Molecular Microbiology (1991) 5(7), 1621–1625 ADONIS 0950382X91001797

Thirty-three nucleotides of 5' flanking sequence including the 'TATA' box are necessary and sufficient for efficient U2 snRNA transcription in *Schizosaccharomyces pombe*

T. Dandekar and D. Tollervey*
European Molecular Biology Laboratory, Postfach 10.22.09, D-6900 Heidelberg, Germany.

Summary

We have sequenced the 5' flanking region of the U2 gene and compared this with the 5' flanking sequences of other snRNA genes from *Schizosaccharomyces pombe*. This revealed no regions of clear homology 5' to a region surrounding the 'TATA' box at –32 to –29. Deletion analysis shows that a 5' flanking region extending to only –33 is sufficient for accurate and efficient transcription of U2 in *Schizosaccharomyces pombe*.

Introduction

The small nuclear RNAs (snRNAs), a group of stable low molecular-weight RNAs, are found in the nuclei of all eukaryotes examined to date. The most-studied snRNAs – U1, U2, U4, U5 and U6 – are all required for pre-messenger RNA splicing *in vitro* (reviewed by Maniatis and Reed, 1987) and *in vivo* (reviewed by Guthrie, 1988; Woolford, 1989). In recent years the transcription signals of genes encoding snRNAs have attracted considerable interest. Sequence comparisons and functional analyses show that snRNA genes of higher eukaryotes lack the 'TATA' sequences present upstream of most genes transcribed by RNA polymerase II, and instead have two important regions designated the proximal sequence element (PSE) and the distal sequence element (DSE). It is likely that the PSE and DSE play roles related to the TATA box and enhancer elements, respectively, of mRNA coding genes (Tebb and Mattaj, 1988). In contrast, snRNA genes in the budding yeast *Saccharomyces cerevisiae* have TATA sequences located between –80 and –100 (see Bally *et al.*, 1988, for a compilation of such sequences), although results of their functional analysis have not been reported. In the fission yeast *Schizosaccharomyces pombe*, the polymerase II-transcribed snRNA genes (U1–U5) also contain TATA sequences (Brennwald *et al.*, 1988; Dandekar *et al.*, 1989; Porter *et al.*, 1990). As is the case with mRNA coding genes (Russell, 1983), the TATA sequences in *S. pombe* snRNA genes lie closer to the transcription initiation site than do those of *S. cerevisiae*. Here we report the first functional analysis of the transcription signals in a *S. pombe* snRNA gene.

Received 15 January, 1991; revised 19 March, 1991. *For correspondence. Tel. (6221) 387264; Fax (6221) 387306; E-mail Dandekar@EMBL.Bitnet and Tollervey@EMBL.Bitnet.

Results and Discussion

The cloning of the *snu2* gene has previously been reported (Dandekar and Tollervey, 1989), and the sequence of the coding region together with the 5' and 3' flanking sequences is shown in Fig. 1. The sequence of the U2 RNA has been reported (Brennwald *et al.*, 1988); the sequence shown in Fig. 1 differs in three positions from that previously reported (see the legend to Fig. 1).

