Architecture of the U5 Small Nuclear RNA

DANIEL N. FRANK,[†] HELI ROIHA, AND CHRISTINE GUTHRIE*

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received 1 October 1993/Returned for modification 9 November 1993/Accepted 9 December 1993

We have used comparative sequence analysis and deletion analysis to examine the secondary structure of the U5 small nuclear RNA (snRNA), an essential component of the pre-mRNA splicing apparatus. The secondary structure of *Saccharomyces cerevisiae* U5 snRNA was studied in detail, while sequences from six other fungal species were included in the phylogenetic analysis. Our results indicate that fungal U5 snRNAs, like their counterparts from other taxa, can be folded into a secondary structure characterized by a highly conserved stem-loop (stem-loop 1) that is flanked by a moderately conserved internal loop (internal loop 1). In addition, several of the fungal U5 snRNAs include a novel stem-loop structure (ca. 30 nucleotides) that is adjacent to stem-loop 1. By deletion analysis of the *S. cerevisiae* snRNA, we have demonstrated that the minimal U5 snRNA that can complement the lethal phenotype of a U5 gene disruption consists of (i) stem-loop 1, (ii) internal loop 1, (iii) a stem-closing internal loop 1, and (iv) the conserved Sm protein binding site. Remarkably, all essential, U5-specific primary sequence elements are encoded by a 39-nucleotide domain consisting of stem-loop 1 and internal loop 1. This domain must, therefore, contain all U5-specific sequences that are essential for splicing activity, including binding sites for U5-specific proteins.

Although the U5 small nuclear ribonucleoprotein (snRNP) is an essential cofactor in pre-mRNA splicing, its precise function is unknown. The U5 snRNP has long been hypothesized to play a role in the second catalytic step of splicing, perhaps as a factor required for the recognition of the 3' splice site (6, 28, 38). Indeed, we have recently identified several proteins that genetically interact with the U5 small nuclear RNA (snRNA) and are required for the second step of splicing (8). One of these proteins, the product of the *SLU7* gene, functions in the selection of distant 3' splice sites (7). Whether any of these proteins is an integral U5 snRNP protein remains an open question; nevertheless, they provide strong evidence linking U5 function to the recognition and utilization of the 3' splice site. Moreover, several studies indicate that conserved sequences in the U5 snRNA, itself, can interact with exon sequences at the intron-exon junctions. Newman and Norman (25) have demonstrated that mutations in the *Saccharomyces cerevisiae* U5 snRNA can suppress mutations in both 5' and 3' splice sites, suggesting that the U5 snRNA can form hydrogen bonds with both 5' and 3' exon sequences. In agreement with this finding, a subset of the U5 nucleotides implicated in the 5' splice site-U5 interaction have been cross-linked to exonic and intronic RNA sequences near the 5' splice site, indicating a close, physical juxtaposition of U5 and 5' splice site nucleotides during splicing (41, 44). Taken together, these studies suggest that the U5 snRNA plays an active role in the function of this snRNP.

Aside from the few nucleotides implicated in the exon-U5 interaction, only the conserved Sm core protein binding site of U5 has a defined function (13, 14, 21). Little else is known about how the structure of the U5 snRNA relates to its function. Consequently, we have sought to define the structural elements that make up the U5 snRNA and assess their

functional significance. An important first step towards this goal is the determination of U5 secondary structure, which is most effectively accomplished through phylogenetic analysis; this approach delineates structures that are presumed, by their evolutionary conservation, to be functionally significant (27). Hypotheses concerning function can then be readily tested by genetic and biochemical means, in the appropriate system.

Existing models of U5 secondary structure have been derived and tested primarily in mammalian systems (1, 3). The gene encoding the *S. cerevisiae* U5 snRNA, *SNR7*, was initially identified on the basis of a 9-nucleotide (nt) sequence element (nt 93 to 101) that is invariant in all U5 snRNA sequences (29). Aside from short regions flanking this element and the conserved Sm protein binding site, the remainder of the *S. cerevisiae* sequence could not be reliably aligned with other, highly divergent U5 sequences. Thus, a phylogenetically rigorous secondary structure could not be established. To broaden the scope of U5 structural studies, we have begun to analyze the U5 snRNAs isolated from a variety of fungal species, beginning with representative yeast species. Our reasons for focusing upon the yeasts are threefold: (i) the yeasts are phylogenetically diverse (indeed, they are a polyphyletic group spanning several phyla of fungi [5]); (ii) yeasts are only moderately divergent from plant and animal taxa, so meaningful sequence alignments including all three groupings should be possible; and (iii) several yeast species, particularly *S. cerevisiae*, offer powerful genetic systems with which to test structural models. In this paper we present five new U5 snRNA sequences obtained from ascomycetous yeast species and compare them with other plant, metazoan, and fungal U5 sequences.

We show that fungal U5 snRNAs, like their counterparts from other taxa, can be folded into a secondary structure characterized by a highly conserved stem-loop (stem-loop 1 [SL1]) that is flanked by a moderately conserved internal loop (internal loop 1 [IL1]). Several of the fungal U5 snRNAs include a novel stem-loop structure (ca. 30 nt) that is adjacent to SL1. Furthermore, we have taken advantage of the *S. cerevisiae* genetic system in order to examine, via deletion analysis, the functions of the structural regions defined by phylogenetic analysis. Our goal was to define the minimal RNA structure

* Corresponding author. Mailing address: Department of Biochemistry and Biophysics, University of California, San Francisco, 513 Parnassus Ave., San Francisco, CA 94143-0448. Phone: (415) 476-2321. Fax: (415) 476-0943. Electronic mail address: Guthrie@cgl.ucsf.edu.

[†] Present address: Department of Biology, Indiana University, Bloomington, IN 47405.

that provides U5 function, the essence of the U5 snRNA. The minimal U5 snRNA that can complement the lethal phenotype of a U5 gene disruption consists of (i) SL1, (ii) IL1, (iii) a stem-closing IL1, and (iv) the conserved Sm protein binding site. Remarkably, all essential, U5-specific primary sequence elements are encoded by a 39-nt domain consisting of SL1 and IL1. This domain must, therefore, contain all U5-specific sequences that are essential for splicing activity, including binding sites for U5-specific proteins.

MATERIALS AND METHODS

Strains. *Saccharomyces darensis* (NRRL Y-12639) and *Pachysolen tannophilus* (NRRL Y-2460) were provided by Cletus Kurtzman (USDA Northern Regional Research Center). *Saccharomyces exiguus* (IH925), *Saccharomyces kluyveri* (IH400), and *Pichia canadensis* (IH207) were provided by Ira Herskowitz (University of California, San Francisco). *Kluyveromyces lactis* (ATCC 32143) was obtained from the American Type Culture Collection. U5 snRNA deletion mutations were screened in a haploid *S. cerevisiae* strain, Δ SNR7::LEU2, in which chromosomal SNR7 was disrupted by LEU2 (14). The lethal SNR7 deletion was complemented by either the 2 μ m plasmid pGAL1-SNR7, which carries a GAL1-SNR7 fusion, or the centromeric plasmid pBWHis600, which carries the SNR7 gene under the control of its own promoter (29). Growth assays were performed in the former strain, while RNA analysis utilized the latter.

Plasmids. The following plasmids were employed in this study. pSE360 was the gift of S. Elledge (Baylor College of Medicine, Houston, Tex.). pSP-U5 (human U5 gene) was the gift of D. Black (University of California, Los Angeles). The *K. lactis* genomic library (on pBR322) was provided by J. Shuster (Chiron, Inc.). Bluescript II SK(+) was purchased from Stratagene. Plasmid pBS-7EB was created by subcloning a *HpaI*-*NcoI* fragment containing the SNR7 gene into the *SmaI* site of pUC19. The *EcoRI*-*BamHI* fragment of this clone, which contains the SNR7 gene, was subsequently subcloned in the *EcoRI*-*BamHI* sites of Bluescript II to create pBS-7EB. pBS-7 Δ PVU was created by oligonucleotide-directed deletion of the sequence encoding nt 7 to 141 of U5; this deletion creates a *PvuII* site consisting of nt 4, 5, 6, 142, 143, and 144 of U5.

