch, H. 1988. Small nuclear RNAs: RNA sequences, structure, and modifications. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Max L. Bernstiel, Ed. Springer-Verlag. Berlin, Heidelberg, New York, London, Paris, Tokyo. pp.1-35.

61. Riedel, N., Wolin, S., and Guthrie, C. 1987. A subset of yeast snRNA's contains functional binding sites for the highly conserved Sm antigen. Science 235:328-31.

62. Riedel, N., Wise, J. A., Swerdlow, H., Mak, A. and Guthrie, C. 1986. Small nuclear RNAs from Saccharomyces cerevisiae: unexpected diversity in abundance, size and molecular complexity. Proc. Natl. Acad. Sci. USA 83:8097-9001.

63. Rinke, J., Appel, B., Digweed, M., and Luhrmann, R. 1985. Localization of a base paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. J. Mol. Biol. 185:721-31.

64. Rogers, J., and Wall, R. 1980. A mechanism for RNA splicing. Proc. Natl. Acad. Sci. USA 77:1877-79.

65. Ruskin, B., and Green, M. R. 1985. Specific and stable intron-factor interactions are established early during in vitro pre-mRNA splicing. Cell 43:131-42.

66. Ruskin, B., Greene, J. M., and Green, M. R. 1985. Cryptic branch-point activation allows accurate in vitro splicing of human β-globin intron mutants. Cell 41:833-44.

67. Ruskin, B., Zamore, P. D., and Green, M. R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52:207-19.

68. Rymond, B. and Rosbash, M. 1986. Differential nuclease sensitivity identifies tight contacts between yeast pre-mRNA and spliceosomes. EMBO J. 5:3517-23.

69. Rymond, B., Torrey, D. and Rosbash, M. 1987. A novel role for the 3' region of introns in pre-mRNA splicing of Saccharomyces cerevisiae. Genes and Develop. 1:238-46.

70. Schmelzer, C. and Schweyen, R. 1986. Self-splicing of group II introns in vitro: mapping of the branchpoint and mutational inhibition of lariat formation. Cell 46:557-65.

71. Siliciano, P. G., Brow, D. A., Roiha, H., and Guthrie, C. 1987. An essential snRNA from S. cerevisiae has properties predicted for U4, including interaction with a U6-like snRNA. Cell 50:585-92.

72. Siliciano, P. G., Jones, M. H., and Guthrie, C. 1987. Saccharomyces cerevisiae has a U1-like small nuclear RNA with unexpected properties. Science 237:1484-87.

73. Sillekens, P. T. G., Habets, W. J., Beijer, R. P. and van Venrooij, W. J. 1987. cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. EMBO J. 6:3841-48.

74. Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendewey, D., and Keller, W. 1988. Functions of the abundant U-snRNPs. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Max L. Birnsteil, Ed. Springer-Verlag. Berlin, Heidelberg, New York, London, Paris, Tokyo. pp 115-54.

75. Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. A Protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47:755-66.

76. Tollervey, D. and Guthrie, C. 1985. Deletion of a yeast small nuclear RNA gene impairs growth. EMBO J. 4:3873-78.

77. Tollervey, D., and Mattaj, I. W. 1987. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs EMBO J. 6:469-76.

78. Tollervey, D., Wise, J. A. and Guthrie, C. 1983. A U4-like small nuclear RNA is dispensable in yeast. Cell 35:753-62.

79. van Santen, V.L., and Spritz, R.A. 1987. Nucleotide sequence of a bean (Phaseolus vulgaris) U1 small nuclear RNA gene: Indications for plant pre-mRNA splicing. Proc. Natl. Acad. Sci USA. 84:9094-98.

80. Vankan, P., and Filipowicz, W. 1988. Structure of U2 snRNA genes of Arabidopsis thaliana and their expression in electroporated plant protoplasts. EMBO J. in press.

81. Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., et al. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicng pathway, particularly assembly of the spliceosome. EMBO J. 5:1683-95.