Computer searches revealed no sequences in *snu2* with clear homology to consensus sequences for the DSE (TATGCAAAY; Carbon *et al.*, 1987) or PSE (TC(A/C)CTCTCCYYARG; Carbon *et al.*, 1987) of vertebrate snRNA genes. In an effort to identify other transcription signals, the 5' flanking sequence of *snu2* was aligned against the 5' flanking regions of *snu1*, *snu3* and *snu4*. For U1 and U2, 5' flanking sequences extending to –386 were compared. The complete *snu1* sequence determined, including the newly sequenced upstream region (an additional 147 upstream nucleotides), has been deposited in the EMBL/GenBank/DDBJ Nucleotide Sequence Data Libraries under the accession number X55773. No differences were detected within the region previously sequenced (Porter *et al.*, 1990). In the case of *snu4*, a region extending to position –230 is sufficient for efficient expression of the gene (Dandekar *et al.*, 1989) and was used for comparison. For *snu3* the published sequence extends to –315 (Porter *et al.*, 1988), and this region was used for comparison. A region of homology was detected between the first two sequences available, i.e. *snu2* and *snu4*. In the *snu2* sequence this region lies at positions –71 to –56. However, homologous sequences are not found in other snRNA genes, and deletion of this sequence between

```
  -220       -210       -200       -190       -180
snu2 TATATCATAT TGTTTTAAAA AATGTATGAA AATTTTTAGT GTAAAAAAAA

  -170       -160       -150       -140       -130
snu2 ATAATAATAA TGAGTAGCAG TTTTTGTTAA TTTCATGGCC CATGATCCCT
                                                    |
                                                   U2Δ1

  -120       -110       -100        -90        -80
snu2 TCTGGTAACA AGTTTGTCCA AGGCCTCTGA CCACAAAATT GGGACTGCGA
                                                    |
                                                   U2Δ2

   -70        -60        -50        -40        -30
snu2 AGTAATCTCA GGGTTCGGGT TTATTTATGA GATTTACTAT ATATACAGGT
                     |                      |         |
                    U2Δ3                   U2Δ4      U2Δ5

   -20        -10         +1         10         20         30
snu2 ACTCGTGCAT TTGGTTCGGT ATTCTCTCTT TGCCTTTTGG CTTAGATCAA
U2                    Gppp AUUCUCUCUU UGCCUUUUGG CUUAGAUCAA

             40         50         60         70         80
snu2 GTGTAGTATC TGTTCTTTTC AGTTTAATCG CTGAAATCAC CTCACTGAGG
U2   GUGUAGUAUC UGUUCUUUUC AGUUUAAUCG CUGAAAUCAC CUCACUGAGG
         |
         C in M2

             90        100        110        120        130
snu2 TGTTTCCGAT TAATCTTGTT TTTGGTTTGA GTTGGAAAGC CTCTGGCTTG
U2   UGUUUCCGAU UAAUCUUGUU UUUGGUUUGA GUUGGAAAGC CUCUGGCUUG

            140        150        160        170        180
snu2 CTATGCTTTC CGACACTGGT GTTCTTGCTA TTGCACTACT GGCAAGCGAC
U2   CUAUGCUUUC CGACACUGGU GUUCUUGCUA UUGCACUACU GGCAAGCGAC

         187          +10        +20        +30        +40
snu2 GCCGAAT     CTTCTTCTGT TGACAGATTT GTGACGCATG AATTTGACAA
U2   GCCGAAU_OH

            +50        +60
snu2 CATTGGATGG TTTAGGATGT GAGG
                |
               U2-3′Δ
```

Fig. 1. Sequence of *snu2*. The coding sequence of *snu2* is shown together with the 5' flanking sequence extending to –220, and the 3' flanking sequence extending to +64. The end-points of deletions Δ1–Δ5 are shown. A vertical line denotes the 3' C of the *Eco*RI site at the 5' end of each PCR fragment and the 5' T of the *Hin*dIII site at the 3' end of the PCR fragment. The mutation M2 at position +36 of *S. pombe* U2 is also indicated. The three differences from the published RNA sequence (Brennwald *et al.*,1988) are as follows: nucleotide (nt) 84, an additional U; nt 167, only one U instead of two; nt 187, 3' end at U. The predicted U2 RNA sequence is also shown; the end-points of the RNA coding region were determined by direct enzymatic RNA sequencing from the 3' end and by reverse transcriptase dideoxy-nucleotide sequencing of the RNA (Inoue and Cech, 1985) at the 5' end.

```
                  -30
                   |
snu1:  CGTCGTTACTATAAATATGGGTGCTTACACTTACGGTTCTTT:^^^^^^^^^

                  -30
                   |
snu2:  GAGATTTACTATATATACAGGTACTCGTGCATTTGGTTCGGT:^^^^^^^^^
                |        |
              U2Δ4     U2Δ5

                    -30
                     |
snu3:  GTTATTTACTATATATACGGGTCATTATCCTACGAACAAAACACT:^^^^^^

                  -30
                   |
snu4:  CAATTTTACTATAAATGGATGTACAAGAGTTTACCATTCCCC:^^^^^^^^^

-------------------------------------------------------------

       gttaTTTACTATAaATAcRGGTaCTtRtgCTTacgRTTC--Y:^^^^^^^^^
       c            t
```

Fig. 2. Alignment of 5' flanking regions of *snu1*, *snu2*, *snu3* and *snu4*. The alignment of the TATA box regions of the snRNA genes is shown together with the end-points of deletions Δ4 and Δ5. The start of the coding region is denoted by ^^^. Under the four sequences their consensus is given: lower-case letters show nucleotides common to the two sequences, and upper-case letters indicate that three out of four sequences match. Underlining indicates four out of four matches.