Oligonucleotides. The following oligonucleotides were used: T-oligo, 5'-CCTCGAGGCC(T)₁₇; PCR-U5, 5'-GGGGATCCA CATGGTTCTTGCCTTTTAC; SNR7, 5'-AAGTTCCTCAAAA ATATGGCAAGC; SNR14, 5'-AAGGTATTCCTCAAAAATTC CCTAC; Δ 3, 5'-AAATATGGCAAGCGCAGCTTTACCTGT; 7 Δ PVU, 5'-CCCACAGTAACGGACAGCTGCTTCAGGT TTT; 7DBW, 5'-GAGAAACCCAGAGTACTACGATCCTC TA; Δ VSL, 5'-CATGTTCTGTTATGATGTTGACC; YH.B, 5'-GGCGAAAGATTGTTTCGTTATTGAAGAGAAACC; Δ L1, 5'-CCCGGATGGTTCTCTCTAGAACCATTGTCG; Δ IL1.3', 5'-GGAGACAACACCATGGTTCTGGTAA; Δ IL 1.5', 5'-GGCAAGAACCATGATGTTGACCTCCC; Δ 4, 5'-CC ATGTTCTGTTATGATCTGTAAAGC; Δ ISR, 5'-GACAGCT TTACAGATCCGGATGGTTCTG; YH.C, 5'-GCAAGAACC ATATACGATCGATGTTGACCTC; MIN1, 5'-ATAACGAA CATGGTTCTTGCCTTTTACCAGAACCATCCG; REV1, 5'-CCGATGGTTCTCTGGTA; MIN2, 5'-TCATAACGAACATG GTTCTTGCCTTTTACCAGAACCATCCGGA; REV2, 5'-T CCGGATGGTTCTGG; MIN3, 5'-AGATCATAACGAAC ATGGTTCTTGCCTTTTACCAGAACCATCCGGATCT; REV3, 5'-AGATCCGGATGGTTC; MIN4, 5'-TACAGATCA TAACGAACATGGTTCTTGCCTTTTACCAGAACCATC CGATCTGTA; and REV4, 5'-TACAGATCCGGATGG.

Bacterial and yeast genetic methods. All yeast genetic

techniques, including transformations and media, were as described by Guthrie and Fink (10). Bacterial transformations, oligonucleotide kinasing, and basic DNA cloning techniques were as described by Maniatis et al. (20).

Sequencing. (i) *K. lactis*. The gene for *K. lactis* U5 was cloned as described in Roiha et al. (33) and sequenced by dideoxy DNA sequencing (34).

(ii) **All other species.** U5 snRNA sequences were determined by two methods: (i) PCR amplification of 3' halves of U5 molecules and (ii) oligonucleotide-directed RNA sequencing of 5' halves of U5 molecules. RNA sequencing was by the protocol of Roiha et al. (33).

For the PCR amplification, snRNAs were purified from total cellular RNA (43) by immunoprecipitation with anti-trimethyl guanosine cap antibodies (32). Precipitated snRNA (15 μ g) was poly(A) tailed with Poly(A) polymerase (Bethesda Research Laboratories) for 10 min at 37°C, under the manufacturer's reaction conditions. Poly(A)-tailed RNA was then primed with the oligonucleotide T-oligo and replicated with reverse transcriptase (by the Bethesda Research Laboratories protocol, at 37°C for 60 min). Second-strand synthesis was carried out by PCR amplification of 1/20 of the sample recovered from primer extension. PCRs, with oligonucleotides PCR-U5 and T-oligo, followed the protocol provided with *Taq* polymerase (Amersham). Thirty PCR cycles were carried out as follows: (i) 1 min at 94°C, (ii) 1 min at 42°C, and (iii) 1 min at 72°C. Following amplification, overhanging ends were filled in with *Taq* polymerase at 72°C for 15 min. One-third of each PCR sample was phenol-chloroform extracted, ethanol precipitated (with glycogen carrier), and then digested with *BamHI* and *XhoI*. Digested fragments were isopropanol precipitated (to separate large DNA fragments from smaller cleavage products) and then ligated with *BamHI*-*XhoI*-digested Bluescript(+) plasmid. Recombinant plasmids were transformed into DH5 α -competent cells (Bethesda Research Laboratories) and selected on Luria-Bertani plates plus ampicillin, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal), and isopropyl- β -D-thiogalactopyranoside (IPTG). Positive transformants were sequenced by dideoxy DNA sequencing (34).

Deletion mutagenesis. (i) **Oligonucleotide-directed mutagenesis.** Deletion mutants Δ VSL, Δ IL1.5', Δ IL1.3', Δ L1, Δ 3, Δ 4, and MIN5 were constructed via oligonucleotide-directed mutagenesis of plasmid pBS-7EB (for Δ VSL) or pBS- Δ VSL (all other deletions) (22). The nucleotides deleted by each oligonucleotide are as follows: Δ VSL, nt 41 to 74; Δ IL1.5', nt 75 to 83; Δ IL1.3', nt 111 to 113; Δ 3, nt 146 to 157; and Δ 4, nt 18 to 40. In construct Δ L1, nt 92 to 102 are replaced by the sequence GAGA.

(ii) **Bal 31 deletion.** Deletion mutants Δ 1 and Δ 2 were created by digestion of pBS-7EB with the exonuclease *Bal* 31. Mutant Δ (1+2) was created by ligating the *KpnI*-*HpaII* fragment of construct Δ 1 with the *HpaII*-*BamHI* fragment of construct Δ 2. The nucleotides deleted in each construct are as follows: Δ 1, nt 7 to 27; Δ 2, nt 126 to 141; and Δ (1+2), nt 7 to 27 and 126 to 141.

(iii) **MIN series construction.** MIN constructs were created by inserting synthetic DNA fragments encoding SL1, IL1, and stem 2 (S2) of MIN1 through MIN4 into the *PvuII* site of plasmid pBS-7 Δ PVU. To create the double-stranded DNA, oligonucleotides REV1 through REV4 were used to prime second strand synthesis templated by oligonucleotides MIN1 through MIN4. Template and primer oligonucleotides (200 pmol of template and 50 to 70 pmol of primer) were mixed in annealing buffer (200 mM Tris-HCl [pH 7.9], 20 mM MgCl₂, 500 mM NaCl), heated at 68°C for 3 min, and then rapidly chilled on ice (23). To this mixture was added 2 μ l of 10 $\times$

	Stem 3	Loop 2	Stem 2	Variable Stem/Loop	
		20	40	60	
1 saccar	-----AA GCAGCUUUA	AGAUCAAUGGCGGAG	GGA-GGUCAACAUC	AAGAACUGUGGGC-----CUUUUAUU----	
2 sacklu	-----CG CCCUCUCU-	-----UCUUUAGC	AGGAGAUCAUAUC	AGAAAGAGCGGCAGUU-----CUAU-----	
3 klulac	-----AGAG AUUGACGUC-		UGGCGAUCAGUAUC	AAUGGGGGCGUCCAG-----CUUUUUA-----	
4 pactan	-----C GACCA-		-GAAGAUCAAAUUC		
5 piccan	-----G AGCCAU-		-GGAGAUCAAAUUC		
6 sacexu	-----?CUUUAC	AGAUCAAUGGCGGAG	-GGAGGUCAACAUC	AAGAACUGUGGGC-----C??UUAUU----	
7 schpom	-----AUAUCC GUCAAA-	-----GC	-----ACUUUGCAAA		
8 aratha	GGGAGUAAAAUCAC	GCA-----GCCAUG	-----UGGUGAGUAACA		
9 pissat5a	-----GGA GCCG-----	-----UG	-----UGAUGAUGACA		
10 crycoh	-----GAUC ACA-----	-----G	-----UGUUCACUUCA		
11 tetpyr5a	-----AUC ACA-----	-----G	-----AACUCAGCUCA		
12 tetthe51	-----AUC ACA-----	-----G	-----AACUCAGCUCA		
13 caeele	-----AAUCA ACUC-----	-----UG	-----GUUCCUCUGCA		
14 dromel	-----AU ACUCU-----	-----G	-----GUUUCUCUUCA		
15 xenlae	-----AU ACUC-----	-----UG	-----GUUUCUCUUCA		
16 galgal5arn	-----AU ACUC-----	-----UG	-----GUUUCUCUUCA		
17 musmus5	-----GAU ACUC-----	-----UG	-----GUUUCUCUUCA		
18 ratnor5a	-----AU ACUC-----	-----UG	-----GUUUCUCUUCA		
19 homsap5b	-----AU ACUC-----	-----UG	-----GUUUCUCUUCA		
Cons.			u.....y.		
Yeasts 1-7			grUcRy.auc		
Plants 8-9	caa GCA	UG	UGaUGAggACA		
Prot. 10-12	AUC ACA	G	aacUCAgcUCA		
Met. 13-19	au ACUC	G	GUUuCUcUuCA		