82. Wise, J. A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E. J., and Guthrie, C. 1983. Yeast contains small nuclear RNAs encoded by single copy genes. Cell 35:743-51.

83. Zhuang, Y., and Weiner, A. M. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell 46:827-35.

EPILOGUE

The last several years have witnessed dramatic advances in our understanding of the process of nuclear mRNA splicing (Peebles et al., 1986). The discovery that self-splicing group II introns undergo the same covalent reactions as nuclear pre-mRNA splicing has led to the widely held view that the primary role of the trans-acting components is to fold the pre-mRNA into an active configuration. Thus the burden of investigation has fallen to understanding how the splicing machinery identifies introns and assembles them.

The most accessible of the splicing factors have been the snRNPs. The general paradigm for their action has arisen from the finding that the U1 snRNP recognizes the 5' splice site via a basepairing interaction involving the 5' end of the U1 snRNA (Rogers and Wall, 1980; Lerner et al., 1980, Zhuang and Weiner, 1986). U2 has been shown to engage in a similar type of recognition event with the branchsite (Parker et al., 1986). Despite this progress, our understanding of the assembly map of the spliceosome even at this level (factor-site interactions) is still woefully inadequate. We are still determining how and when other snRNP-substrate interactions take place. This in turn will represent the groundwork for the final understanding of the splicing process: the regulation of the snRNP-substrate interactions and the rules that drive the higher order interactions which bring factor-site assemblies into the active configurations for the covalent steps of splicing.

In the past, the role of the U5 snRNP in splicing has been particularly difficult to assess for several reasons: 1) it is not

subject to oligonucleotide-directed RNAse H cleavage (Black et al., 1985), 2) there are no specific antibodies to its proteins, 3) while it is thought to interact with the polypyrimidine stretch via one of its proteins, this region of the substrate has been identified as the binding site for several other protein factors, complicating the interpretation of this finding. Nonetheless, there is an accumulating body of suggestive evidence implicating U5 as important for the late steps of the splicing reaction. In both mammalian and yeast systems, the U5 snRNP joins the spliceosome after the U1 and U2 snRNPs (Cheng and Abelson, 1987; Konarska and Sharp, 1987; Lamond et al., 1987)--and thus after commitment of the substrate to an assembly pathway (Ruskin et al., 1988). Therefore one would not anticipate it to have a primary role in intron identification, but rather in the mechanics of the splicing process or its resolution.

An association of the U5 snRNP with the polypyrimidine stretch in mammals was first suggested by correlating the micrococcal nuclease resistance of the U5 snRNA with that of a trimethylguanosine-capped species that associates with the polypyrimidine stretch (Chabot et al., 1985). Subsequent to this finding was the discovery that Intron Binding Protein or IBP, a protein factor that specifically interacts with the polypyrimidine stretch, is a component of the U5 snRNP (Tazi et al., 1986; Gerke and Steitz, 1986). However, unlike the case of the U1-5' splice site interaction, positing a binding site for U5 on the substrate does not immediately illuminate a role in the splicing process (Ruskin et al., 1988). This is because an early, essential role for this substrate sequence has already been demonstrated in mammalian systems--the

binding of U2AF, a protein factor required for U2 recruitment to the substrate. U2AF binds to the polypyrimidine stretch very early in the splicing process. What role, then, does the U5 snRNP play in intron recognition or splicing? Clearly, the most reasonable suggestion is that it binds to the polypyrimidine stretch to perform a role subsequent to that of U2AF. The role which presents itself for an element binding at the 3' splice site (given that the polypyrimidine stretch has already played a role in intron identification) is marking, manipulating or utilizing the 3' splice junction during 3' splice site cleavage/exon ligation.