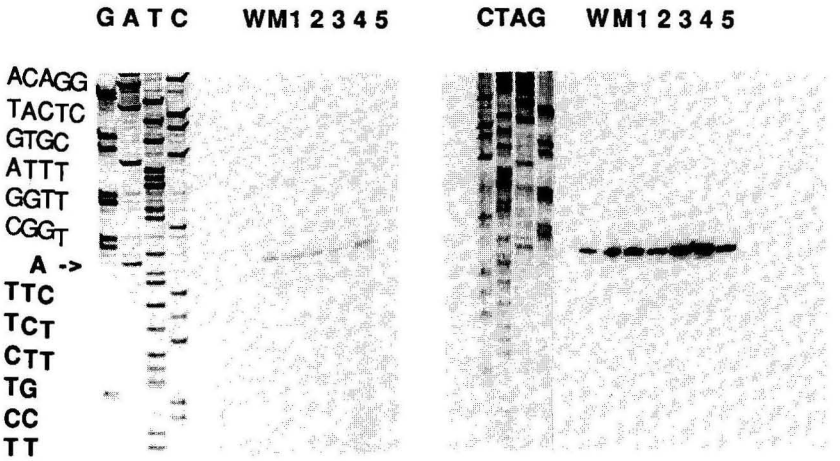

Fig. 3. Primer extension of *snu2* deletion mutants. Left panel. Primer extension using a primer oligonucleotide specific for the M2 mutation present in the *snu2* gene expressed from the promoter deletion constructs.
Right panel. Primer extension using a primer oligonucleotide capable of being extended on the mutant or wild-type U2. On the left margin the U2 gene sequence is given, the start of the RNA coding region is highlighted in bold, and the 5' end of the U2 RNA is marked by an arrow. In both experiments DNA was sequenced in parallel using the same oligonucleotide as for the primer extension (lanes labelled GATC on the left panel and CTAG on the right, according to the dideoxy nucleotide used). The RNA used for each experiment was extracted from the wild type (lanes labelled W), or contained mutant RNA with undeleted upstream region (lanes labelled M) or having progressive deletions (lanes labelled 1 to 5 for the deletions Δ1 to Δ5). A darker exposure of the right panel is shown to allow the DNA sequence to be read.

mutations Δ2 (–75) and Δ3 (–55) (see below) does not reduce transcription. Comparison of the alignments of *snu2* with this region of *snu4*, and with the 5' flanking regions of *snu1* and *snu3*, did not reveal any regions in which the homology is shared by more than two snRNA genes, with the exception of the homology surrounding the TATA box (Fig. 2).

The *snu2* gene is present as a single copy in the *S. pombe* genome (Dandekar and Tollervey, 1989) and gene-disruption experiments show it to be essential for viability (D. Tollervey, unpublished data). In order to follow the transcription of U2 from mutant promoter regions, it was therefore necessary to mark the coding region so as to allow it to be distinguished from the endogenous U2 gene. To do this, a single point mutation was introduced into the *snu2* coding region (Fig. 1). Polymerase chain reaction (PCR) was used to generate a series of fragments containing the mutant *snu2* gene with end-points at various sites in the 5' flanking region (Figs 1 and 2). These fragments were then cloned into the *S. pombe* replicating vector YEpU4 (Dandekar *et al.*, 1989) and transformed into *S. pombe*. This plasmid carries the 2μ origin of replication; in *S. pombe*, such vectors replicate with low copy numbers (2–5 copies per cell) (Beach *et al.*, 1982).