	Variable Stem/Loop	Inner Loop 1	Stem/Loop 1	Inner Loop 1	Stem 2	Inner Loop 2
		80	100		120	
1 sacc	---GCCUAUAGAACUU	--AUAACGAAC	AUG-GUUCU.UGCCUUUUACC.AGAAC-CAU	CCG	GGUGUUGUCUCC-	AUAGAAACAG
2 sacklu	---AACUGUCGCCAAUCU	--AUAUCGAAC	AUG-GUUCU.UGCCUUUUUACC.AGAAC-CAU	CCG	GGUACUGCUCUUU	CAUACAAUGU
3 klulac	CUGGUUUUGCCUCUGU	-----UCGAAC	AUG-GUUCU.UGCCUUUUUACC.AGAAC-CAU	CCG	GGUACUGUCCCA-	ACUAUG----
4 pactan	-----UUUAACGUAG	AUG-GUUCU.UGCCUUUUUACC.AGAGC-CAU	CCG	GGUACUGUCUCC-	AUACCAA----	
5 piccan	---AAUUUAAAA?CUC	--AUAUCGAUC	AUG-C-?CU.UGCCUUUUUACC.AGAG-CAU	CCG	GGUACUGUCUCC-	ACUCCAA----
6 sacexu	---GCUUAUAGAACUU	--AUAACGAAC	AUG-GUUCU.UGCCUUUUUACC.AGAAC-CAU	CCG	GGUGUUGUCUCC-	AUAGAAACAG
7 schpom	-----AGCUAACGUUAU	CUG-UUUUCU.UGCCUUUUUACC.AGAAA-CAG	CCG	-UUUGUAAGGU--	GUGCUAA----	
8 aratha	-----AAGCGAAC	UAUUUUUUU.CGCCUUUUUACU.AAAGA-AUA	CCG	UGUGCUC-UCG--	ACGCUAAG--	
9 pissat5a	-----UAGCGAAC	UAU-UCUUU.CGCCUUUUUACU.AAAGA-AUA	CCG	UGUCAGCGUCA--	CAAUUAG--	
10 crycoh	-----ACCGAAU	CAA-UCUUU.CGCCUUUUUACU.AAAGG-UUG	CCG	UGAAUGGGGACA--	CAUCAA----	
11 tetpyr5a	-----AUACGCUU	UAA-UUUUU.CGCCUUUUUACU.AAAGA-UUA	CCG	UGGGCUGGGUU--	CUACAA----	
12 tetthe51	-----UUACGCUU	UAA-UUUUU.CGCCUUUUUACU.AAAGA-UUA	CCG	UGGGCUGGGUU--	UACCAA----	
13 caeele	-----UUUAACCGUGA	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGCAAAGGAGC--	AUUUACU----	
14 dromel	-----AUGUCGAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAAC--	ACUCUA----	
15 xenlae	-----AAUUCGAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAAC--	GACCAU----	
16 galgal5arn	-----GAUCGUUAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAAC--	AACUCU----	
17 musmus5	-----GAUCGUUAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAAC--	AACUCU----	
18 ratnor5a	-----GAUCGUUAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAAC--	AACUCU----	
19 homsap5b	-----GAUCGUUAU	AAA-UCUUU.CGCCUUUUUACU.AAAGA-UUU	CCG	UGGAGAGGAU--	AACUCU----	
Cons.	CG.y	yUyU.YGCCUUUYAYY.ARAR.	CCG	.g		
Yeasts 1-7	CG.ac	aUG.....CU.UGCCUUUUUACC.AGAR..CAU	CCG	.gu.yugucyy..	ry...a...	
Plants 8-9	aAGCGAAC	UAU.UCUUU.CGCCUUUUUACU.AAAGA.AUA	CCG	UGUcagC.UCa..	aaacUaa...	
Prot. 10-12	uaCGcuU	uAA.UUUUU.CGCCUUUUUACU.AAAGa.UUA	CCG	UGggcuGGguu..	cayCAA....	
Met. 13-19	rruCG.au	AAA.UCUUU.CGCCUUUYyYy.AAAGA.UUU	CCG	UGgAgAGGAac..	a.yy.....	

	Stem 3		Sm Site		3' Stem/Loop
	140	160	180	200	
1 saccar	GUAAAGCUGU	CCGUUACUGUGGGCUUGCCA----	UAUUUUUUGG	AACUUUUC--	UGCCUUUUUUCUCAAUUG---AGUAAGGAGGGCG
2 sacklu	GGGAUGAGGG	AAGAUCGUGCUUCG-----	UAUUUUUUGG	AACUUUU--	
3 klulac	GACUCAAU--	ACUCUGUCU-----	AAUUUUU--GG	CAGCCUU--	
4 pactan	-UGGUC----	UAUCCU-----	AGUUUUU--GG	AAACC-----	
5 piccan	AUGGUU--	GUGUGAUUA-----	CAUUUUUUGG	AAACCC-----	
6 sacexu	GUAAAGCUGU	CCGUUACUGUGGGCUUGCCA----	UAUUUUUUGG	AACUUUUC--	UGCCUUUUUUCUCAAUUG---AGUAAGGAGGGCG
7 schpom	UUUGAC----	UGUAU-----	AGUUUUU--GA	AUC-----	UUUU-----UCUU-----GAAA
8 aratha	UGC-----	AUACGCC-----	UAUUUUU--GG	A-----	GGGCUCCAC--UUUCU-----GUGGAACCC
9 pissat5a	CGGC-----	AUACGCU-----	AGUUUUU--GG	A-----	AGAGUUCUCA--AUUU-----UGAGGGCUCU
10 crycoh	UGU-----	GAAUCUCUC-----	AAUUUUU--GA	A-----	GGG-----CUCUGC-----CCC
11 tetpyr5a	UGUGA----	AUUUUUAA-----	AAUUUUU--GA	A-----	GGAUU-----GUGUG-----AAUCC
12 tetthe51	UGUGA----	AUUUUUAA-----	AAUUUUU--GC	A-----	GGAUUC-----UUUU-----GAAUCC
13 caeele	-GAGU-----	AUUACAUA-----	AAUUUUU--GG	A-----	GACUC-CUU--GAGA-----AAGCGGGUC
14 dromel	AGAGU-----	CUAAAAACU-----	AAUUUUUUGAG	A-----	UCAGUCUUG--UCG-----CAAGACUGG
15 xenlae	-GAGU-----	UUUCGUUC-----	AAUUUUUUGA	A-----	GCCUGGU--UCAC-----ACAGGU
16 galgal5arn	-GAGU-----	CUUAAACC-----	AAUUUUUUGA	A-----	GCCUUGU--UCCG-----GCAAGGC
17 musmus5	-GAGU-----	CUUAAACC-----	AAUUUUUUGA	A-----	GGCCUUGU--CUUG-----GCAAGGCU
18 ratnor5a	-GAGU-----	CUUAAACC-----	AAUUUUUUGA	A-----	GCCUUG--UUCCGG-----CAAGGC
19 homsap5b	-GAGU-----	CUUAAAGCU-----	AAUUUUUUGA	A-----	GCCUUG--CUCCGA-----CAAGGC
Cons.			.RUUUUU.gr		
Yeasts 1-7	...r.....	...y.....	.RUUUUU.Gg	aa.....	aGacUcCaC...aUcU.....GaGgaacCc
Plants 8-9	cGC.....	AUACGCc.....	aaUUUUU.GG	A.....	Gga.....yUyU.....ucc
Prot. 10-12	UGU.....	auuauUaa.....	AAUUUUU.Ga		g.cyy.....yy.....rrg.y
Met. 13-19	.GAGU.....	yUya.rc.....	AAUUUUU.gr		