However, given the essential early role of the polypyrimidine stretch in mammalian splicing, its hypothetical second role is inaccessible at this time. The yeast system offers a unique opportunity to investigate late roles of the polypyrimidine stretch in splicing. Beyond the oft-invoked "power of yeast genetics" is the clear phenomenon that yeast splicing has significantly different requirements at the 3' splice site. Spliceosome assembly and the first step of splicing occur in vitro on transcripts that terminate just 3' of the branchsite (Rymond et al., 1987). Obviously such experiments cannot easily be performed in vivo, but there is evidence that the same general situation holds. The two introns of the MATa1 gene have branchpoint-to-3' splice site spacings of 10 and 11 nucleotides (Miller, 1984)--clearly suggestive of introns lacking a polypyrimidine stretch altogether. Nonetheless, these introns are both recognized and removed by the yeast splicing machinery. Thus from the point of view of the yeast intron, the U2AF-polypyrimidine stretch interaction found in mammals is not critical to the

initiation of splicing. If the above conjecture about the temporal significance of the U5-polypyrimidine stretch interaction is true, then at least this aspect of U5 snRNP function should be dispensable for the first step of the splicing reaction. Our results in Chapter I go well beyond this prediction: the snRNA itself is dispensable for the first step of the reaction, at least for the actin intron.

The above represents the intellectual groundwork for the experiments described in Chapters II, III, IV and V: further characterization of the U5 snRNP and of the 3' splice site in yeast. Initial efforts at elucidating the role of the U5 snRNP by mutagenesis of the snRNA component were unproductive: while we now possess several U5 snRNA mutants that result in slowed growth and lethality at 30°, we have not been able to demonstrate a splicing defect in these mutants, much less define the mode of action of the snRNP. The ominous possibility is that the mutants curtail the splicing of only a small number of pre-mRNAs, eluding our detection. For this reason, and to broaden the possible range of U5-related defects we can recover, we initiated a screen designed to identify factors that exacerbate the defects caused by the U5 point mutations. Having demonstrated that this screen can recover _ts_ alleles of bona fide splicing factors, we are now in a position to begin intensive characterization of these gene products and to determine if any are actual components of the U5 snRNP. Identifying the protein components of the U5 snRNP is essential because: 1) the mammalian U5 particle contains at least one protein that directly interacts with substrate, 2) the mammalian snRNP appears to contain a tremendous amount of protein, consistent with a greater relative importance of

protein to function than the other snRNPs, and 3) the very limited sequence conservation of the U5 snRNA is not suggestive of extensive, essential roles for the RNA component of the snRNP.

The type and number of roles played by the U5 snRNP's putative binding site, the polypyrimidine stretch are also unknown. Earlier arguments in this chapter have suggested that the polypyrimidine element is dispensable for early splicing steps; there are those who feel the element is not even there. Nonetheless, as shown in Chapter III there was a clear statistical argument that yeast 3' splice sites are uridine-enriched, which lead us to propose a role for this element in 3' splice site identification during the second step of splicing. The work in Chapters IV and V is centered around actually demonstrating the existence and function of the yeast polyuridine stretch. In Chapter IV we showed that in a branchsite mutation (A->C) a uridine-rich 3' splice site allowed efficient conversion of the lariat intermediate to mature RNA, whereas an adenine-rich element accumulated the intermediate. There was no effect on the first step of splicing, again consistent with the importance of this element being restricted to the second step (in yeast).

The further characterization of 3' splice site selection in Chapter V provides the most definitive look at the actual role of the polypyrimidine stretch. First, we have been able to confirm our identification of a yeast polyuridine stretch in that the ability of a site to compete for selection by the splicing machinery is completely correlated with its adherence to the consensus derived in Chapter III. The failure of the MATa1-like splice site to compete effectively and of all the adenine-preceded proximal sites to shut

down the distal sites showed that the "first AG" rule applies only in the absence of a uridine-rich sequence preceding distal PyAG motifs. Nonetheless it remains true that the splicing machinery at some level can identify AG dinucleotides that are distant (30-60 ntes.) from the branchsite but not preceded by significant polyuridine runs.

