

An essential domain in *Saccharomyces cerevisiae* U14 snoRNA is absent in vertebrates, but conserved in other yeasts

Dmitry A. Samarsky, Gregory S. Schneider and Maurille J. Fournier*

Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, MA 01003, USA

Received February 26, 1996; Revised and Accepted April 18, 1996

GenBank accession no. U29583

ABSTRACT

U14 is a small nucleolar RNA (snoRNA) required for early cleavages of eukaryotic precursor rRNA. The U14 RNA from *Saccharomyces cerevisiae* is distinguished from its vertebrate homologues by the presence of a stem-loop domain that is essential for function. This element, known as the Y-domain, is located in the U14 sequence between two universal sequences that base pair with 18S rRNA. Sequence data obtained for the U14 homologues from four additional phylogenetically distinct yeasts showed the Y-domain is not unique to *S.cerevisiae*. Comparison of the five Y-domain sequences revealed a common stem-loop structure with a conserved loop sequence that includes eight invariant nucleotides. Conservation of these features suggests that the Y-domain is a recognition signal for an essential interaction. Several plant U14 RNAs were found to contain similar structures, though with an unrelated consensus sequence in the loop portion. The U14 gene from the most distantly related yeast, *Schizosaccharomyces pombe*, was found to be active in *S.cerevisiae*, showing that Y-domain function is conserved and that U14 function can be provided by variants in which the essential elements are embedded in dissimilar flanking sequences. This last result suggests that U14 function may be determined solely by the essential elements.

INTRODUCTION

Several small nucleolar RNAs (snoRNAs) are now known to be involved in processing of ribosomal RNA, and many others are expected to have roles in this and other aspects of ribosome biogenesis (reviewed in 1–3). The snoRNA populations in vertebrates and yeast are complex and a broad range of potential functions is imagined, including folding of pre-rRNA as RNA chaperones, direct roles in cleavage, modification of rRNA nucleotides and assembly of rRNP subunits (1,4,5).

U14 is a phylogenetically conserved snoRNA identified thus far in *Saccharomyces cerevisiae*, vertebrates and plants. Genetic depletion and mutation studies with *S.cerevisiae* showed that U14

is required for growth and that loss or inactivation of U14 disrupts cleavages which form a 20S precursor to 18S RNA (6,7). Comparison of U14 sequences revealed a common helix formed by the 5' and 3' ends, as well as four universal sequence elements, called boxes C and D, and domains A and B. Functional mapping in yeast has shown each of the conserved sequences and terminal stem to be essential (8,9). The box C and box D elements and terminal stem are required for U14 accumulation and are presumed to function in U14 maturation, most likely processing, or incorporation of U14 into a stable snoRNP complex (9). The box D element in *Xenopus* U14 has been shown to be involved in formation of a 5' trimethylguanosine cap, although under unnatural conditions; this phenomenon was observed for *in vitro* synthesized RNA containing a monomethylguanosine cap, following injection into oocyte nuclei (10). This modification might not be relevant for *S.cerevisiae* U14 as the mature RNA lacks a 5' terminal cap. Box C has also been implicated in binding of the nucleolar protein fibrillarin in studies of human U3, and both box elements influence the metabolic stability of *Xenopus* U8 snoRNA (10–12). The two other universal sequences known as domains A and B (13 and 14 nt, respectively) are complementary to conserved, well separated sequences in 18S rRNA. *In vivo* genetic analysis has shown that U14 base pairs with the corresponding sequences in rRNA (13).

The middle portion of the *S.cerevisiae* U14 sequence between domains A and B is considerably longer than the corresponding segment in vertebrate U14 RNAs and accounts for nearly all of the size difference between the larger *S.cerevisiae* RNA and the smaller vertebrate homologues (130 nt for *S.cerevisiae* versus 87 nt for mouse). Chemical and enzymatic probing data showed this segment, called the Y-domain (yeast domain), to be highly structured in solution, forming an extended stem-loop structure (14). The high degree of secondary folding suggests that the Y-domain might participate in either intra- or intermolecular interactions. While no information is yet available about the function of this structure, deletion and substitution mutations have shown that the Y-domain is essential for U14 activity in yeast. Variants with altered Y-domains accumulate normally, but the U14-dependent processing reactions do not occur (8, and W. Liang and M. Fournier, in preparation). Especially impressive was the finding that the Y-domain endows mouse U14 with activity in yeast. Mouse U14, which differs substantially from the

* To whom correspondence should be addressed

yeast sequence in the non-conserved regions, is inactive in *S.cerevisiae* despite the fact that it accumulates at good levels. However, when the Y-domain was spliced into the appropriate region of the mouse U14 RNA, the hybrid RNA could functionally substitute for yeast U14 (15). This finding showed that the Y-domain of *S.cerevisiae* can function in a different U14 context, and further suggested that U14 function may rely solely on the conserved elements.