synthesis buffer (100 mM Tris-HCl [pH 7.4], 50 mM MgCl₂, 20 mM dithiothreitol, 5 mM each deoxynucleoside triphosphate) and 1 to 2 U of T4 DNA polymerase (New England Biolabs). Extension reactions were carried out at 37°C for 60 to 90 min, and reaction products were separated on nondenaturing 20% polyacrylamide gels. DNA was eluted by incubating crushed gel slices in elution buffer (500 mM ammonium acetate, 1 mM EDTA) overnight. The eluate was phenol-chloroform extracted once, and the DNA was ethanol precipitated. The MIN DNA fragments were then ligated with 1 µg of pBS-7ΔPVU DNA that had been linearized by *Pvu*II digestion.

(iv) **Chimeric U5 construction.** A portion of the human U5 gene (nt 1 to 85) was first subcloned into the *Hind*III and *Eco*RI sites of Bluescript to create pBS-HU5. A *Sca*I restriction site was then created (by oligonucleotide mutagenesis with oligonucleotide 7DBW) by insertion of a guanosine nucleotide between positions 1 and 2 of the U5 coding region. The *Xba*I-*Hin*FI fragment encompassing the human U5 sequence was blunted with T4 DNA polymerase and inserted into vector pBS-7ΔPVU, which had been linearized by digestion with *Xba*I and *Pvu*II. This construct fuses human U5 nt 3 to 78 to the yeast U5 sequence 142 to 215 (and yeast 3' sequences). The *Sca*I-*Bam*HI fragment encompassing this chimera was then subcloned into vector pBS-7ΔPVU, which had been linearized by digestion with *Pvu*II and *Bam*HI. This construct, pBS-YH.A, fuses the upstream region of yeast U5, ending at position 6, to nt 3 of human U5. The final, chimeric U5 coding region is shown in Fig. 5.

(v) **U5 expression.** All of the U5 variants described above were constructed in Bluescript(+) vectors. To express the U5 snRNAs in yeast cells, *Eco*RI-*Bam*HI fragments containing the mutant U5 alleles were subcloned into the *Eco*RI-*Bam*HI sites of pSE360 (a centromeric plasmid carrying the *URA3* gene). These plasmids were then transformed into strain Δ*SNR7* (p*GAL1-SNR7*) for analysis of growth phenotypes or strain Δ*SNR7* (pBWHis600) for RNA analysis.

Growth assays. Doubling times of strains homozygous for U5 deletion mutant alleles were determined by monitoring the optical density at 600 nm of log-phase cultures grown in liquid yeast extract-peptone-dextrose (YEPD) medium at 30°C. For each culture, a graph of time versus ln(optical density at 600 nm) was plotted; the slope of such a plot gives the value of ln(2)/(doubling time). For each mutant, doubling times were determined from three independent transformants. Mean doubling times are reported in the text and figures.

Plate growth assays were accomplished by streaking strains heterozygous for p*GAL1-SNR7* and deletion mutant U5 genes for single colonies on YEPD plates at 18, 25, 30, 33, and 37°C. Cellular growth was scored by noting the growth rates of individual colonies relative to those of positive and negative controls streaked on the same plate. The plate growth assays whose results are shown in Fig. 4 and 5 represent cultures patched onto YEPD plates at the specified temperatures. For each patch, an approximately equivalent amount of cells was suspended into 100 µl of sterile TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). The results of this assay can be

interpreted only qualitatively and are presented simply as a survey of the temperature effects on U5 deletion mutant-bearing strains. Three independent transformants in each plate assay were scored, and identical results were observed within each set of mutants.

RNA analysis. Yeast strains heterozygous for wild-type and deletion mutant U5 genes were grown on minimal medium (minus uracil and histidine) at 30°C. Cells were harvested at mid-log phase (as determined by optical density at 600 nm), and RNA was prepared as described by Frank et al. (8). For each sample, 16 µg of RNA was separated on an 8% polyacrylamide gel and then transferred to Hybond-N membranes (Amersham) by being electroblotted in 25 mM sodium-phosphate (pH 6.5). RNA was fixed to Hybond-N filters by UV cross-linking (UV Stratalinker 1800, Stratagene). Filters were hybridized at 25°C in a buffer of 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 5× Denhardt's solution, and 0.2% sodium dodecyl sulfate. 5'-end-labeled oligonucleotides complementary to U5 and U4 snRNAs were added at a concentration of 4 × 10⁵ cpm/ml. Filters were washed four times at 37°C in hybridization buffer (minus Denhardt's solution) and then autoradiographed. snRNA levels were quantitated by PhosphorImager analysis of the labeled filter (Molecular Dynamics).

RESULTS AND DISCUSSION

Comparative analysis of fungal U5 structure. Given the divergence between the *S. cerevisiae* U5 snRNA and all other U5 sequences (discussed above), comparative analysis of U5 structure necessitated obtaining U5 sequences from other, more closely related, yeast species. We have now determined the U5 sequences from five additional species of ascomycetous yeasts (see Materials and Methods for details). During the course of this study, the U5 sequence of *Schizosaccharomyces pombe*, another ascomycetous yeast, was also reported (37). These sequences are aligned in Fig. 1, along with several other representative sequences. This alignment provides the primary data for devising and testing structural models (see below).

Our current model of *S. cerevisiae* U5 snRNA secondary structure is presented in Fig. 2A, along with the closely related U5 structure of *S. kluyveri*. For comparison, an archetypal metazoan U5 secondary structure is shown in Fig. 2B. The models of fungal U5 secondary structure were based upon the identification of covarying positions within the U5 sequence alignment. Such instances of covariation provide strong evidence for the existence of duplexed RNA structure within the U5 molecule. What follows is a detailed description of fungal U5 secondary structure, including the phylogenetic evidence in support of the various substructures that define U5 (the nomenclature used to describe these elements is summarized in Fig. 2C):

(i) **Loop 1 (L1)-S1.** The fungal sequences spanning nt 84 to 110 (*S. cerevisiae* numbering will be used unless otherwise noted) can be unambiguously aligned with the existing U5 sequences. The central, highly conserved block of 11 nt is

FIG. 1. Alignment of U5 snRNA sequences. U5 sequences reported in this work, along with previously determined sequences, are shown in the alignment. Abbreviations are as follows: saccer, *S. cerevisiae* (29); sacklu, *S. kluyveri*; klulac, *K. lactis*; pactan, *P. tannophilus*; piccan, *P. canadensis*; sacexu, *S. exiguus*; schpom, *S. pombe* (37); aratha, *Arabidopsis thaliana* (40); pissat, *Pisum sativum* (12); crycoh, *Cryptocodinium cohnii* (31); tetpyr, *Tetrahymena pyriformis* (4); tetthe, *Tetrahymena thermophila* (26); caeele, *Caenorhabditis elegans* (39); dromel, *Drosophila melanogaster* (24); xenlav, *Xenopus laevis* (16); galgal, *Gallus gallus* (17); musmus, *Mus musculus* (15); ratnor, *Rattus norvegicus* (17); homsap, *Homo sapiens* (17). Boxed regions denote secondary structure elements as named in Fig. 2. Consensus sequences were derived from the entire alignment as well as from several taxonomic groups (yeasts, plants, protozoa [Prot.], and metazoa [Met.]). In the consensus (Cons.) sequence, uppercase denotes an invariant nucleotide, while lowercase denotes identity in all but one sequence.

flanked, in each case, by two complementary sequences; this region can thus form a stem-loop structure. The existence of this stem is supported by the occurrence of several covarying base pairs, examples of which are boxed in Fig. 2A. In terms of primary sequence conservation, the fungal sequences that make up S1 are conserved to a much greater extent than are the sequences found in the other stems of U5.