The most important fruit borne of the demonstration of a polypyrimidine competitive system is the provision of a tool for understanding the factors involved in recognition and utilization of a 3' splice site. The example of the +A series is particularly instructive. While in the +A-PyDOWN construct the branchsite-proximal site is favored (presumably because neither site is pyrimidine-enhanced and the proximal site is simply closer), in the +A-PyUP construct the distal site is the primary selection. Thus by altering the strength of the signal in cis we have been able to drive 3' splice site choice. Just as varying the strength of the polypyrimidine stretch alters the ability of a branchsite-distal site to compete against a proximal, non-polyuridine enhanced site, one would anticipate that decreased efficiency in the machinery which recognizes and manipulates the polyuridine element should mimic "turning down" PyUP to PyDOWN. Mutants in the polypyrimidine machinery, can be sought using a construct which detects the product of the proximal splice site in +A-PyUP. Cells would then be mutagenized and screened for mutants which increase the activity from this site (see below).

Clearly we lack a full understanding of the role of the U5-polypyrimidine stretch interaction in splicing. Nonetheless, we have made significant inroads into understanding both elements

individually. The U5 snRNA is not absolutely required for the first step of splicing, but its depletion does not noticably alter the splicing patterns for any of the proximal constructs combined with the wild-type distal site (+0-WT, +A-WT, +T-WT, +AA-WT and +TT-WT). Preliminary work with the U5-A98 mutation does not show a significant alteration in the splicing of the +TT-PyUP, +0-PyUP, +AA-PyUP or +AA-A295 constructs.

Beyond examining the U5 snRNA itself, we are generating a set of mutants derived in a strain bearing the U5-C97,C99 mutation which can be tested in this assay system. As argued above, mutation in factors related to the U5 snRNA may well be a more productive place to search for splice site selection mutants. Finally, we are setting up to perform the mutant hunt described above which should give us mutants in the polypyrimidine stretch recognition machinery whether or not the U5 snRNP is involved (Dan Frank, pers. comm.). Clearly, any mutants recovered in this screen will be examined in light of the above theory

Finally, having spent the last 3 years attempting to develop a variety of genetic schemes to further the pursuit of an understanding of pre-mRNA splicing in yeast, it seems appropriate to ponder what seem to the the hard-won lessons regarding what does not and what might work. The difficulties with splicing genetics arise because the application of classical genetics to the problem of pre-mRNA splicing presents several novel problems beyond the adaptation of selection schemes to the splicing process. Classical genetics has been designed and tested largely on non-essential processes such as biosynthetic or regulatory pathways.

Pre-mRNA splicing, on the other hand, is an indispensable process. If any one of a large number of mRNAs is not produced, the cell will be inviable. The entire process takes place in a large macromolecular complex, and any given actor may play several roles and act in several steps. Omissions or defects amongst the actors often manifest as a lack of any observable complex assembly rather than giving rise to illuminating intermediates. Thus a wide variety of mutants have failed to be informative because they tell us nothing about _why_ they fail to splice pre-mRNAs.

The upshot of this reasoning is that we are seeking mutations that alter functions rather than knocking them out. Mutants that can tell us something about splicing are limited to those which can assemble the complex but fail to function within it, or which have aberrant behaviors within the context of the spliceosome. One important ramification of this concept is that this strikes severely into the target size for useful, splicing-related mutations. Given that any genetic screen will give rise to a background of non-specific revertants, this means that increasingly powerful secondary screens and more subtle/insightful primary screens will be required to sort the wheat from the chaff. However, the rewards will be mutants whose defects should inform us about the process of splicing, rather than reiterating the now established fact that a spliceosome assembly is easily blocked by alteration of its components.