To determine the level of expression of U2 from the mutant promoters, primer extension analyses were performed on total RNA using an oligonucleotide whose 3' end corresponds to the site of the internal point mutation. This oligonucleotide therefore acts as a primer for reverse transcriptase on the mutant U2 (Fig. 3, left panel, lanes labelled M, 1, 2, 3, 4, and 5), but not the wild type (Fig. 3, left panel, lane W). Surprisingly, the level of synthesis of U2 obtained by the primer extension reaction *in vitro* from RNA extracted from respective strains is similar to that from strains carrying plasmids containing 700 bp of 5' flanking sequence (Fig. 3, left panel, lane M) or the Δ1, Δ2, Δ3 or Δ4 mutations (Fig. 3, left panel, lanes labelled 1–4, respectively). In the Δ4 mutation, the 5' flanking sequence of *snu2* commences one nucleotide 5' to the TATA box. This was confirmed by sequencing of the plasmid carrying *snu2*-Δ4, after recovery from the transformed strain of yeast. Deletion of the TATA box in Δ5 abolishes detectable transcription of *snu2* (Fig. 3, left panel, lane 5). Primer extension using an oligonucleotide complementary to both the mutant and wild-type U2 (Fig. 3, right panel) shows that the differences observed in the left panel are not due to variations in RNA recovery. The greater intensity of signal in the right panel lanes M, 1, 2, 3, and 4 is due to the presence of both the wild-type and M2 mutant U2 transcripts in these strains. The 5' ends of the U2 transcripts from Δ1–Δ4 are identical and correspond to the previously reported 5' end of U2 (Ares, 1986).

Northern hybridization using a short oligonucleotide probe complementary to the mutant U2 was also used to assess the expression from Δ1–Δ4 (Fig. 4) *in vivo*. This confirms the results shown in Fig. 3. In addition, the hybridizing bands are identical in size, showing that the 3' ends of the transcripts are also likely to be correct at the nucleotide level. In this experiment the expression from the undeleted plasmid (with 700 bp of upstream sequence) is approximately two-fold higher than from the plasmid carrying the deletion mutations, indicating that sequences further upstream may also influence the level of transcription. However, up to deletion Δ4 there is still stable and efficient transcription of the U2 mutant RNA, which is abolished in deletion Δ5. The sequence around the TATA box, which is conserved between *snu1*, *snu2*,

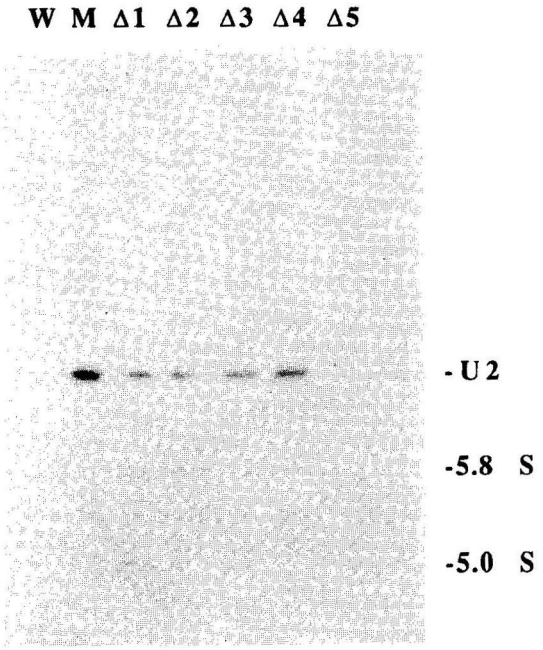

Fig. 4. Northern analysis of *snu2* deletion mutants. RNA was extracted from the wild-type strain (W), or from strains carrying the *snu2* plasmid with 700 bp of 5' flanking region (M) or with upstream deletions (Δ1 to Δ5). RNA was extracted, separated by polyacrylamide gel electrophoresis, and transferred for Northern hybridization with an oligonucleotide specific for the mutant RNA M2.

snu3 and *snu4*, is therefore necessary for efficient transcription of *snu2*. The region of 3' flanking sequence extending to +50 is also sufficient to direct accurate 3' end formation.

We do not know whether the TATA region alone is sufficient to direct transcription, or whether additional sequences are required. These might lie between the TATA box and the transcription start site or within the coding sequence. Many genes transcribed by RNA polymerase III have promoter elements which lie within the coding sequence. However, we have detected no homology to such promoter elements in the snRNA genes described here. Moreover, U1, U2, U3, and U4 from *S. pombe* all have trimethyl guanosine 5'-cap structures (Dandekar and Tollervey, 1989), a feature typical of snRNAs transcribed by RNA polymerase II, rather than the γ-methyl phosphate found on the U6 snRNA (Singh and Reddy, 1989), which is transcribed by RNA polymerase III (Mattaj *et al.*, 1988, reviewed by Parry *et al.*, 1989). Some homology can be detected amongst the *S. pombe* snRNA genes in the region between the TATA box and the transcription start (Fig. 2), suggesting that this region might contain novel promoter elements. It also remains a formal possibility that sequences in the vector act in concert with the TATA box as activators of transcription.