(ii) **IL1.** Both the 5' and 3' halves of IL1 can be readily aligned with existing sequences. The level of conservation is highest in the 3' half of IL1, where the sequence is invariably CCG in the fungi. In contrast, the fungal and nonfungal sequences that define the 5' halves of IL1 are only moderately conserved in comparison to L1 and the 3' halves of IL1. Nevertheless, the 5' halves of IL1 in all species of U5 include an invariant CG dinucleotide at positions 79 and 80.

(iii) **VSL.** The fungal nucleotides immediately upstream of the conserved SL1-IL1 region cannot be aligned with the sequences of nonfungal species of U5. Despite the lack of primary sequence conservation in this region, stem-loop structures can be drawn in the case of four of the seven fungal U5 snRNAs, including *S. cerevisiae* and *S. kluyveri* (Fig. 2A). The existence of this novel stem-loop structure is supported by the existence of covariations between *S. cerevisiae*, *S. exiguus*, and *S. kluyveri* sequences. Furthermore, the results of chemical and enzymatic structure probing studies, of both *S. cerevisiae* and *K. lactis* U5 snRNAs, are consistent with the existence of such a stem-loop structure (8a). To date, this novel structure has been observed in only a subset of fungal sequences; no examples are known in nonfungal sequences. Consequently, we have named this domain the variable stem-loop (VSL) in order to reflect the fact that it is not an obligate feature of all U5 snRNAs.

(iv) **S2.** The regions defining the 5' and 3' halves of S2 (nt 28 to 40 and 114 to 125) do not bear strong sequence similarity to their nonfungal counterparts. However, the existence of this stem is strongly supported on phylogenetic grounds by the existence of many covarying base pairs (Fig. 2A). The length and placement of this putative stem relative to IL1 further suggest that this structure is homologous to S2 of metazoan U5.

(v) **IL2.** In general, the lengths of both halves of IL2 are quite variable in the fungi, ranging from 0 to 15 nt within the 5' half and 6 to 10 nt within the 3' half of this structure. Furthermore, these regions are highly divergent at the primary sequence level and cannot be reliably aligned.

(vi) **S3.** Despite the great primary sequence diversity in this region, potential S3 structures can be identified in all U5 species (Fig. 2A). In addition to the sequence heterogeneity, the lengths of these stems are quite variable relative to the lengths of S1 and S2.

(vii) **Sm linker region.** The sequences extending from the base of S3 to the Sm binding site vary greatly in length in the fungi and could not be aligned with any precision. Because of the lack of sequence conservation and absence of obvious secondary structures, this region may serve primarily as a variable linker region joining the U5-specific domain to the Sm binding site.

(viii) **Sm site.** Putative Sm protein binding sites could be identified in all U5 sequences, thus highlighting the importance of this sequence motif in snRNP assembly (14, 21). The consensus sequence of the fungal U5 Sm sites is R(U)₅NGgaa (see the legend to Fig. 1).

(ix) **3' stem-loop.** Aside from fungal U5 snRNAs, the Sm site is invariably followed by a stem-loop structure of variable size, termed the 3' stem-loop. In contrast, this structure is found in fewer than half of the yeast species that we have examined

FIG. 3. U5 deletion analysis. The extent and placement of deletions introduced into the *S. cerevisiae* U5 snRNA are depicted. Details of mutant construction are given in Materials and Methods.

(three of seven examples possess this structure). Furthermore, Northern (RNA) blot analysis revealed that the U5 snRNAs of *S. cerevisiae* and *S. exiguus* are expressed in two forms, termed U5 short (U5-S) and U5 long (U5-L) (29, 33). In the case of *S. cerevisiae*, both U5 species are products of the *SNR7* gene that differ in the presence or absence of the 3' stem-loop (29).

(x) **Summary.** Overall, the structural organization of the fungal U5 snRNAs is similar to that found in other taxa. The fungal U5 structures differ from these other U5 structures primarily in the existence of extra fungal nucleotides. The fungi have exploited two basic strategies to accommodate these nucleotides: (i) the extension, or enlargement, of universally conserved structures (e.g., IL2, S3, and the linker region defined above) and (ii) the creation of domains that are phylogenetically sporadic, such as the VSL structures of *S. cerevisiae* and *S. kluyveri*.

U5 deletion analysis. To test the functional necessity of the various structural domains of the U5 snRNA, we systematically deleted portions of the *S. cerevisiae* molecule and tested the ability of the resulting deletion mutants to replace the biological activity of wild-type U5. Since the U5 gene is essential for growth, the phenotypes of U5 mutations can be readily assessed by a genetic complementation assay. To this end, deletion mutants, carried on centromeric *URA3* plasmids, were transformed into a yeast strain in which wild-type U5 activity was provided by a gene fusing the repressible *GAL1* promoter to the *SNR7* gene (29). Shifting this strain to glucose medium causes the repression of wild-type U5 synthesis; cellular growth thus becomes dependent on the activity of the U5 allele provided.

The set of deletion mutants that we have characterized is shown in Fig. 3, while the growth phenotypes of the strains carrying the U5 deletion mutants are summarized in Table 1. For ease of analysis, all of the deletions were constructed in an *SNR7* gene that expresses only the short form of U5 (U5-S), which is truncated immediately upstream of the 3' stem-loop (Δ3'SL). This gene produces wild-type growth under all conditions tested, indicating that the 3'SL is dispensable for U5

TABLE 1. Growth phenotypes of U5 deletion alleles

Construct	Doubling time ^a	Cell growth ^b
Wild type	1.0	+++
Δ VSL	1.1	+++
Δ IL1.5'	—	—
Δ IL1.3'	—	—
Δ L1	—	—
Δ 1	1.0	+++
Δ 2	1.2	+++
Δ (1+2)	1.2	+++
Δ 3	1.0	+++
Δ 4	—	+

^a Doubling times (in arbitrary units) were determined from cultures growing in liquid YEPD medium at 30°C and normalized to wild-type levels. —, inviability.

^b Cell growth on plates was assayed qualitatively by streaking cultures for single colonies on solid YEPD medium at 30°C. +++, wild-type growth; +, colony growth after prolonged incubation; —, inviability.

function (28a). The functional dispensability of the 3'SL was anticipated by two previous observations: (i) similarly truncated U5 snRNAs exist naturally in several fungal species, including *S. cerevisiae* (29, 33; this work), and (ii) the U5-S molecule is incorporated into spliceosomes (30). Similarly, we found that deletion of the VSL domain (Δ VSL) did not affect cell growth, meaning that this structure is also dispensable for U5 function. Since one goal of this analysis was to create the smallest functional U5, we constructed the remainder of the deletions in a U5 context in which both the VSL and the 3'SL had been excised (this construct also supported wild-type growth). We hoped that two additional benefits would be derived from this approach: (i) combining deletions might exacerbate the phenotypes of individual mutants and thus allow us to uncover subtle phenotypes, and (ii) smaller U5 molecules might ameliorate RNA folding problems caused by the deletions.

Surprisingly, the only changes in U5 snRNA structure that caused lethal phenotypes were Δ IL1.5', Δ IL1.3', and Δ L1 (Fig. 3, Table 1; note that L1 was replaced by a GAGA tetra-loop in order not to radically disrupt S1 folding). These three regions encompass the majority of the highly conserved nucleotides of U5, aside from the Sm protein binding site. The functional necessity of these sequence elements is therefore reflected in their evolutionary conservation.

