RNA (1997), *3*:17–26. Cambridge University Press. Printed in the USA.

Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing

LING XIA, NICHOLAS J. WATKINS, and E. STUART MAXWELL
Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695-7622, USA

ABSTRACT

Vertebrate U14 snoRNAs are encoded within hsc70 pre-mRNA introns and U14 biosynthesis occurs via an intron-processing pathway. We have shown previously that essential processing signals are located in the termini of the mature U14 molecule and replacement of included boxes C or D with oligo C disrupts snoRNA synthesis. The experiments detailed here now define the specific nucleotide sequences and structures of the U14 termini that are essential for intronic snoRNA processing. Mutagenesis studies demonstrated that a 5, 3-terminal stem of at least three contiguous base pairs is required. A specific helix sequence is not necessary and this stem may be extended to as many as 15 base pairs without affecting U14 processing. The spatial positioning of boxes C and D with respect to the terminal stem is also important. Detailed analysis of boxes C and D revealed that both consensus sequences possess essential nucleotides. Some, but not all, of these critical nucleotides correspond to those required for the stable accumulation of nonintronic yeast U14 snoRNA. The presence of box C and D consensus sequences flanking a terminal stem in many snoRNA species indicates the importance of this "terminal core motif" for snoRNA processing.

Keywords: boxes C and D; intron-encoded snoRNA; ribosomal RNA processing; small nucleolar RNAs; terminal core motif

INTRODUCTION

The small nucleolar RNAs (snoRNAs) play important roles in eukaryotic ribosomal RNA processing and ribosome biogenesis. Specific snoRNA species in both vertebrates (Kass et al, 1990; Savino & Gerbi, 1990; Peculis & Steitz, 1993; Tycowski et al., 1994) and yeast (Tollervey, 1987; Li et al., 1990; Hughes & Ares, 1991; Morrissey & Tollervey, 1993; Schmitt & Clayton, 1993; Lygerou et al., 1996) have been shown to be critical for various rRNA processing events. It has also been suggested that snoRNAs may serve as "RNA chaperones" to assist in folding the rRNA precursor for processing and ribosome assembly (Bachellerie et al., 1995; Maxwell & Fournier, 1995; Steitz & Tycowski, 1995). More recently, it has been demonstrated that many snoRNAs utilize their rRNA-complementary sequences to direct the base-specific 2′-*O*-methylation of ribosomal RNA (Kiss-Laszlo et al., 1996; Nicoloso et al., 1996; Tollervey, 1996; J. Ni & M.J. Fournier, pers. comm.).

The snoRNAs can be classified into two major groups based upon the presence of nucleotide consensus sequences. Most snoRNAs characterized to date possess nucleotide boxes C and D, first defined in the analysis of the U3 snoRNA (Wise & Weiner, 1980; Maxwell & Fournier, 1995). With the exception of MRP, all of the remaining species belong to the second group and possess an ACA sequence motif positioned near the 3′ terminus (Balakin et al., 1996). Both the box C and D and the ACA consensus sequences have been shown to be important for snoRNA processing events and the accumulation of mature snoRNA species (Huang et al., 1992; Peculis & Steitz, 1994; Balakin et al., 1996; Caffarelli et al., 1996; Watkins et al., 1996). In addition, boxes C and D have been implicated in nuclear retention and 5′ cap hypermethylation (Baserga et al., 1992; Peculis & Steitz, 1994; Terns et al., 1995).

Examination of snoRNA gene structure has revealed that most vertebrate snoRNAs are encoded within introns of protein-coding genes (reviewed in Sollner-Webb, 1993; Bachellerie et al., 1995; Maxwell & Fournier, 1995; Steitz & Tycowski, 1995). SnoRNA biosynthesis involves the transcription of the snoRNA species as part of the pre-mRNA followed by intron-processing to produce a mature snoRNA molecule (Leverette et al., 1992; Fragapane et al., 1993; Kiss & Filipowicz, 1993, 1995; Prislei et al., 1993; Tycowski et al., 1993).

Reprint requests to: E. Stuart Maxwell, Department of Biochemistry, Box 7622, North Carolina State University, Raleigh, North Carolina 27695-7622, USA; e-mail: maxwell@bchserver.bch.ncsu.edu.

We demonstrated recently that the 5′- and 3′-terminal regions of the mouse U14 that contain nucleotide boxes C and D are required for processing (Watkins et al., 1996). In addition, replacement of either box C or D with oligo C disrupts U14 synthesis (Watkins et al., 1996). In the present study, we have defined with detailed mutagenesis those specific nucleotides as well as structures of the U14 molecule that are essential for processing this intronic snoRNA. We define a "terminal core motif" consisting of a base paired, 5′,3′-terminal stem with flanking box C and D consensus sequences. Critical nucleotides found within boxes C and D are similar, but not identical, to those required for the accumulation of nonintronic yeast U14 snoRNA. The phylogenetic conservation of the terminal core motif in both intronic and nonintronic U14 snoRNAs, as well as many other intronic snoRNA species, indicates the importance of this RNA element for snoRNA processing.

RESULTS

The sequences/structures required for U14 processing are located in the 5′ and 3′ termini of the mature snoRNA molecule and include nucleotide boxes C and D (Watkins et al., 1996). Specific nucleotide sequences and structural requirements for U14 processing were assessed by cloning mutant U14 snoRNAs into a *Sma* I cloning site engineered within intron 5 of the mouse hsc70 pre-mRNA (Fig. 1A) (Watkins et al., 1996). As seen in Figure 1C, mature U14 snoRNA was processed efficiently from both the native and engineered hsc70 precursor RNAs when injected into *Xenopus* oocyte nuclei. The 5′ terminus of U14 processed from the engineered hsc70 pre-mRNA was identical to the 5′ terminus of U14 processed from the wild-type precursor for all constructs (data not shown). This implies that the slight increase in heterogeneity seen with the U14 processed from the engineered hsc70 pre-mRNA is at the 3′ terminus of the processed snoRNA. In this and previous work, we observe no effect upon splicing when U14 is mutated or eliminated from the intron. Although equal amounts of each precursor were injected, variations in amount of spliced exons and processing intermediates were often observed among the different U14 mutants, as noted previously (Watkins et al., 1996). The lack of a 5′-capped nucleotide on the precursors, the varying stabilities of different precursors, and variations in processing efficiencies of different oocyte preparations account for these differences.

A base paired terminal helix is required for U14 processing

The effect of altering terminal helix size, sequence, and integrity upon intronic U14 snoRNA processing was examined first (Fig. 1B,C). Replacement of the U14 5′ terminus with the U14 3′-terminal sequence disrupted base pairing of the terminal stem and resulted in the loss of U14 processing (StemNO). U14 processing was restored when compensatory mutations were made in the 3′ end by substituting the U14 5′ terminal sequence (StemINV). U14 processing also occurred when the wild-type helix sequence was replaced with alternative base pairs (StemALT). Although a specific helix sequence is clearly not required, alterations in terminal helix sequence can affect U14 processing efficiency. Disruption of individual base pairs within the mature four-base pair helix revealed that hydrogen bonding of those three base pairs closest to boxes C and D (pairs 2, 3, and 4) were required for U14 synthesis.