There is also a related problem arising from the generality of the splicing process--a given defective spliceosomal component may be incompetent to splice only a subset of pre-mRNAs, reflecting their

different requirements or limitations. If these specific messages are not identified and examined, the defect will not be found, and its relevance to splicing will remain obscure. Indeed, the scenario of splicing machinery mutants with growth phenotypes but no known splicing defects has been played out for U1, U2, U5, and U6 (P. Siliciano, M. Ares, and R. Bordonne, pers. comm.). While in vitro systems offer partial solutions to many of these problems, they cannot substitute for the classical power of an open-ended genetic approach in terms of unveiling new components or new roles for previously identified ones.

The nature of the splicing process also limits the classical genetic approach of selecting suppressors that restore the activity of an altered gene product. In this case, the classical alteration is a point mutation in one of the conserved splicing signals in the intron. Empirically, such suppressors are few and far between. The heroic efforts of Tamm, Couto and Parker to suppress a host of such mutations yielded only one suppressor, the _rna_16-1 allele, which suppresses an A to C transversion at the branchpoint (Couto et al., 1987). While this mutant promises to be a powerful tool in understanding the splicing process, we must ask whether hindsight can provide us with insight for the future.

In general, there seem to be two factors conspiring to limit the availability of mutations in the splicing machinery that allow recognition of altered substrates. First, there is the problem of all the other spliced messages in the cell. These must be spliced with sufficient accuracy and efficiency to support viability in the

same cellular context that allows efficient splicing of mutant pre-mRNAs.

A second, more striking limitation on gain-of-function supressors of intron mutations arises from the demands on the splicing machinery to splice a novel substrate. Beyond the general concern that there may simply be no mutation which will restore a given interaction is the fact that many of the substrate signals play multiple roles in the splicing process. An excellent example of this problem is the A1 mutation of the actin intron. This mutation in the 5' splice site consensus sequence (GTATGT->ATATGT) results in the accumulation of precursor (45% of transcripts) and lariat intermediate (55%) (Vijayrhagavan et al., 1986). Much effort has gone into screening for suppressors of this mutation with no notable successes. One reason for this difficulty is now clear. While, as expected, the compensatory change in U1 that restores the basepairing suppresses this mutation, it does so only at the level of the first step of the splicing reaction--in the presence of the suppressor 75% of transcripts now progress to lariat intermediate, but no further (Siliciano and Guthrie, 1988). Thus assays for mature RNA cannot detect this change. The potential that this type of probelm will exist at other sites in the intron is extensive: in vitro, U1 requires UACUAAC sequences for binding, but this requirement does not arise from the characterized U2-UACUAAC recognition event (Ruby and Abelson, 1988). Alternately, the polypyrimidine stretch (at least in mammals) can be recognized by IBP (Tazi et al., 1986; Gerke and Steitz, 1986), 200Kd (J. Beggs and P. Sharp, pers. comm.), hnRNP C (Choi et al., 1986), and U2AF (Ruskin et al., 1988). Thus there is

no guarantee (and indeed, a powerful counterargument) that a given lesion can be healed by a single mutation in trans.

One well-studied system that poses genetic problems very similar to those of the spliceosome is the ribosome. The ribosome also represents a macromolecular complex that directs a multistep process on a tremendous variety of substrates. Unfortunately, an overview of the ribosomal genetics literature leaves one with the striking conclusion that one approach makes up the majority of ribosomal genetics: characterization of an astounding variety of naturally occurring antibiotics which specifically derange specific ribosomal functions. The latter technique and the isolation of resistance alleles appears to be the major window providing illumination of the ribosomes inner secrets. Unfortunately, this approach is not currently applicable to the study of pre-mRNA splicing. Thus we find that the attempt to do good, instructive genetics on the process of pre-mRNA splicing as a process is largely without precedent. It is the intent of this section to describe solutions that are designed (or proven) to address these issues.

One obvious alternative to gain-of-function mutations is to go the other direction--to seek mutations which impair rather than enhance the performance of a given splicing factor. This approach has two significant advantages: 1) Such mutations must exist, in that it is trivial to imagine myriad ways to "disimprove" a functioning apparatus. 2) The target size of possible mutations that reduce the efficiency or fidelity of a process is far broader than the target size for specific alterations which allow novel recognition properties.