In contrast to the lethal deletions, deletions Δ 1 and Δ 2, which excise all of IL2 and the majority of S3, produced no noticeable changes in growth phenotypes [a combination of these two deletions, Δ (1+2), had a phenotype that was similar to those of the single deletions]. Thus, neither the sequences nor the structures predicted in these regions (i.e., S3 and IL2) are crucial for U5 activity. This result was surprising, because the IL2 structure, though not sequence, of mammalian U5 influences the affinity of Sm core proteins for this snRNA (13). While we have not directly tested whether the Δ 1, Δ 2, or Δ (1+2) U5 snRNAs bind Sm core proteins, the wild-type growth phenotypes and steady-state U5 snRNA levels (see below) produced by these constructs indicate that Sm binding is not drastically affected by the loss of yeast internal loop IL2. The contrasting results of our study and that of Jarmolowski and Mattaj (13) suggest that the mode of Sm protein-U5 snRNA interaction differs between yeast and metazoan species.

In *S. cerevisiae* U5, the Sm site is spaced approximately 12 nt further downstream of S3 than in nonyeast U5 molecules. To test whether these extra nucleotides constitute an important functional constraint on U5 structure, deletion mutant Δ 3 was created by excising 12 nt immediately downstream of S3 (nt 146

to 157). In effect, this construct mimics the S3-Sm spacing of nonyeast U5 snRNAs. Deletion of this region, however, did not affect U5 activity in the complementation assay (Table 1). Thus, the spacing between the Sm binding site and the rest of the U5 structure must not be tightly constrained in *S. cerevisiae*.

A more subtle constraint on U5 structure was revealed by deletion Δ 4, which removes the entire 5' half of S2, and all but five nucleotides of IL2. This U5 snRNA variant conferred a strong deleterious effect on cellular growth. While Δ 4 was viable at 30°C on solid media (i.e., YEPD plates), its doubling time was too long to reproducibly measure in liquid culture. Furthermore, Δ 4 was unable to support growth at 37°C on any medium. Since IL2 is not required for U5 activity, this result suggests that the integrity of S2 is important, though not essential, for U5 function.

To determine whether the sequence, or simply the structure, of S2 is important for proper U5 snRNA function, we created a series of constructs in which the essential domain of U5 (IL1 and SL1) was fused to an artificial stem. We reasoned that such constructs would provide U5 activity only if S2 has no sequence-specific role in U5 function. Five constructs, named MIN1 to MIN5 (for minimal functional U5), were made, each differing in the length of the stem-closing IL1 (Fig. 4). We refer to this engineered stem as S2/3, since it must replace the functionalities of wild-type S2 and S3. Plasmids bearing the minimal U5 constructs were transformed into the pGAL1-SNR7, strain and the functions of the MIN U5 snRNAs were tested in the complementation assay.

The results of the growth assay revealed a wide spectrum of growth phenotypes. In general, the defects conferred by these MIN genes became more severe as the length of S2/3 was decreased. This effect can be seen most clearly by comparing the doubling times of cells bearing the MIN U5 series (Fig. 4B). For instance, while the doubling time of MIN5 was 1.8-fold that of the wild type at 30°C, the MIN1 construct was unable to support growth. Furthermore, as the length of S2/3 was decreased, the snRNAs became increasingly more sensitive to higher temperatures; while MIN5- and MIN4-expressing cells were viable at 37°C, MIN3 and MIN2 cells were inviable above 33 and 25°C, respectively.

The primary conclusion derived from our analysis of the MIN series is that S2 can be replaced by an artificial stem, with only modest effects on U5 activity (in the case of MIN5 and MIN4). Since disruption of S2 (by deletion mutant Δ 4) produces a strong effect on growth, we conclude that wild-type U5 activity requires that IL1 be terminated by a stem but that the sequence of this stem is relatively unimportant. In addition, analysis of the MIN U5 series indicates that the length of this stem is also an important determinant of U5 function. We suggest three explanations for how the length of S2/3 might influence MIN U5 function: (i) the nucleotides at the base of S2/3 may not by themselves form stable base pairs in vivo, while further extensions of this stem provide added stability; (ii) S2/3 may function as a sequence-nonspecific protein binding site (for Sm proteins for instance) with a minimum length requirement; and (iii) the stem may provide a spacing function, separating regions of the snRNA, such as L1 and the Sm site, from one another. The flexibility of the linker region, as indicated by the lack of a growth defect conferred by Δ 3, suggests that model iii is unlikely; distinguishing models i and ii will require additional analyses.

Analysis of a yeast-metazoan U5 chimera. The deletion mutants Δ (1+2), MIN5, and MIN4 produced U5 snRNA molecules with growth characteristics similar to those of wild-type U5. At least superficially, the secondary structures of these snRNAs are similar to that of mammalian U5 snRNA.

FIG. 4. Summary of U5 MIN series. (A) Structures of the MIN series of U5 constructs (see Materials and Methods for construction details); (B) results of growth assays comparing the phenotypes conferred by this series of constructs at a variety of temperatures. Plate assays were carried out on YEPD plates, while doubling times of strains bearing these constructs were measured at 30°C in liquid YEPD medium. W.T., wild type.

We were thus prompted to express a mammalian U5 snRNA in yeast cells and determine whether it too could support viability. We chose to test the function of a chimeric U5 snRNA (termed YH.A), in which the conserved structure of human U5 was joined to the base of *S. cerevisiae* S3; this construct thus encodes the *S. cerevisiae* Sm site linked to human U5-specific sequences (Fig. 5A). The chimeric U5 gene was then inserted into a cassette encoding the yeast U5 transcriptional initiation and termination sequences. The effects on secondary structure of such a fusion were predicted to be minor, since the fusion junctions were constructed in regions that were entirely dispensable for growth. The function of this construct was then tested in our growth assay.

Surprisingly, despite the similarity in sequence and structure with some of the functional yeast U5 deletion mutants, the chimeric U5 was not functional under any of the conditions tested (Fig. 5B). YH.A provides all of the invariant nucleotides found in the U5 snRNA but lacks a moderately conserved sequence element: the 5' half of IL1 (IL1.5'), which is essential for growth in yeast cells. To test the hypothesis that the human IL1 sequence is responsible for the inviability conferred by YH.A, we replaced the human IL1.5' sequence of the chimeric U5 with its yeast counterpart (to create YH.B). Interestingly, this modified chimera was able to partially complement the U5 gene disruption, with a growth rate 4.6 times that of the wild type at 30°C (growth defects were more severe at 18, 33, and

37°C). The lethal growth phenotype conferred by YH.A could thus be partially suppressed simply by replacing its IL1 region with *S. cerevisiae* IL1. Experiments described below suggest that it is unlikely that suppression occurs by an increase in the expression or stability of the YH.B RNA. Thus, IL1 constitutes a sequence-specific element that is required for wild-type U5 activity. It is tempting to speculate that this loop is the binding site for a U5-specific protein, such as PRP8 (see discussion below) (19).

To complement the analysis of YH.B, we replaced the IL1 region of *S. cerevisiae* U5 (construct ΔVSL) with the mammalian IL1 region, to form construct YH.C. This construct was designed to assess the effect of altering the IL1 sequence in an otherwise wild-type U5 context. Surprisingly, YH.C conferred only a slight decrease in cellular growth (doubling time, 1.1 times that of the wild type at 30°C). Thus, in an otherwise wild-type background (i.e., *S. cerevisiae* sequence), the invariant nucleotides of IL1 are sufficient to provide function. A likely explanation for these observations is that IL1 constitutes only a portion of a larger functional unit (e.g., a U5 snRNP protein binding site). In this case, the activity that is lost by substituting the nonconserved nucleotides in IL1 would be negligible if the remainder of the U5 structures were wild type. However, substitutions within IL1 would be detrimental if other U5 structures or sequences were perturbed, as is the case in chimeras YH.A and YH.B.