Frequently, base pairing of the terminal helix in the precursor transcript extends beyond the 5′ and 3′ ends of the mature U14 snoRNA (Liu & Maxwell, 1990). Therefore, we assessed the processing of U14 snoRNAs with terminal stems of varying length (Fig. 2A). Each U14 precursor with extended stem structure was processed correctly, independent of base pair composition or the length of the terminal stem extension (Fig. 2B). Increasing terminal helix length did, however, increase the amount of the 5′ processed intermediate (approximately 250 nt), probably as a result of the extended stem stabilizing this transient intermediate. Shortening of the wild-type stem demonstrated that a minimum helix length of three base pairs was required. This result confirms the importance of base pairs 2, 3, and 4 in the mature terminal stem (see above). Primer extension analysis confirmed that the 5′ terminus for all processed U14 snoRNAs matched that for the wild-type molecule (data not shown).

Nucleotide boxes C and D contain specific, phylogenetically conserved nucleotides critical for U14 processing

Preliminary results from our laboratory have shown that both box C and D consensus sequences are required for U14 processing (Watkins et al., 1996). To guide a more detailed mutagenesis study of these consensus sequences, a phylogenetic analysis of box C (box C′ in U3) and D sequences from 49 different snoRNA species was first conducted (Table 1). Two nucleotides in box C/C′ are 100% conserved (`.GA...`) and a third base is strongly conserved (`....G.`) to establish a consensus sequence of nGAnGn. Analysis of box D revealed a single perfectly conserved nucleotide (`.....A`) and three strongly conserved nucleotides (`..CUG.`) to establish a consensus sequence of nnCUGA. Mutagenesis of both box C and D sequences was then performed by replacing selected nucleotides with a C residue (or an A residue when appropriate). Results demonstrated that individual nucleotides critical for U14 processing are present in each consensus

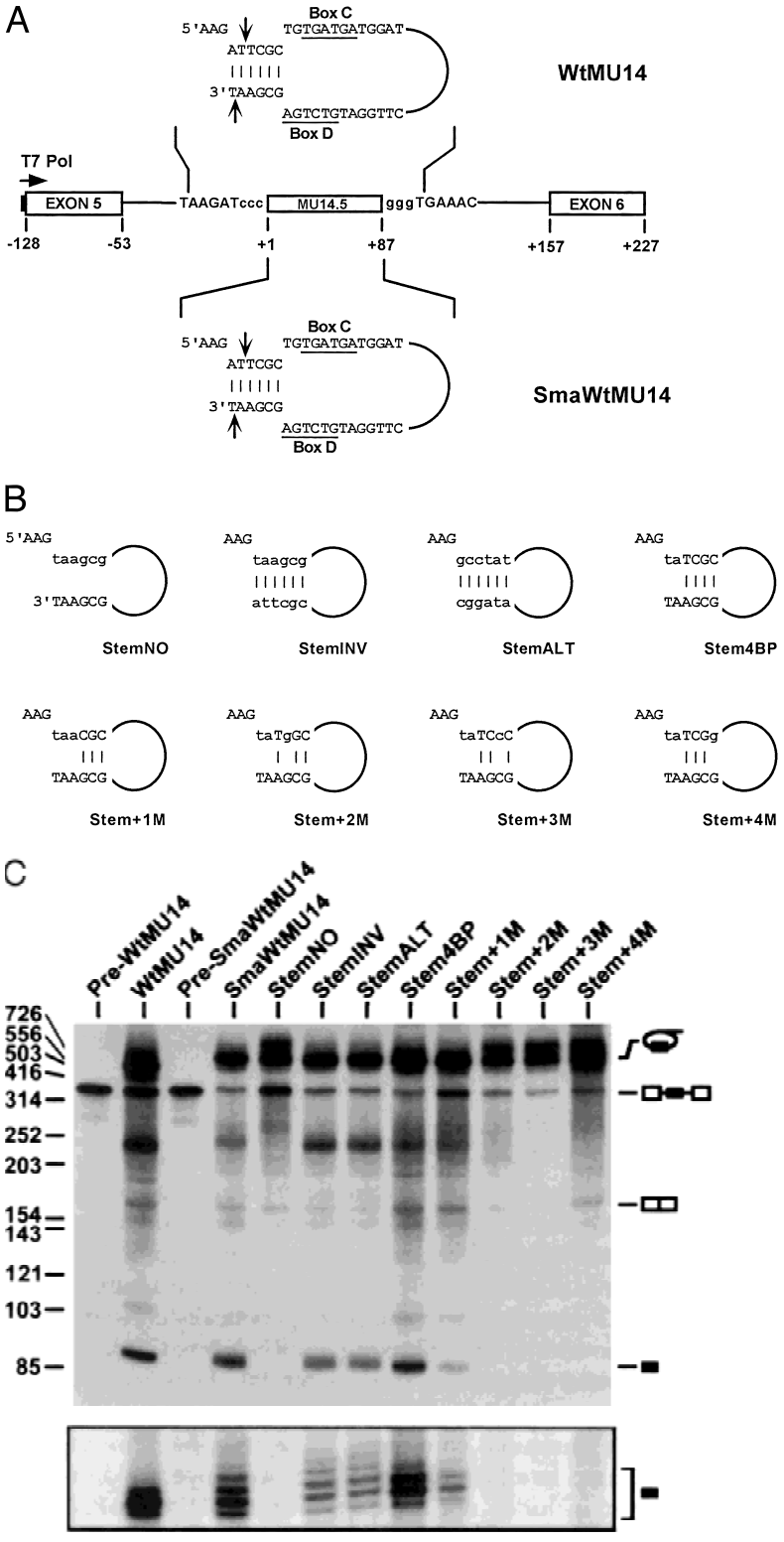

FIGURE 1. Processing of U14 snoRNAs with altered terminal stem sequences and structures. **A:** Schematic representation of the hsc70 pre-mRNA: U14 snoRNA cloning vector with wild-type U14 snoRNA sequence and *Sma* I cloning site. Mouse hsc70 pre-mRNA exons 5 and 6 (open boxes) flank hsc70 intron 5 with included wild-type U14 snoRNA (WtMU14). The terminal core motif consisting of base paired 5′ and 3′ termini with flanking nucleotide box C and D sequences is indicated. Cloning plasmid vector pBS+hsc70ΔS with *Sma* I cloning site and included mouse U14.5 snoRNA wild-type sequence is indicated (SmaWtMU14). Capital letters are the wild-type sequence and small letters are the *Sma* I cloning site. Arrows designate the mature 5′ and 3′ ends of mouse U14.5. **B:** U14 snoRNA mutant constructs with altered terminal stem structure. Capital letters are the U14 wild-type sequence and small letters are mutated nucleotides. **C:** Processing of U14 snoRNAs with altered terminal stems. ^{32}P-radiolabeled hsc70 pre-mRNA: U14 snoRNA precursors containing the various altered terminal stem sequences/structures were injected into *Xenopus* oocyte nuclei and, after 18–20 h incubation, total RNA was prepared and processed transcripts were analyzed by PAGE. Molecular weight markers and processing products are represented at the sides (exons are open boxes and U14 snoRNA is a solid box). Uninjected/unprocessed precursor controls (pre-WtMU14 and pre-SmaWtMU14) as well as injected and processed radiolabeled transcripts are designated at the top. The lower panel (boxed) is a high-resolution polyacrylamide sequencing gel to reveal the size heterogeneity of the processed U14 snoRNAs.

sequence (Fig. 3). These essential box C and D nucleotides match the phylogenetically conserved nucleotides shown in Table 1. Many of these same nucleotides are important for the in vivo stability of yeast U14, a nonintronic snoRNA species (Huang et al., 1992; see the Discussion).