The sequences required for accurate and efficient transcription of *snu2* are shorter than those of many other eukaryotic genes transcribed by RNA polymerase II, including other snRNA genes (Tebb and Mattaj, 1988; reviewed by Parry *et al.*, 1989; Dahlberg and Lund, 1988). However, efficient transcription from short promoter regions is known from other systems, e.g. for the vaccinia virus promoters (Davison and Moss, 1989). Moreover, some (10% of wild-type) transcription from a 44-nucleotide promoter region is observed for U2 RNA transcription in the plant *Arabidopsis thaliana* (Vankan and Filipowicz, 1988).

It has been proposed (Dahlberg and Lund, 1988) that the specialized transcription signals of vertebrate snRNAs are adaptations to the very high rate of snRNA synthesis. However, the rate of transcription of *S. pombe* snRNA genes (2000 molecules per cell from one gene in 150 min doubling time, or about 1 snRNA per 4.5 s) is not very different from that of HeLa cells (1 million molecules of U1 per cell from 60 genes per diploid cell in 18 h doubling time, or about 1 snRNA per 4 s per gene). Thus, while the snRNA transcription signals of vertebrate UsnRNAs may well be adapted to allow high transcription rates, these can also be obtained from TATA-box-containing snRNA genes.

Experimental procedures

Strains and media

The growth and handling of *S. pombe* has been described previously (Gutz *et al.*, 1974). The strain used for transformation and expression of *snu2* carries h$^+$, *ura4*-Δ18, *leu1-32*, and *ade6-704*, and was kindly provided by Dr J. Kohli.

Mutant construction and plasmids

To introduce mutation M2, a 1.5 kb genomic fragment containing the *snu2* gene in bluescript (KS+) (Stratagene) was converted to the single-stranded form and mutagenized using an oligonucleotide (GGCTTAGATCAAGTGTA**C**TATCTGTTCTTTTCAG) and a site-directed mutagenesis kit (Amersham, UK). This resulted in the change G^{19}–C (the site of the mutation is shown in bold) in the *snu2* coding region. Sequence analysis demonstrated that this was the only mutation in *snu2*. The 1.5 kb genomic fragment was subcloned into the *S. pombe* replicating plasmid, YEpU4 (Dandekar *et al.*, 1989) to generate plasmid pM2, which contains ≈700 bp of 5'- and 600 bp of 3' flanking sequence. The Δ1–Δ5 constructs were generated using PCR with an oligonucleotide (AA**AAGCTT**CCATCCAATGTTGTCAAATTC) which introduces a *Hin*dIII site (bold) at position +50 in the 3' flanking region of *snu2* and oligonucleotides which introduced *Eco*RI sites (bold) into the 5' flanking region of *snu2* at positions –123 (Δ1; AAA**GAATTC**CCTTCTGGTAACAAG), –73 (Δ2; AAA**GAATTC**GCGAAGTAATCTCAGGGTTC), –53 (Δ3; AAA**GAATTC**GGGTTATTTATGAGA), –33 (Δ4; AAA**GAATTC**TATATATACAGGTACTCGTG), and –25 (Δ5; AAA**GAATTC**AGGTACTCGTGCATTTGGTT).

DNA was amplified in a Hybaid thermal cycler for 30 cycles of 15 s at 94°C, 30 s at 40°C, and 1 min at 72°C. The products were cleaved with *Eco*RI and *Hin*dIII, gel-purified and cloned into YEpU4 (Dandekar *et al.*, 1989). Plasmids were introduced into *S. pombe* using the Li-acetate method (Ito *et al.*, 1983).

RNA extraction and analysis

RNA extraction and Northern hybridization were as previously described (Tollervey and Mattaj, 1987). The hybridization probe, specific for the M2 mutation, was of sequence CAGATA**G**TACAC. Primer extension was performed as previously described (Inoue and Cech, 1985) using an oligonucleotide specific for the M2 mutant RNA (CTGAAAAGAACAGATA**G**) or capable of being extended on either RNA (GAACAGATACTACACTTG).