FIG. 5. Analysis of yeast-human U5 chimeric snRNAs. (A) Structures of the chimeric U5 snRNAs. Boldface, yeast nt; italics, human nt. (B) Growth phenotypes conferred by these constructs at a variety of temperatures. Plate assays were carried out on YEPD plates, while doubling times of strains bearing these constructs were measured at 30°C in liquid YEPD medium. W.T., wild type.

Analysis of deletion mutant RNA. The preceding interpretation of U5 snRNA structure and function data is based on the premise that differences in the growth phenotypes conferred by the various deletion mutants directly reflect functional differences in U5 snRNA structure. However, the U5 snRNA deletions might disrupt U5 snRNP biosynthesis or stability rather than function per se. If so, the steady-state levels of the U5 variants should be correlated with their activities. A precedent for this hypothesis is the observation that reductions in the steady-state levels of U5 Sm site mutant snRNAs are proportional to the severities of their growth phenotypes (14). Thus, snRNAs not capable of binding Sm core proteins are rapidly degraded in vivo.

To test the hypothesis that the deletion mutants disrupt U5 biosynthesis or stability, we examined cellular U5 snRNA levels, via Northern blots, in strains heterozygous for wild-type and deletion mutant U5 genes. These strains differed from those used in the growth assays in only one respect: a centromeric plasmid carrying the wild-type U5 gene (pBWHis600) was used in place of pGAL1-SNR7. Steady-state snRNA levels of strains grown on selective medium (SD without Ura and His) at 30°C were quantitated by PhosphorImager analysis of Northern blots. As seen in Fig. 6 and Table 2, we observed that even phenotypically lethal deletions in U5 caused little or no

FIG. 6. Northern blot analysis of selected U5 constructs. RNA was extracted from strains heterozygous for the deletion U5 constructs and pBWHis600 (the source of wild-type U5) and probed with U5 and U4 oligonucleotides. Strains were grown at 30°C in liquid SD medium without Ura and His. Both U5-S and U5-L forms are shown (see text for details).

TABLE 2. Steady-state levels of mutant U5 snRNAs

Construct	Relative RNA level ^a (%)	Growth ^b at 30°C
Wild type	ND	+++
ΔVSL	100	+++
Δ(1+2)	81	+++
ΔIL1.5'	91	—
ΔL1	96	—
ΔIL1.3'	111	—
Δ4	59	+
MIN5	37	++
MIN4	36	++
MIN3	31	+
MIN2	36	+
MIN1	24	—
YH.A	43	—
YH.B	38	+
YH.C	139	+++

^a Northern blots shown in Fig. 7 were quantitated by phosphorimaging, and the U5 snRNA levels were normalized to U4 levels. The quantity of ΔVSL RNA was arbitrarily set to 100%. ND, not determined (the quantity of wild-type U5 snRNA produced by construct pSE-7EB could not be accurately determined because it comigrates with endogenous U5).

^b Cell growth was assayed qualitatively by streaking cultures for single colonies on solid YEPD medium at 30°C. +++, wild-type growth; ++ and +, growth slower than wild type; —, inviability.

substantial decreases in snRNA levels. Thus, such recessive lethal constructs as ΔL1, ΔIL1.5', and ΔIL1.3' produced levels of variant U5 snRNAs that were comparable to those of wild-type constructs [e.g., ΔVSL and Δ(1+2)]. These results strongly support the premise that the phenotypes conferred by the deletion mutants arise primarily from disruptions in U5 snRNA function rather than biosynthesis or stability. Taken together with the results of Jones and Guthrie (14), our examination of deletion mutant snRNA levels indicates that the Sm protein binding site is the primary determinant of U5 snRNP stability. In contrast, deletion of IL1 or L1 structures, which presumably disrupts the binding of U5 snRNP-specific proteins, does not greatly affect snRNP stability. In keeping with this observation, Jarmolowski and Mattaj (13) have demonstrated biochemically that deletion of the homologous structures from *Xenopus* U5 (i.e., IL1 and L1) does not affect Sm core protein binding to this snRNA.

In the case of the MIN U5 series, progressively larger deletions of S2/3, which produced increasingly more-severe growth phenotypes, did not result in concomitant decreases in MIN U5 levels. For example, the levels of RNA produced by the MIN2 construct, which is unable to support growth at 30°C, were equivalent to those produced by the viable MIN5 and MIN4 constructs (36 versus 37 or 36% of wild-type levels, respectively). Thus, all of these truncated U5 snRNAs are apparently equally able to assemble core snRNP particles. The differences in the U5 activities of these alleles must therefore be due to defects in higher-order U5 snRNA functions, such as binding U5 snRNP proteins or interacting with other factors.

In the case of the chimeric U5 snRNAs, the nonfunctional YH.A construct produced a level of U5 equivalent to that produced by the viable YH.B construct (43% YH.A versus 38% YH.B). Thus, the level of suppression conferred by the yeast IL1.5' sequence was not caused simply by an increase in the steady-state level of YH.B; IL1.5' must be a determinant of U5 snRNP activity.

Conclusions and implications. Using phylogenetic analysis, we have proposed a generic model of U5 snRNA secondary structure that is consistent with the previous analyses of metazoan U5 secondary structure (1, 3, 11). We find that

several fungal U5 snRNAs encode nonconserved regions of variable structure. As is the case with the variable regions of RNAs such as RNase P RNA, rRNA, U2 snRNA, and U1 snRNA (9, 18, 35, 36, 42), the U5 variable regions are not essential for cell growth under standard conditions. However, the conservation of the VSL (present in four of seven fungal species examined) suggests that this domain has a subtle function in splicing that we cannot presently identify (18).

Through deletion analysis, we have begun to define the functionality of the U5 snRNA. We find that the minimal U5 structure consists of the conserved SL1 structure flanked by a moderately conserved IL1 and linked to the Sm protein binding site. IL1 must be closed by a stem of minimally 6 bp for U5 activity, but the sequence of this stem is relatively unimportant. Similarly, the sequence of S1 is not critical for U5 function, since the highly divergent S1 of human U5 functioned, at least partially, in *S. cerevisiae*. Finally, an Sm protein binding site must be present in order to bind the snRNP core proteins. The results presented here and by Jones and Guthrie (14) suggest that Sm protein binding is the primary determinant of the overall stability of the U5 snRNP.

Our analysis of the chimeric U5 constructs is consistent with the possibility that the IL1 structure functions as a binding site for a U5-specific protein, such as PRP8 (19). Bach and Lührmann (1) proposed a similar model of U5 snRNP organization based on the observation that U5-specific proteins protect the IL1 region of mammalian U5 from chemical modification (suggesting steric occlusion of IL1 by this set of proteins, if not direct binding). Our work extends these results by demonstrating that IL1 provides an essential function for the U5 snRNP. The isolation of *trans*-acting genetic suppressors of IL1 mutations could provide strong evidence that this RNA structure does indeed serve as a protein binding site.

In contrast to IL1, the bases of L1 are probably not bound by U5 snRNP-specific proteins and instead, interact with other splicing factors. Biochemical evidence indicates that the bases of L1 are accessible to chemical modification, even in the U5 snRNP (1, 3). Furthermore, genetic evidence suggests that L1 nucleotides can interact directly with the pre-mRNA (25). Finally, the identification of several splicing factors that genetically interact with L1 further underscores the functional importance of this structure; such proteins may contact L1 along the phosphate backbone or transiently interact with L1 nucleotides during splicing (8).

In conclusion, the U5 snRNA is defined by two conserved sequence elements (L1 and IL1) in addition to the Sm binding site. The nucleotides that make up these loop regions must, therefore, provide all U5 snRNA sequence-specific activities (i.e., binding domains for U5-specific proteins and interactions with other RNAs). Given the myriad of proteins that make up the U5 snRNP (2) and the relative paucity of potential protein binding sites on the U5 snRNA, it is likely that assembly of the U5 snRNP entails a complex hierarchy of protein-RNA and protein-protein interactions. Elucidation of the components that make up this snRNP will be necessary to understand the function of the U5 snRNP.