Spatial positioning of boxes C and D with respect to the terminal stem is important for U14 processing

The spatial relationship or positioning of boxes C and D with respect to the terminal stem was examined next. Switching the respective positions of boxes C

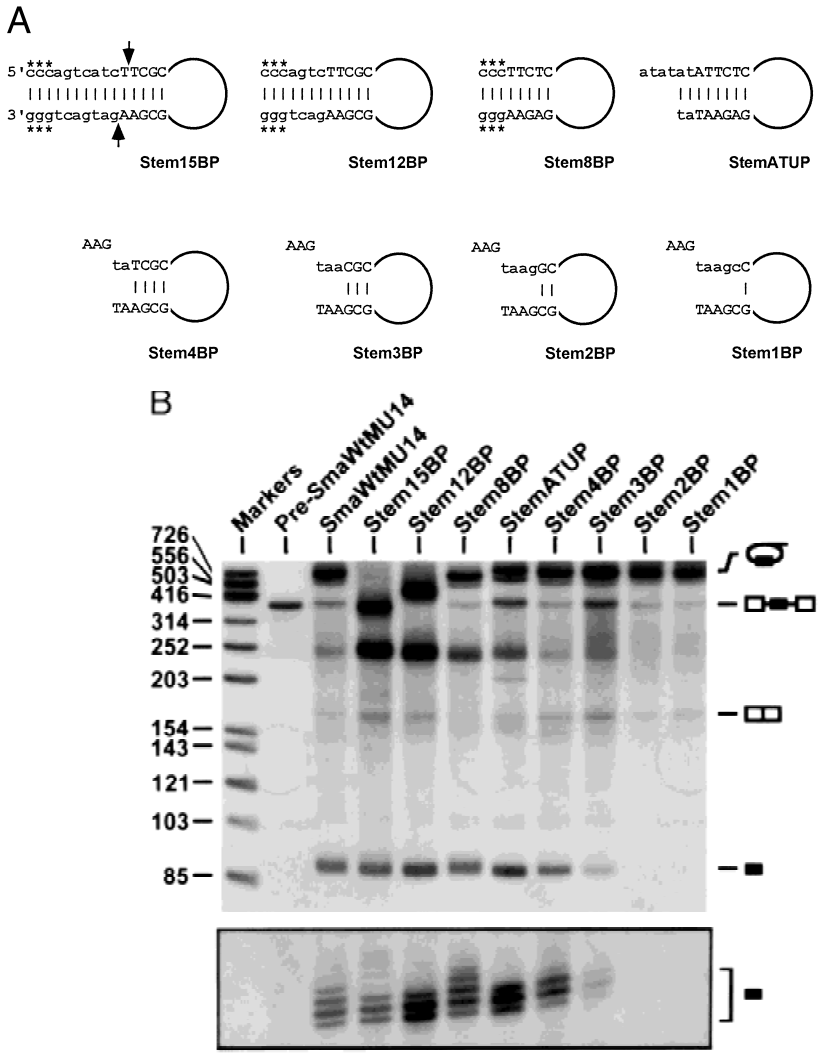

FIGURE 2. Processing of U14 snoRNAs with varying length terminal stem structures. **A:** Schematic representation of U14 snoRNAs with varying length terminal stem structures. Asterisks designate the GC bases of the *Sma* I cloning site and arrows designate the mature U14 5′ and 3′ termini. Capital letters are wild-type sequence and small letters are the mutated nucleotides. **B:** Processing of U14 snoRNAs with variable terminal stem lengths. ^{32}P-radiolabeled hsc70 pre-mRNA:U14 snoRNA precursors possessing terminal stems of varying lengths were injected into *Xenopus* oocyte nuclei and, after 18–20 h incubation, total RNA was prepared and processed transcripts were analyzed by PAGE. Molecular weight markers and processing products are represented at the sides (exons are open boxes and U14 snoRNA is a solid box). Uninjected/unprocessed precursor control (pre-SmaWtMU14) as well as injected and processed radiolabeled transcripts are designated at the top. The lower panel (boxed) is a high-resolution polyacrylamide sequencing gel to reveal the size heterogeneity of the processed U14 snoRNAs.

and D resulted in the loss of U14 processing, demonstrating that each box could not substitute for the other (Fig. 4B). Comparison of available snoRNA sequences revealed that box C and the terminal stem are always separated by two nucleotides of nonconserved sequence (data not shown). Removal of these two spacer nucleotides or insertion of two additional nucleotides eliminated U14 processing (Fig. 4B). Similar inspection of box D positioning revealed the absence of spacer nucleotides between the terminal helix and this consensus sequence (data not shown). U14 processing did occur, although at a reduced level, when two nucleotides were inserted between the stem and box D, but was disrupted when six nucleotides were inserted (Fig. 4B).

DISCUSSION

We have shown previously that the RNA elements necessary for intronic U14 processing are contained within the terminal regions of the U14 snoRNA molecule itself and that the replacement of either box C or D with oligo C disrupts U14 processing (Watkins et al., 1996). The experiments reported here demonstrate the importance of individual sequences and structures within these terminal regions required for intronic U14 synthesis. Defined critical elements include a base paired, 5′,3′-terminal stem and phylogenetically conserved nucleotides within flanking snoRNA box C and D consensus sequences. These essential elements constitute a minimal U14 snoRNA element that we now define as the "terminal core motif." The terminal core motif is found in all 30 U14 snoRNAs sequenced to date. This list includes not only the intronic U14 snoRNAs of vertebrates, but the nonintronic species of yeast and plants. Although yeast and plant U14 snoRNAs are nonintronic species, they are also processed RNAs derived from precursor, multi-snoRNA transcripts (Zagorski et al., 1988; Leader et al., 1994). Previous work in yeast has shown that a terminal stem and boxes C and D are important for U14 accumulation (Huang et al., 1992). Thus, similar elements are required for both intronic and nonintronic U14 snoRNA biosynthesis.