DNA sequencing

DNA was sequenced by the dideoxy chain termination technique (Sanger *et al.*, 1977) using a Sequenase kit (USB).

Acknowledgement

T.D. was supported by a grant from Boehringer Ingelheim Fonds for Basic Medical Research.

References

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

Bally, M., Hughes, J., and Cesareni, G. (1988) snR30: a new, essential small nuclear RNA from *Saccharomyces cerevisiae*. *Nucl Acids Res* **16:** 5291–5303.

Beach, D., Piper, M., and Nurse, P. (1982) Construction of a *Schizosaccharomyces pombe* gene bank in a yeast bacterial shuttle vector and its use to isolate genes by complementation. *Mol Gen Genet* **187:** 326–329.

Brennwald, P., Porter, G., and Wise, J.A. (1988) U2 small nuclear RNA is remarkably conserved between *Schizosaccharomyces pombe* and Mammals. *Mol Cell Biol* **8:** 5575–5580.

Carbon, P., Murgo, S., Ebel, J.P., Krol, A., Tebb, G., and Mattaj, I.W. (1987) A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA Polymerase III and U2 snRNA by RNA Polymerase II. *Cell* **51:** 71–79.

Dahlberg, J.E., and Lund, E. (1988) The genes and transcription of the major small nuclear RNAs. In *Structure and Function of Major and Minor Small Ribonucleoprotein Particles.* Birnstiel, M.L. (ed.). Heidelberg: Springer-Verlag, pp. 38–70.

Dandekar, T., and Tollervey, D. (1989) Cloning of *Schizosaccharomyces pombe* genes encoding the U1, U2, U3 and U4 snRNAs. *Gene* **81:** 227–235.

Dandekar, T., Ribes, V., and Tollervey, D. (1989) *Schizosaccharomyces pombe* U4 small nuclear RNA closely resembles vertebrate U4 and is required for growth. *J Mol Biol* **208:** 371–379.

Davison, A., and Moss, B. (1989) Structure of vaccinia virus late promoters. J. Mol. Biol. **210:** 771–784.

Guthrie, C. (1988) Genetic analysis of yeast small nuclear RNAs. In *Structure and Function of Major and Minor Small Ribonucleoprotein Particles.* Birnstiel, M.L. (ed.). Heidelberg: Springer-Verlag, pp. 196–213.

Gutz, H., Heslot, H., Leupold, U., and Loprieno, N. (1974) *Schizosaccharomyces pombe.* In *Handbook of Genetics.* Volume 1. King, R.C. (ed.). New York: Plenum Press, pp. 395–446.

Inoue, T., and Cech, T.R. (1985) Secondary structure of the circular form of the tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA* **82:** 648–652.

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* **153:** 163–168.

Maniatis, T., and Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325:** 673–678.

Mattaj, I.W., Dathan, N.A., Parry, H.D., Carbon, P., and Krol, A. (1988) Changing the RNA polymerase specificity of U snRNA gene promoters. Cell, **55:** 435–442.

Parry, H.D., Scherly, D., and Mattaj, I.W. (1989) 'Snurpogenesis': the transcription and assembly of UsnRNP components. *Trends Biochem Sci*, **14:** 15–19.

Porter, G.L., Brennwald, P.J., Holm, K.A., and Wise, J.A. (1988) The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes. *Nucl Acids Res* **16:** 10131–10152.

Porter, G.L., Brennwald, P., and Wise, J.A. (1990) U1 small nuclear RNA from *Schizosacchaaromyces pombe* has unique and conserved features and is encoded by an essential single copy gene. *Mol Cel Biol* **10:** 2874–2881.

Russell, P. (1983) Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. *Nature* **301:** 167–169.

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

Singh, R., and Reddy, R. (1989) γ-monomethyl phosphate: A cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci USA* **86:** 8280–8283.

Tebb, G., and Mattaj, I.W. (1988) Positionally exact initiation is required for the formation of a stable RNA polymerase II transcription complex *in vivo*. *EMBO J* **7:** 3785–3792.

Tollervey, D., and Mattaj, I.W. (1987). Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J* **6:** 469–476.

Vankan, P., and Filipowicz, W. (1988) Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts. *EMBO J* **7:** 791–799.

Woolford, J.L. (1989) Nuclear pre-mRNA splicing in yeast. *Yeast* **5:** 439–457.