ACKNOWLEDGMENTS

We thank Bruce Patterson, Shelly Haltiner-Jones, and Beate Schwer for valuable critiques of all phases of this work and Elizabeth Blackburn, Jon Staley, and Peter Walter for comments on the manuscript. We thank D. Black and A. Pinto for providing a clone of human U5. Saira Mian and Bryn Weiser provided valuable support in the analysis of U5 sequences. We also thank Lucita Esperas for superb technical assistance.

This work was supported by National Institute of Health (NIH) grant GM21119. D.N.F. was supported by a predoctoral fellowship

from the National Science Foundation and grants from NIH and the Markey Charitable Trust. C.G. is an American Cancer Society Research Professor of Molecular Genetics.

REFERENCES

- Bach, M., and R. Lührmann. 1991. Protein-RNA interactions in 20S U5 snRNPs. *Biochim. Biophys. Acta* **1088**:139–143.
- Bach, M., G. Winkelmann, and R. Lührmann. 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl. Acad. Sci. USA* **86**:6038–6042.
- Black, D. L., and A. L. Pinto. 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol. Cell. Biol.* **9**:3350–3359.
- Branlant, C., A. Krol, E. Lazar, B. Haendler, M. Jacob, L. Galego-Dias, and C. Pousada. 1983. High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. *Nucleic Acids Res.* **11**:8359–8367.
- Bruns, T. D., T. J. White, and J. W. Taylor. 1991. Fungal molecular systematics. *Annu. Rev. Ecol. Syst.* **22**:525–564.
- Chabot, B., D. L. Black, D. M. LeMaster, and J. A. Steitz. 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science* **230**:1344–1349.
- Frank, D., and C. Guthrie. 1992. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.* **6**:2112–2124.
- Frank, D., B. Patterson, and C. Guthrie. 1992. Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. *Mol. Cell. Biol.* **12**:5197–5205.
- Frank, D. N. Unpublished data.
- Gutell, R. R., B. Weiser, C. R. Woese, and H. F. Noller. 1985. Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **32**:155–216.
- Guthrie, C., and G. R. Fink (ed.). 1991. Guide to yeast genetics and molecular biology. Academic Press, Inc., San Diego, Calif.
- Guthrie, C., and B. Patterson. 1988. Spliceosomal snRNAs. *Annu. Rev. Genet.* **23**:387–419.
- Hanley, B. A., and M. A. Schuler. 1991. cDNA cloning of U1, U2, U4 and U5 snRNA families expressed in pea nuclei. *Nucleic Acids Res.* **19**:1861–1869.
- Jarmolowski, A., and I. W. Mattaj. 1993. The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical. *EMBO J.* **12**:223–232.
- Jones, M. H., and C. Guthrie. 1990. Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site. *EMBO J.* **9**:2555–2561.
- Kato, N., and F. Harada. 1981. Nucleotide sequence of nuclear 5S RNA of mouse cells. *Biochem. Biophys. Res. Commun.* **99**:1468–1476.
- Kazmaier, M., G. Tebb, and I. W. Mattaj. 1987. Functional characterization of *X. laevis* U5 snRNA genes. *EMBO J.* **6**:3071–3078.
- Krol, A., H. Gallinaro, E. Lazar, M. Jacob, and C. Branlant. 1981. The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. *Nucleic Acids Res.* **9**:769–787.
- Liao, X. L., L. Kretzner, B. Seraphin, and M. Rosbash. 1990. Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. *Genes Dev.* **4**:1766–1774.
- Lossky, M., G. J. Anderson, S. P. Jackson, and J. Beggs. 1987. Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. *Cell* **51**:1019–1026.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mattaj, I. W. 1988. UsnRNP assembly and transport, p. 100–114. In M. L. Birnstiel (ed.), Structure and function of major and minor small nuclear ribonucleoprotein particles. Springer-Verlag, Berlin.
- McClary, J. A., F. Witney, and J. Geisselsoder. 1989. Efficient site-directed *in vitro* mutagenesis using phagemid vectors. *Biotechniques* **3**:282–289.
- Milligan, J. F., and O. C. Uhlenbeck. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**:51–62.
- Myslinski, E., C. Branlant, E. D. Wieben, and T. Pederson. 1984. The small nuclear RNAs of *Drosophila*. *J. Mol. Biol.* **180**:927–945.
- Newman, A. J., and C. Norman. 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**:743–754.
- Orum, H., H. Nielsen, and J. Engberg. 1991. Spliceosomal small nuclear RNAs of *Tetrahymena thermophila* and some possible snRNA-snRNA base-pairing interactions. *J. Mol. Biol.* **222**:219–232.
- Pace, N. R., D. K. Smith, G. J. Olsen, and B. D. James. 1989. Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA—a review. *Gene* **82**:65–75.
- Parker, R., and B. Patterson. 1987. Architecture of fungal introns: implications for spliceosome assembly, p. 133–149. In M. Inouye and B. S. Dudoek (ed.), Molecular biology of RNA: new perspectives. Academic Press, Inc., San Diego, Calif.
- Patterson, B., D. N. Frank, and C. Guthrie. Unpublished data.
- Patterson, B., and C. Guthrie. 1987. An essential yeast snRNA with a U5-like domain is required for splicing *in vivo*. *Cell* **49**:613–624.
- Pikielny, C. W., B. C. Rymond, and M. Rosbash. 1986. Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature (London)* **324**:341–345.
- Reddy, R., D. Spector, D. Henning, M. H. Liu, and H. Busch. 1983. Isolation and partial characterization of dinoflagellate U1-U6 small RNAs homologous to rat U small nuclear RNAs. *J. Biol. Chem.* **258**:13965–13969.
- Riedel, N., S. Wolin, and C. Guthrie. 1987. A subset of yeast snRNA's contains functional binding sites for the highly conserved Sm antigen. *Science* **235**:328–331.
- Roiha, H., E. O. Shuster, D. A. Brow, and C. Guthrie. 1989. Small nuclear RNAs from budding yeasts: phylogenetic comparisons reveal extensive size variation. *Gene* **82**:137–144.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
- Shuster, E. O., and C. Guthrie. 1988. Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. *Cell* **55**:41–48.
- Siliciano, P. G., W. J. Kivens, and C. Guthrie. 1991. More than half of yeast U1 snRNA is dispensable for growth. *Nucleic Acids Res.* **19**:6367–6372.
- Small, K., P. Brennwald, H. Skinner, K. Schaefer, and J. A. Wise. 1989. Sequence and structure of U5 snRNA from *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **17**:9483.
- Tazi, J., C. Alibert, J. Temsamani, I. Reveillaud, G. Cathala, C. Brunel, and P. Jeanteur. 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.
- Thomas, J. D., K. Lea, E. M. Zucker-Aprison, and T. Blumenthal. 1990. The spliceosomal snRNAs of *Caenorhabditis elegans*. *Nucleic Acids Res.* **18**:2633–2642.
- Vankan, P., D. Edoh, and W. Filipowicz. 1988. Structure and expression of the U5 snRNA gene of *Arabidopsis thaliana*. Conserved upstream sequence elements in plant U-RNA genes. *Nucleic Acids Res.* **16**:10425.
- Wassarman, D. A., and J. A. Steitz. 1992. Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. *Science* **257**:1918–1925.
- Waugh, D. S., C. J. Green, and N. R. Pace. 1989. The design and catalytic properties of a simplified ribonuclease P RNA. *Science* **244**:1569–1571.
- Wise, J. A. 1991. Preparation and analysis of low molecular weight RNAs and small ribonucleoproteins, p. 405–415. In C. Guthrie and G. R. Fink (ed.), Guide to yeast genetics and molecular biology. Academic Press, Inc., San Diego, Calif.
- Wyatt, J. R., E. J. Sontheimer, and J. A. Steitz. 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**:2542–2553.