TABLE 1. Phylogenetic comparison of snoRNA box C and box D nucleotide sequences.[a]

RNA	Organism	Box C(C′) UGAUGA	Box D GUCUGA	RNA	Organism	Box C(C′) UGAUGA	Box D GUCUGA
U3	Human	G..A..	.....	U26	Mouse	G.....	C.....
	Rat	G..C..	.....		Human	G.....	U.....
	Mouse	G..C..	.....	U27	Mouse	.....	UA....
	X. laevis	G..A..	AG....		Human	.....	UA....
	T. aestivum	C	.....	U28	Mouse	....UU	U.....
	L. esculentum	.....	.....		Human	....UU	U.....
	A. thaliana	.....	.....	U29	Mouse	.....	AC....
	D. discoideum	.....	U..GU.		Human	.....	AC....
	S. cerevisiae	.....	.....	U30	Mouse	.....	A.....
	S. pombe	.....	U.....		Human	.....	A.....
	T. thermophilia	....U.	U.....	U31	Mouse	.....	U.....
	T. brucei	.....	UC.A..		Human	.....	U.....
	T. cruzi	.....	UC.A..	U32	Mouse	....C.	C.....
	L. collosoma	...A..	UC.A..		Human	.....	C.....
U8	Human	.....	A.....	U33	Mouse	.....	A.....
	Mouse	.....	A.....		Human	.....	A.....
	Rat	.....	A.....	U34	Mouse	.....U	.G....
	X. laevis	.....	U.....		Human	.....U	.G....
U13	Human	.....	.....	U35	Mouse	.....U	AG....
U14	Human (2)	.....	.....		Human	.....U	AG....
	Human	.....	..U...	U36	Mouse	.....U	CC....
	Mouse (3)	.....	.....		Chicken	.....U	CA....
	Rat (3)	.....	.....	U37	Hamster	.....	.....
	Hamster (3)	.....	.....		Human	.....	.....
	X. laevis (4)	.....	.....	U38	Mouse	.....	A.....
	X. laevis	A.....	.....		Human	.....	.....
	S. cerevisiae	.....	.....	U39	Mouse	.....	AG....
	Trout (5)	.....	.....		Human	.....	AG....
	Maize	.....U	.....	U40	Mouse	.....	CA....
	Maize (4)	.....	.....		Human	.....	CA....
	Potato	.....	.....	U41	Human	.....	GG....
U15	Human A	...A..	.....	U42	Human	.....G	CA....
	Human B	.....	.....	U43	Human	.....	U.....
	X. laevis (4)	.....	.....	U44	Human	.....	AA....
U16	*X. laevis* (2)	.....U	U.....	U45	Human	.....U	C.....
	Human	.....U	U.....		Human	.....U	CA....
U18	*X. laevis* (5)	.....	U.....	U46	Human	.....	CA....
	X. tropicalis (3)	.....U	U.....	U47	Human	.....U	U.....
	Human	.....	U.....	U48	Human	.....	C.....
	S. cerevisiae	.....	A.....	U49	Human	.....	.....
U20	Human	.....	U.....	U50	Human	.....	U.....
	Mouse	.....	U.....	U51	Human	.....	U.....
	Rat	.....	U.....	U52	Human	.....	U.....
	X. laevis	.....	U.....	U53	Human	.....	.....
U21	Human	.....	.A....	U54	Human	...G..	U.....
	Mouse	.....	.A....	U55	Human	.....	AG....
	Chicken	.....	.A....	U56	Human	.....G	.....
U22	Human	...A..	.....	U57	Human	.....	.C....
	X. laevis	.....	.....	U58	Human	.....	U.....
U24	Chicken	.....U	CG....		Human	.....G	C.....
	Human	.....U	CG....	U59	Human	.....	UA....
	S. cerevisiae	.....U	C.....	U60	Human	.....	.A....
U25	Mouse	.....	C.....	U61	Human	.....	U.....
	Human	.....	C.....	snR190	*S. cerevisiae*	.....	A.....

(*continued*)

TABLE 1. *continued.*

Nucleotide conservation in box C

	Box C sequence					
Nucleotide	U	G	A	U	G	A
A	0.7%	0.0%	100.0%	3.7%	0.0%	81.5%
G	4.4%	100.0%	0.0%	0.7%	97.0%	3.3%
C	0.0%	0.0%	0.0%	1.4%	0.7%	0.7%
U	94.8%	0.0%	0.0%	94.0%	2.2%	15.6%

Nucleotide conservation in box D

	Box D sequence					
Nucleotide	G	U	C	U	G	A
A	14.1%	10.4%	0.0%	2.2%	0.0%	100.0%
G	43.7%	8.1%	0.0%	0.7%	99.3%	0.0%
C	13.3%	5.2%	98.3%	0.0%	0.0%	0.0%
U	28.9%	76.3%	0.7%	97.1%	0.7%	0.0%

[a]*Abbreviations*: *X. laevis*, *Xenopus laevis*; *T. aestivum*, *Triticum aestivum*; *L. esculentum*, *Lycopersicon esculentum*; *A. thaliana*, *Arabidopsis thaliana*; *D. discoideum*, *Dictyostelium discoideum*; *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces pombe*; *T. thermophilia*, *Tetrahymena thermophilia*; *T. brucei*, *Trypanosoma brucei*; *T. cruzi*, *Trypanosoma cruzi*; *L. collosoma*, *Leptomonas collosoma*. U3 sequences compiled in Hartshorne and Agabian (1994). U8 sequences from Reddy et al. (1985), Tyc and Steitz (1989), and Peculis and Steitz (1993). U13 sequence from Tyc and Steitz (1989). U14 sequences from Zagorski et al. (1988), Liu and Maxwell (1990), Zafarullah et al. (1992), Leader et al. (1994), and Xia et al. (1995). U15 sequences from Pellizzoni et al. (1994) and Tycowski et al. (1993). U16 sequences from Fragapane et al. (1993). U18 sequences from Prislei et al. (1993) and A.G. Balakin and M.J. Fournier (pers. comm.). U20 sequences from Nicoloso et al. (1994). U21 sequences from Qu et al. (1994). U22 sequences from Tycowski et al. (1993) and Tycowski et al. (1994). U24 sequences from Qu et al. (1995). U25–U31 sequences from Tycowski et al. (1996). U32–U40 sequences from Nicoloso et al. (1996). U41–U61 sequences from Kiss-Laszlo et al. (1996). SnR190 sequence from Zagorski et al. (1988). Numbers in parentheses indicate number of snoRNAs with identical box C and D primary sequences.

