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 (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.

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 (TATGCAAAV; Carbon et al., 1987) or PSE (TC(A/C)CTCCTCYYARG; 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
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 EcoRl site at the 5' end of each PCR fragment and the 5' T of the HindIII 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 dideoxynucleotide sequencing of the RNA (Inoue and Cech, 1985) at the 5' end.

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.
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,
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+, ura4Δ18, 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 (GGCTTAGCTATCAGGTATATGTTCTTTCCAG) and a site-directed mutagenesis kit (Amersham, UK). This resulted in the change G10→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 1.5kb 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 (AAAGACTCAGGTATGTCACAATT) which introduces a HindIII site (bold) at position +50 in the 3' flanking region of snu2 and oligonucleotides which introduced EcoRI sites (bold) into the 5' flanking region of snu2 at positions −123 (Δ1; AAAGAATTCCCTTCTGGTAAAGAG), −73 (Δ2; AAAGAATTCCGCGAAGTAAATCTCAGGGTTC), −53 (Δ3; AAAGAATTCCCGTATTTGAGA), −33 (Δ4; AAAGAATTCCATATACAGGTTACCTCGT), and −25 (Δ5; AAAGAATTCCAGGTACTCGTCATTTGTT).
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 EcoRI and HindIII, gel-purified and cloned into YEp4 (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 CAGATACTAC. Primer extension was performed as previously described (Inoue and Cech, 1985) using an oligonucleotide specific for the M2 mutant RNA (CTGAAAAGAACAGATAQ) or capable of being extended on either RNA (GAACAGATACTACCTTG).

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