HOWEVER, the search for useful down mutations contains its own pitfalls (and hopefully, a cognate set of remedies). First and foremost among these is limiting mutations to the process of interest.

How could one then pursue greater specificity in the search for down mutations? There are two general approaches we have derived that seem to offer promise. The first entails employing a competitive system. Generally speaking, a competitive system circumvents most of the problems because decreased function of one element, rather than resulting in loss of product per se, appears as increased use of the second competing pathway. Thus by assaying for increased use of an alternative pathway much of the non-specific background of revertants is eliminated. Any mutations that non-specifically impair splicing should affect both pathways simultaneously, and thus will not be recovered. Instead, recovery of mutants with partial or altered function is emphasized. Finally, since the appropriate mutations are anticipated to be recessive, a rapid second screen is to test whether mating to a wild type cell restores the initial phenotype.

To illustrate this idea, I suggest two (currently) hypothetical examples. The first involves the polypyrimidine stretch. As suggested earlier in this chapter, by making a construct in which both 3' splice sites are in cis with the same branchpoint sequence, one creates a competitive situation vis a vis 3' splice site choice. Since both sites are used to some measurable extent one can screen for mutants that favor the proximal site, presumptively by decreasing the efficiency of the distal, polypyrimidine enhanced site.

Technically, this assay is easily implemented because the two splice sites result in translation of the 3' exon in two different reading frames. By fusing a reporter gene in frame with the proximal site, one can assay its absolute usage, which reflects the ability of the distal site to be a successful competitor. Therefore, by screening for mutations which increase net utilization of the proximal site, we are in a position to isolate down mutations in the machinery that is required for efficient utilization of the distal, polypyrimidine enhanced 3' splice site. Hopefully, this provides us with a much broader target size (because we are searching for non-specific or down mutations in a process) while demanding partially functional products (the proximal intron must be completely spliced, and other introns in the cell, that have distal 3' splice sites, must be spliced with reasonable efficiency to support viability). Further, since we are dealing with a competitive system, subtle changes in the actual functions of the machinery should be amplified because of the possible alternative event.

A second strategy which is being developed in the lab is designed to approach the problem of recruitment of U2 to the spliceosome. The initial impetus for this project comes from the work of Felder et al., (1987) who observed that in the yeast actin intron, if the cryptic TACTAAG sequence (14 nucleotides 5' of the functional branchpoint sequence) is repaired to TACTAACA, and the normal branchpoint sequence altered to TACTAAG, then both branchsites are utilized roughly equally. Further, and equally important, they were able to determine that half the lariats formed at the internal branchsite result in 3' splice site utilization at the AG

dinucleotide in the distal branchsite (TACTAAG). Clearly, all lariats formed at this AG are products of the internal branch sequence. Thus this construction represents a situation in which branchsite formation can be assayed by 3' splice site selection. Presumptively, utilization of the newly created AG is directly proportional to the utilization of the internal branchsite. As before, this product can be uniquely assayed by fusing a reporter gene in frame with this AG but out of frame with the distal AG.

How can this assay system be used to study the recruitment of U2 snRNAs to the spliceosome? If one accepts the current model that U2 snRNAs bind to the spliceosome independent of their ability to interact with the branchsite (which presumptively occurs subsequently) then by designing U2 snRNAs which can uniquely interact with one or the other branch sequence one can assay for their competitive ability to bind to the spliceosome. By using an informational suppressor derived from U2 (the U2-U37 mutation) and altering the internal TACTAAC sequence to the cognate site for U2-U37 (UAAUAAC) we can hope to establish a system which in essence reports the efficiency of recruitment of U2-U37 to the spliceosome. Following mutagenesis of U2-U37, we can screen for mutations which decrease utilization of the proximal site. Help, I'm a bug! To screen out non-specific down mutations of the U2-U37 allele, we can also screen candidates for the ability of U2-U37 to function on a substrate which requires the supressor (i.e., has a UACUAAC->UAAUAAC mutation) but which does not have a competitive alternative. In this way we can hope to isolate mutations which specifically alter the ability of U2 to join the spliceosome.