Mutagenesis of the mouse terminal core motif, supported by an extensive phylogenetic analysis of 135 snoRNA primary sequences (Table 1), has resulted in a better understanding of those sequences and structures that are important for intronic U14 processing. Previous analysis of nonintronic yeast U14 has defined specific nucleotides in boxes C and D that are required for accumulation as well, and has demonstrated that base pairing of the 5′ and 3′ termini is important. Results presented here are the first to detail the role these same consensus sequences and terminal stem play in intronic snoRNA processing. The GA nucleotides (`.GA...`) of box C were shown previously to be important for nonintronic yeast U14 accumulation (Huang et al., 1992), and we have shown that these same two nucleotides are also required for intronic mouse U14 processing. We have also defined an additional or third nucleotide in box C (`....G.`) required for mouse U14 processing. Mutagenesis of this third nucleotide in yeast had no effect upon nonintronic U14 accumulation. However, this box C nucleotide in yeast was mutated to a U rather than the C substitution of the mouse box C (Huang et al., 1992). Examination of the box C phylogenetic data shown in Table 1 indicates that this position is more tolerant of a U than a C substitution. This suggests that this same nucleotide might also be utilized for nonintronic U14 accumulation. The GA nucleotides of box D (`....GA`) were shown previously to be important for nonintronic yeast U14 accumulation, and we have now demonstrated that these same nucleotides are required for intronic U14 processing. However, mouse U14 processing is dependent upon an additional two nucleotides of box D (`..CU..`) that are not as critical for yeast U14 accumulation. The four essential nucleotides of mouse U14 box D defined in this study (CUGA) match the phylogenetically conserved nucleotides shown in Table 1. No effect has been observed for either mouse or yeast U14 synthesis when the less well-conserved `GU....` nucleotides were mutated. Our recent inspection of box D sequences (data not shown) suggests that nucleotide conservation at these positions is probably due to their utilization in base pairing to rRNA for 2′-*O*-methylation (Kiss-Laszlo et al., 1996; Nicoloso et al., 1996; Tollervey, 1996; J. Ni & M.J. Fournier, pers. comm.) rather than for box D function in processing events.

Work in yeast indicated that base pairing of the 5′ and 3′ termini is important for U14 accumulation (Jarmolowski et al., 1990; Huang et al., 1992). Our detailed analysis of the mouse 5′ and 3′ termini has now extended this initial observation and demonstrated that a minimum of three contiguous base pairs positioned specifically with respect to flanking boxes C and D is necessary. The first base pair of the four-base pair stem typically found in mature U14 snoRNAs (farthest removed from boxes C and D) is dispensable. Our results also demonstrate that the spacer nucleotides between box C and the terminal stem are critical and these results agree with phylogenetic analysis of terminal core motif structure. The lack of spacer nucleo-

A

WtBoxC	UGAUGA
BoxC6c	cccccc
BoxCnGAnnn	c..ccc
BoxCG8c	.c....
BoxCA9c	..c...
BoxCG11c	....c.
BoxCA12c	c
Consensus	nGAnGn

WtBoxD	GUCUGA
BoxD6c	cc.ccc
BoxDnnCnGA	cc.c..
BoxDU78c	.c....
BoxDC79a	..a...
BoxDU80c	...c..
BoxDG81c	....c.
BoxDA82c	c
Consensus	nnCUGA

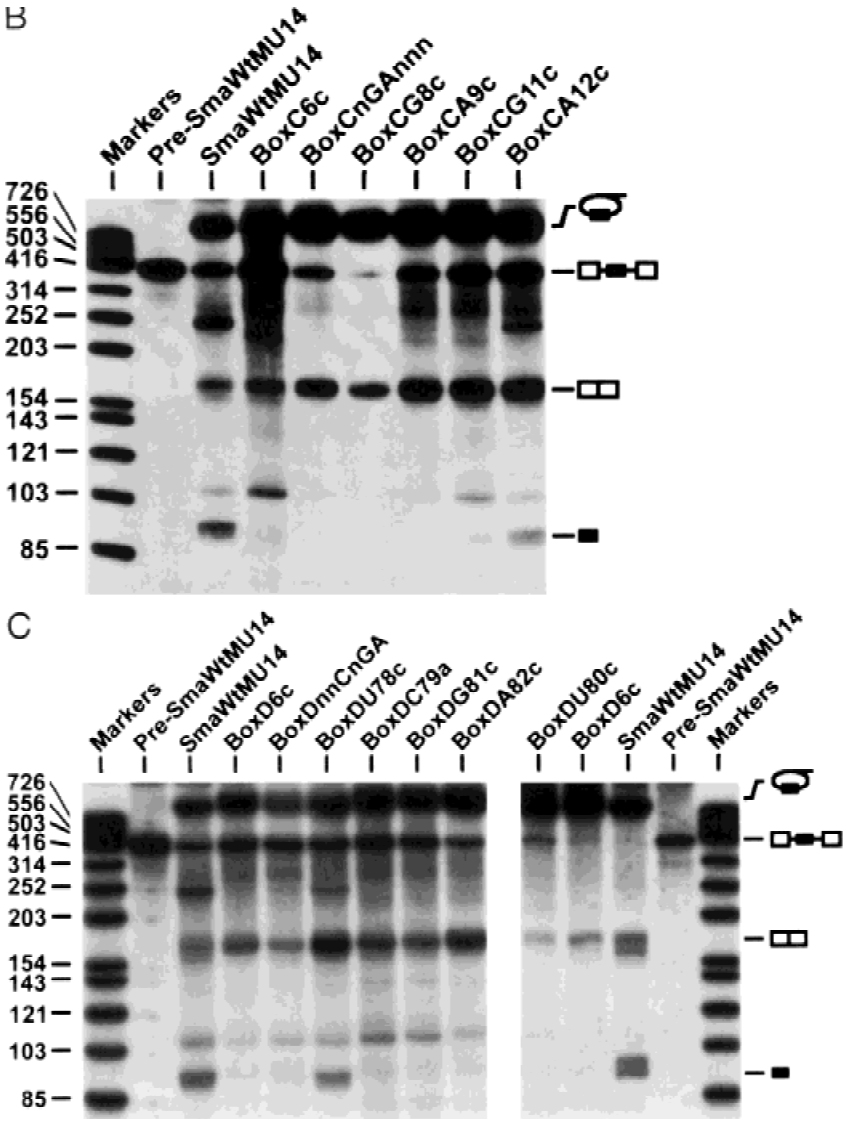

FIGURE 3. Processing of U14 snoRNAs with altered nucleotide boxes C and D. **A:** Presentation of U14 snoRNA box C and D mutants. Wild-type mouse U14.5 snoRNA box C and D sequences are indicated at the top (capital letters). Nonmutated nucleotides are indicated by periods and altered nucleotides are indicated as C or A base mutations (small letters). The box C and D consensus sequences determined from Table 1 are indicated below. **B,C:** Processing of U14 snoRNA precursors with mutated box C (B) or box D (C) nucleotides. ^{32}P-radiolabeled hsc70 pre-mRNA:U14 snoRNA precursors with various mutated box C or D sequences were injected into *Xenopus* oocytes and, after 18–20 h incubation, total RNA was prepared and processed transcripts were analyzed by PAGE. Molecular weight markers (M) are designated at the side. Uninjected/unprocessed precursor (pre-SmaWtMU14) and processed mutant constructs are designated at the top. Processing products are represented at the side (exons are open boxes and U14 snoRNA is a solid box). The additional upper band observed for processed U14 snoRNA from the SmaWtMU14 precursor transcript on the right side of panel C corresponds to incomplete processing at the 3′ terminus of the U14 snoRNA (see also Fig. 4A). This longer product is more predominant with oocytes isolated from some frogs.

tides between box D and the terminal stem is seen for all U14 snoRNAs, although this consensus sequence can be displaced slightly from the terminal stem without disrupting processing. The base pairing of the terminal stem in the precursor transcript may also be extended well beyond the 5′ and 3′ termini of the mature U14 snoRNA without affecting processing. Thus, the structure or double-strandedness of the terminal stem itself does not solely determine the 5′- and 3′-terminal nucleotides of the fully mature U14 snoRNA molecule. We believe that the terminal core motif establishes a recognition site for box C/D-binding proteins. Our earlier work demonstrating the binding of a *trans*-acting factor (protein?) to the U14 terminal core motif (Watkins et al., 1996) is consistent with this hypothesis. We suspect that these proteins protect the U14 termini from trimming exonucleases and thus define the terminal nucleotides of the mature molecule. Therefore, neither base paired extensions of the terminal helix nor unpaired bases at the terminal end of the

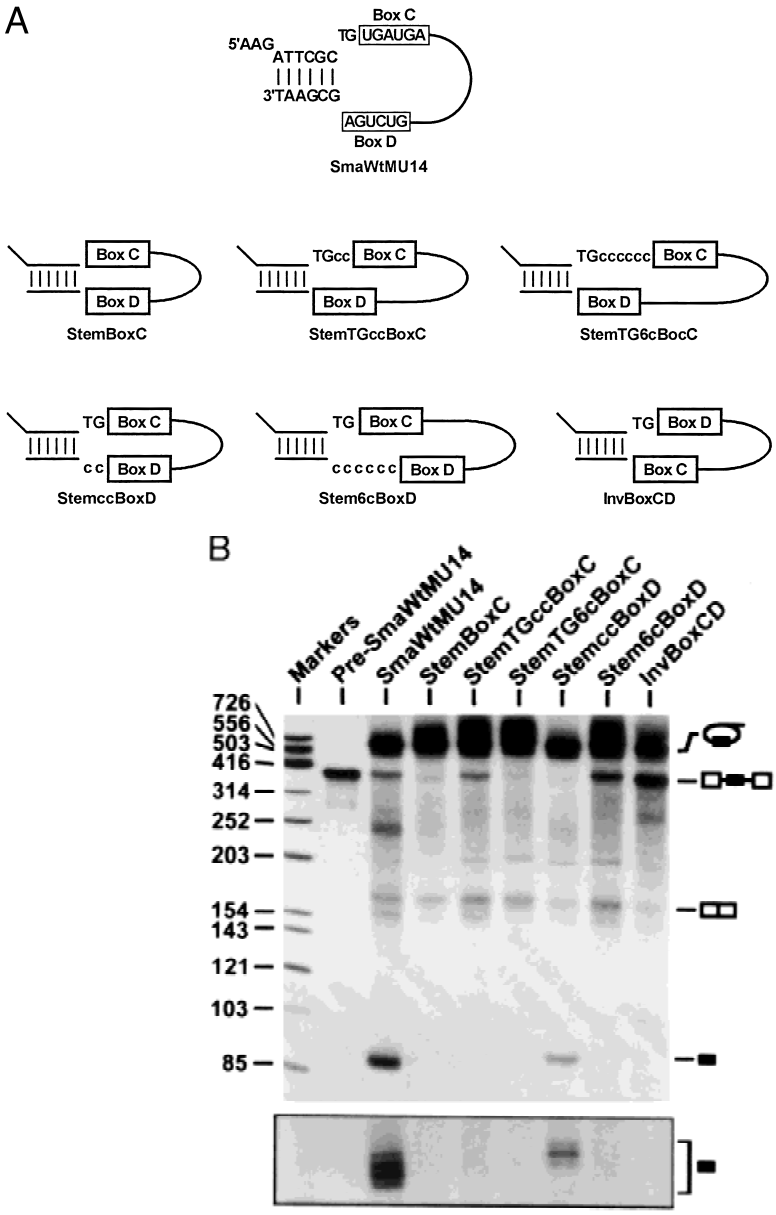

FIGURE 4. Processing of U14 snoRNAs with altered box C and D positioning. **A:** Schematic representation of U14 snoRNAs with altered box C and D positioning. SmaWtMU14; wild-type mouse U14.5 snoRNA box C and D positioning with respect to the terminal stem. Boxes C and D are indicated in the open blocks. Positions of boxes C and D are altered with respect to the terminal stem by removing wild-type nucleotides (capital letters) or inserting additional C bases (small letters). Mutant designations are indicated below each construct. **B:** Processing of U14 snoRNAs with altered box C and D positioning. ^{32}P-radiolabeled hsc70 pre-mRNA:U14 snoRNA precursors with positionally altered boxes C or D were injected into *Xenopus* oocytes and, after 18–20 h incubation, total RNA was prepared and U14 processing was analyzed by PAGE. Molecular weight markers (M) and RNA processing products (exons are open boxes and U14 snoRNA is a solid box) are designated at the side. Uninjected/unprocessed precursor control (pre-SmaWtMU14) as well as injected and processed radiolabeled transcripts are designated at the top. The lower panel (boxed) is a high-resolution polyacrylamide sequencing gel to reveal the size heterogeneity of the processed U14 snoRNAs.

mature helix have an effect upon U14 processing ultimately and, more importantly, upon establishment of U14's 5′- and 3′-terminal nucleotides. A similar processing mechanism has also been proposed for the ACA snoRNAs of both vertebrates and yeast (Cecconi et al., 1995; Balakin et al., 1996).

Finally, the terminal core motif is present in the majority of box C/D-containing snoRNAs, suggesting that this conserved element is important for the processing of many snoRNA species. This list includes both intronic and nonintronic species from both higher and lower eukaryotes. However, not all box C/D snoRNAs possess the 5′,3′-terminal stem structure (Kiss-Laszlo et al., 1996; Nicoloso et al., 1996; Tycowski et al., 1996). We believe that those snoRNAs lacking an external stem utilize an internal secondary/tertiary structure to juxtaposition boxes C and D for protein binding (Watkins et al., 1996). At the present time, it is difficult to differentiate between snoRNA processing events, such as site-specific cleavage and endo/exonuclease activities, and the stabilization or protection of the mature snoRNA from trimming exonucleases by bound snoRNA-binding proteins. Our previous work has indicated that protein binding to the terminal core motif is essential for intronic mouse U14 processing. Thus, we believe that the binding of protein(s) to the termi-

nal core motif to form an RNP complex serves as a key regulatory step or control point in snoRNA biosynthesis.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotide primers used in this work are designated numerically below for reference.