We have also begun developing a second, alternative approach to competitions which we hope will yield down mutations restricted to "interesting" splicing mutations rather than non-specific alterations. The general strategy here is to "specify" the site or nature of the new mutation by developing an assay system around a pre-existing mutation in the pathway of interest. More concretely, we are attempting to isolate mutations relevant to the function or biogenesis of the U5 snRNP. By starting with defective U5 snRNPs, we hope to limit our scope of secondary "down" mutations to those related to U5 function. The screen is designed to ask for mutations that can support viability in the presence of the wild type U5 allele, but cannot do so when placed in conjunction with impaired (but non-lethal) versions of the U5 snRNP. In effect these are a novel form of conditional allele, that responds to the presence/absence of the wild-type U5 rather than external stimuli such as temperature. While work on this project is preliminary (and hampered by the fact that we lack a clear idea of the function of the U5 snRNP), we have preliminary evidence that at least some mutations confer splicing defects, whereas others are involved in an as yet undetermined step in mRNA biogenesis (Dan Frank, pers. comm.).

REFERENCES Black, D.L., Chabot, B., and Steitz, J.A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell 42:737-50.

Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science 230:1344-49.

Cheng, S.-C. and Abelson, J. 1987. Spliceosome assembly in yeast. Genes and Develop. 1:1014-27.

Choi, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. Science 231:1534-40.

Couto, J.R., Tamm, J., Parker, R. and Guthrie, C. 1987. A trans-acting supporessor restores splicing of a yeast intron with a branch point mutation. Genes and Develop. 1:445-455.

Felder, E., Cellini, A and Rossi, J. 1987. Oligonucleotide mediated mutational analysis of recognition sequence effects on lariat formation during messenger RNA splicing in yeast. Nucleosides and Nucleotides 6:409-11.

Gerke, V. and Steitz, J.A. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47:973-84.

Konarska, M.M. and Sharp, P.A. 1987. Interaction between snRNP particles in formation of spliceosomes. Cell 49:763-4.

Lamond, A.I., Konarska, M.M., Grabowski, P.J. and Sharp, P.A. 1988. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc. Natl. Acad. Sci. USA 85:411-15.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. 1980. Are snRNPs involved in splicing? Nature 283:220-4.

Parker, R., Siliciano, P. and Guthrie, C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves basepairing to the U2-like snRNA. Cell 49:220-39.

Peebles, C.L., Perlman, P.S., Kecklenburg, K.L., Petrillo, M.L., Tabor, J.J., Jarrell, K.A., and Cheng, H.-L. 1986. A self splicing RNA excises an intron lariat. Cell 33:509-18.

Rogers, J. and Wall, R. 1980. A mechanism for RNA splicing. Proc. Natl. Acad. Sci. USA 77:1877-79.

Ruby, S.W. and Abelson, J. 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science 242:1028-1035.

Ruskin, B., Zamore, P.D., and Green, M.R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52:207-19.