1. SmaWtMU14 UP, AAGATTCGCTGTGATGATGGAT;
2. SmaWtMU14 DW, ATTCGCTCAGACATCCAAGGAA;
3. StemINV UP, AAGTAAGCGTGTGATGATGGATTCC AAAACC
4. StemINV DW, TAAGCGTCAGACATCCAAGGAAGG AAC;
5. StemALT UP, AAGGCCTATTGTGATGATGGATTCC AAAACC;
6. StemALT DW, GCCTATTCAGACATCCAAGGAAGG AAC;
7. Stem4BP UP, AAGTATCGCTGTGATGATGGATTC;
8. Stem3BP UP, AAGTAACGCTGTGATGATGGATTCC;
9. Stem+2M UP, AAGTATGGCTGTGATGATGGATTCC;
10. Stem+3M UP, AAGTATCCCTGTGATGATGGATTCCAA;
11. Stem+4M UP, AAGTATCGGTGTGATGATGGATTCCAA;
12. Stem15BP UP, AGTCATCTTCGCTGTGATGATGGA;
13. Stem15BP DW, AGTCATCTTCGCTCAGACATCCA;
14. Stem12BP UP, AGTCTTCGCTGTGATGATGGA;
15. Stem12BP DW, AGTCTTCGCTCAGACATCCA;
16. Stem8BP UP, TTCGCTGTGATGATGGATTCCA;
17. Stem8BP DW, TTCGCTCAGACATCCAAGGAAG;
18. AStem8BP UP, ATATATATTCGCTGTGATGATGGA;
19. AStem8BP DW, ATATTCGCTCAGACATCCA;
20. Stem2BP UP, AAGTAAGGCTGTGATGATGGATTCCAA;
21. Stem1BP UP, AAGTAAGCCTGTGATGATGGATTCCAA;
22. BoxC6c UP, AAGATTCGCTGCCCCCCTGGATTCCAA AACCATTCGTAG;
23. BoxD6c DW, ATTCGCGGGGGGATCCAAGGAAGGAA CTAGCCAA;
24. BoxCnGAn UP, AAGATTCGCTGCGACCTGGATTCC AAAACCATTCG;
25. BoxCG8c UP, AAGATTCGCTGTCATGATGGATTCCAA;
26. BoxCA9c UP, AAGATTCGCTGTGCTGATGGATTCC AAA;
27. BoxCG11c UP, AAGATTCGCTGTGATCATGGATTCC AAAA;
28. BoxCA12c UP, AAGATTCGCTGTGATGCTGGATTCC AAAACC;
29. BoxDnGA DW, ATTCGCTCGTGGATCCAAGGAAGG AACTAG;
30. BoxDU78c DW, ATTCGCTCAGGCATCCAAGGAAG GAA;
31. BoxDC79a DW ATTCGCTCATAATCCAAGGAAGGAA;
32. BoxDG81c DW ATTCGCTGAGACATCCAAGGAAGG;
33. BoxDA82c DW, ATTCGCGCAGACATCCAAGGAAGG;
34. BoxDU80c, ATTCGCTCGGACATCCAAGGAAGGAA;
35. StemBoxC UP, AAGATTCGCTGATGATGGATTCCAA AAC;
36. StemTGccBoxC UP, AAGATTCGCTGCCTGATGATGG ATTCCAA;
37. StemTG6cBoxC UP, AAGATTCGCTGCCCCCCTGATG ATGGATTCCAA;
38. StemccBoxD DW, ATTCGCGGTCAGACATCCAAGG AAGGAACTAG;
39. Stem6cBoxD DW, ATTCGCGGGGGGTCAGACATCC AAGGAAGGAACTAGC;
40. InvBox CD UP, AAGATTCGCTGGTCTGATGGATTCC AAAACCATTCG;
41. InvBoxCD DW, ATTCGCTCATCAATCCAAGGAAGG AACTAGCCAA.

Plasmid construction

All U14/hsc70 constructs were made by cloning PCR-amplified wild-type or mutant U14 coding sequences into the *Sma* I site of pBS+hsc70ΔS as described previously (Watkins et al., 1996). The plasmid pBS+hsc70ΔS contains 73 nt of exon 5, intron 5, and 75 nt of exon 6 of the mouse hsc70 gene, where the U14 snoRNA-coding sequence in intron 5 was replaced with a *Sma* I restriction site (Fig. 1A). Primer sets for each mutant were as follows: plasmid SmaWTU14, 1 and 2; plasmid StemINV, 3 and 4; plasmid StemALT, 5 and 6; plasmid Stem4BP, 7 and 2; plasmid Stem3BP, 8 and 2; plasmid Stem+2M, 9 and 2; plasmid Stem+3M, 10 and 2; plasmid Stem 4M, 11 and 2; plasmid Stem15BP, 12 and 13; plasmid Stem12BP, 14 and 15; plasmid Stem8BP, 16 and 17; plasmid AStem8BP, 18 and 19; plasmid Stem2BP, 20 and 2; plasmid Stem1BP, 21 and 2; plasmid BoxC6c, 22 and 2; plasmid BoxD6c, 23 and 1; plasmid BoxCnGAnnn, 24 and 2; plasmid BoxCG8c, 25 and 2; plasmid Box CA9c, 26 and 2; plasmid Box CG11c, 27 and 2; plasmid Box CA12c, 28 and 2; plasmid BoxDnnnnGA, 29 and 1; plasmid BoxDU78c, 30 and 1; plasmid BoxDC79a, 31 and 1; plasmid BoxDG81c, 32 and 1; plasmid BoxDA82c, 33 and 1; plasmid BoxDU80c, 34 and 1; plasmid StemBoxC, 35 and 2; plasmid StemTGcc BoxC, 36 and 2; plasmid StemTG6cBoxC, 37 and 2; plasmid StemccBoxD, 38 and 1; plasmid Stem6cBoxD, 39 and 1; plasmid InvBoxCD, 40 and 41.

In vitro RNA transcription, microinjection of *Xenopus* oocytes, and RNA analysis

All plasmids were linearized with *Ecl*136 II restriction endonuclease. ^{32}P-radiolabeled RNA transcripts were synthesized and injected into stage VI *Xenopus* oocytes as described previously (Watkins et al., 1996). Injected oocytes were incubated for approximately 20 h at 18 °C before total oocyte RNA was prepared (Watkins et al., 1996). U14 snoRNA processing was assessed by resolution of extracted, radiolabeled RNAs on 8% polyacrylamide/7 M urea gels. For a more detailed resolution of processed U14 snoRNA transcripts, electrophoresis was performed using longer RNA sequencing gels.

ACKNOWLEDGMENTS

We thank Matt Andrews, Donna Newman, and Joyce Liu for helpful discussions and technical assistance. This work was supported by National Science Foundation grant MCB 9304572 (E.S.M.).

Received June 21, 1996; returned for revision July 12, 1996; revised manuscript received October 8, 1996

REFERENCES

Bachellerie JP, Michot B, Nicoloso M, Balakin A, Ni J, Fournier MJ. 1995. Antisense snoRNAs: A family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci 20:261–264.*

Balakin AG, Smith L, Fournier MJ. 1996. The RNA world of the nucleolus: Two major families of small RNAs defined by different box elements with related functions. *Cell 86*:823–834.

Baserga SJ, Gilmore-Hebert M, Yang XW. 1992. Distinct molecular signals for nuclear import of the nucleolar snRNA, U3. *Genes & Dev 6*:1120–1130.

Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I. 1996. Processing of the intron-encoded U16 and U18 snoRNAs: The conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J 15*:1121–1131.

Cecconi F, Mariottini P, Amaldi F. 1995. The *Xenopus* intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes. *Nucleic Acids Res 23*:4670–4676.

Fragapane P, Prislei S, Michienzi A, Caffarelli E, Bozzoni I. 1993. A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA. *EMBO J 12*:2921–2928.

Hartshorne T, Agabian N. 1994. A common core structure for U3 small nucleolar RNAs. *Nucleic Acids Res 22*:3354–3364.

Huang GM, Jarmolowski A, Struck JC, Fournier MJ. 1992. Accumulation of U14 small nuclear RNA in *Saccharomyces cerevisiae* requires box C, box D, and a 5′, 3′ terminal stem. *Mol Cell Biol 12*:4456–4463.

Hughes JMX, Ares MJ. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5′ external transcribed spacer of yeast pre-ribosomal RNA and prevents formation of 18S ribosomal RNA. *EMBO J 10*:4231–4239.

Jarmolowski A, Zagorski J, Li HV, Fournier MJ. 1990. Identification of essential elements in U14 RNA of *Saccharomyces cerevisiae*. *EMBO J 9*:4503–4509.

Kass S, Tyc K, Steitz JA, Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of pre-ribosomal RNA processing. *Cell 60*:897–908.

Kiss T, Filipowicz W. 1993. Small nucleolar RNAs encoded by the introns of the human cell cycle regulatory gene RCC1. *EMBO J 12*:2913–2920.

Kiss T, Filipowicz W. 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes & Dev 9*:1411–1424.

Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of pre-ribosomal RNA: A novel function for small nucleolar RNAs. *Cell 85*:1077–1088.

Leader D, Sanders JF, Waugh R, Shaw P, Brown JW. 1994. Molecular characterisation of plant U14 small nucleolar RNA genes, closely linked genes are transcribed as a polycistronic U14 transcript. *Nucleic Acids Res 22*:5196–5203.

Leverette RD, Andrews MT, Maxwell ES. 1992. Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock premessenger RNA. *Cell 71*:1215–1221.

Li HV, Zagorski J, Fournier MJ. 1990. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol 10*:1145–1152.

Liu J, Maxwell ES. 1990. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res 18*:6565–6571.

Lygerou Z, Allmang C, Tollervey D, Seraphin B. 1996. Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science 272*:268–270.

Maxwell ES, Fournier MJ. 1995. The small nucleolar RNAs. *Annu Rev Biochem 35*:897–934.

Morrissey JP, Tollervey D. 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol Cell Biol 13*:2469–2477.

Nicoloso M, Caizergues-Ferrer M, Michot B, Azum MC, Bachellerie JP. 1994. U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. *Mol Cell Biol 14*:5766–5776.

Nicoloso M, Qu LH, Michot B, Bachellerie JP. 1996. Intron-encoded, antisense small nucleolar RNAs: The characterization of nine novel species points to their direct role as guides for the 2′-*O*-ribose methylation of rRNAs. *J Mol Biol 260*:178–195.

Peculis BA, Steitz JA. 1993. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the *Xenopus* oocyte. *Cell 73*:1233–1245.

Peculis BA, Steitz JA. 1994. Sequence and structural elements critical for U8 snRNP function in *Xenopus* oocytes are evolutionarily conserved. *Genes & Dev 8*:2241–2255.

Pellizzoni L, Crosio C, Campioni N, Loreni F, Pierandrei-Amaldi P. 1994. Different forms of U15 snoRNA are encoded in the introns of the ribosomal protein S1 gene of *Xenopus laevis*. *Nucleic Acids Res 22*:4607–4613.

Prislei S, Michienzi A, Presutti C, Fragapane P, Bozzoni I. 1993. Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. *Nucleic Acids Res 21*:5824–5830.

Qu LH, Henry Y, Nicoloso M, Michot B, Azum MC, Renalier MH, Caizergues-Ferrer M, Bachellerie JP. 1995. U24, a novel intron-encoded small nucleolar RNA with two 12 nt. long, phylogenetically conserved complementarities to 28S rRNA. *Nucleic Acids Res 23*:2669–2676.

Qu LH, Nicoloso M, Michot B, Azum MC, Caizergues-Ferrer M, Bachellerie JP. 1994. U21, a novel small nucleolar RNA with a 13 nt. complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals. *Nucleic Acids Res 22*:4073–4081.

Reddy R, Henning D, Busch H. 1985. Primary and secondary structure of U8 small nuclear RNA. *J Biol Chem 260*:10930–10935.

Savino R, Gerbi SA. 1990. In vivo disruption of *Xenopus* U3 snRNA affects ribosomal RNA processing. *EMBO J 9*:2299–2308.

Schmitt ME, Clayton DA. 1993. Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol 13*:7935–7941.

Sollner-Webb B. 1993. Novel intron-encoded small nucleolar RNAs. *Cell 75*:403–405.

Steitz JA, Tycowski KT. 1995. Small RNA chaperones for ribosome biogenesis. *Science 270*:1626–1627.

Terns MP, Grimm C, Lund E, Dahlberg JE. 1995. A common maturation pathway for small nucleolar RNAs. *EMBO J 14*:4860–4871.

Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J 6*:4169–4175.

Tollervey D. 1996. Small nucleolar RNAs guide ribosomal RNA methylation. *Science 273*:1056–1057.

Tyc K, Steitz JA. 1989. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J 8*:3113–3119.

Tycowski KT, Shu MD, Steitz JA. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & Dev 6*:1120–1130.

Tycowski KT, Shu MD, Steitz JA. 1994. Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science 266*:1558–1561.

Tycowski KT, Shu MD, Steitz JA. 1996. A mammalian gene with intron instead of exons generating stable RNA products. *Nature 379*:464–466.

Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES. 1996. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA 2*:118–133.

Wise JA, Weiner AM. 1980. *Dictyostelium* small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. *Cell 22*:109–118.

Xia L, Liu J, Sage C, Trexler EB, Andrews MT, Maxwell ES. 1995. Intronic U14 snoRNAs of *Xenopus laevis* are located in two different parent genes and can be processed from their introns during early oogenesis. *Nucleic Acids Res 23*:4844–4849.

Zafarullah M, Wisniewski J, Shworak NW, Schieman S, Misra S, Gedamu L. 1992. Molecular cloning and characterization of a constitutively expressed heat shock cognate hsc71 gene from rainbow trout. *Eur J Biochem 204*:893–900.

Zagorski J, Tollervey D, Fournier MJ. 1988. Characterization of an SNR gene locus in *Saccharomyces cerevisiae* that specifies both dispensible and essential small nuclear RNAs. *Mol Cell Biol 8*:3282–3290.