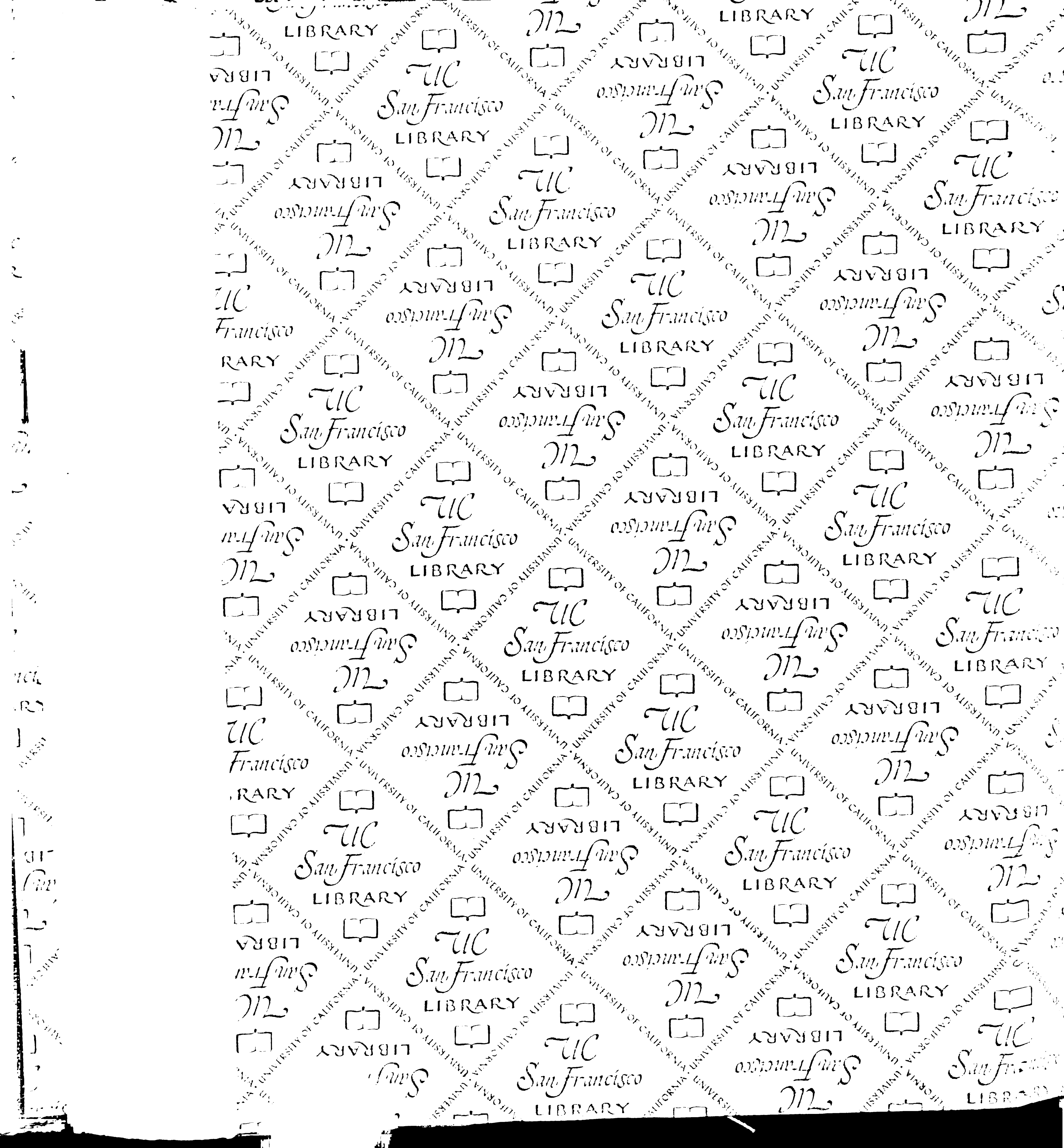
Rymond, B., Torrey, D. and Rosbash, M. 1987. A novel role for the 3' region of introns in pre-mRNA splicing of Saccharomyces cerevisiae. Genes and Develop. 1:238-46.

Siliciano, P.G. and Guthrie, C. 1988. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes and Develop. 2:1258-67.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47:755-66.

Zhuang, Y., and Weiner, A.M. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell 46:827-35.

Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J. and Guthrie, C. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J. 5:1683-95.


UC
San Francisco
LIBRARY
UNIVERSITY OF CALIFORNIA

FOR REFERENCE

NOT TO BE TAKEN FROM THE ROOM