With a view to gaining insights into the function of the Y-domain, we have analyzed partial or full sequences of several other yeast U14 RNAs. The objectives of this study were: (i) to determine if the Y-domain is limited to *S.cerevisiae* and, (ii) if present in other yeasts, to analyze the phylogenetic sequences to obtain information about the structural requirements for Y-domain activity. The phylogenetically disparate yeasts *Kluyveromyces lactis*, *Candida albicans*, *Pichia pastoris* and *Schizosaccharomyces pombe* were all found to possess a Y-domain containing a non-conserved stem and a highly conserved consensus sequence in the loop. Due to its sequence divergence, the gene encoding U14 was cloned from *S.pombe*. Despite differences in sequence and length, *S.pombe* U14 was able to functionally substitute for the *S.cerevisiae* U14, suggesting that U14 homologues are interchangeable in *S.cerevisiae* as long as they contain a Y-domain structure and the universal elements.

MATERIALS AND METHODS

Strains and media

The *K.lactis* IFO 1090, *C.albicans* IFO 1385 and *S.pombe* 975 yeast strains used for total DNA isolation were obtained from M. Fitzgerald-Hayes; the *P.pastoris* GS 115 strain was provided by T. Mason. All strains were grown on YEPD (1% yeast extract, 2% peptone, 2% glucose) medium at 30°C. The *S.cerevisiae* galactose-dependent strain YS153 (MAT α *ura3-167*, *his3*, *trp1-1*, *HIS3::GAL1::U14*) (8) used for heterologous expression of *S.pombe* U14 RNA was grown on YNB (0.67% yeast nitrogen base) selective media containing 2% galactose or 5% glucose supplemented with necessary amino acids and nucleic acid bases.

Escherichia coli strain DH5 α (*supE44 lacU169* (ϕ 80 *lacZ- Δ M15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was used for most cloning procedures. Another *E.coli* strain, JM16 (*lacU169 dam3 rpsL*; M. J. Fournier, unpublished), was used when isolation of non-methylated plasmid DNA was required. Both bacterial strains were grown on LB (0.5% yeast extract, 1% tryptone, 1% NaCl) medium supplemented with ampicillin (50 μ g/ml), IPTG (isopropylthio- β -D-galactoside, 40 μ g/ml) and X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside, 40 μ g/ml) when necessary. Bacterial strains were transformed with plasmids according to the protocol of D. Hanahan (16). The lithium acetate method (17) was used to introduce plasmid DNA into *S.cerevisiae*.

DNA and RNA purification and manipulations

Plasmid DNA from *E.coli* was isolated by a boiling miniprep method (18). The procedure used to isolate yeast genomic DNA for PCR amplification, Southern analysis and library construction is described by Rose *et al.* (17). Total yeast RNA used for northern analysis was purified by a hot-phenol/glass bead method (19).

PCR amplification and sequence analysis of Y-domain regions

PCR amplification of total *K.lactis*, *P.pastoris*, *C.albicans* and *S.pombe* DNAs was carried out with Taq DNA-polymerase (BRL Life Technologies, Inc.) as follows: 35 cycles of 1 min at 95°C, 1 min at 25°C, 1 min at 70°C. The primers were complementary to the universal domain A and B elements (which are complementary to 18S rRNA). The domain A primer includes three additional nucleotides at the 3' end, and the domain B primer includes four additional nucleotides at the 5' end, corresponding to box D. The sequences of the primers were CATTCTG[T or C][A or G]CTTTCCAC (domain A side) and TCAGACATCCAAG-GAAGG (domain B side).

After amplification, the PCR fragments were treated with T4 DNA-polymerase for 20 min at 15°C in the presence of 0.4 mM dNTPs and then subcloned into the pBluescript IISK(-) vector. The sequences of the cloned fragments were determined using the dsDNA Cycle Sequencing System Kit from BRL Life Technologies, Inc. and pBluescript-specific primers SK and KS.

Preparation of radioactive probes and hybridization procedures

PCR fragments were labeled using [α -³²P]dCTP (800 Ci/mmol; DuPont NEN) and a NEBlot Kit (New England Biolabs) and used for northern analyses of total RNA isolated from the corresponding yeast species. *Saccharomyces cerevisiae* U14 was detected with the sequence-specific probe CGATGGGTTTCGTAAGCGT-ACTCCTACCGTGG. The radiolabeled *S.pombe* fragment was also used for Southern analysis of *S.pombe* DNA. Heterologous expression of *S.pombe* U14 in *S.cerevisiae* was evaluated with radioactive *Clal* DNA fragments (see below) labeled with the aid of random primers (NEBlot Kit). All Southern and northern hybridization procedures were carried out essentially as described previously (20,21).

The restriction map of the *S.pombe* U14 genomic locus was developed from Southern analysis of four single and six double restriction digests, after fractionation in a 1% agarose gel and electrotransfer onto a nylon membrane. The enzymes used were *EcoRI*, *HindIII*, *BamHI* and *SalI*.

Isolation and analysis of the *S.pombe* U14 gene

A *S.pombe* subgenomic library of 1500 colonies containing *HindIII-SalI* DNA fragments of 1.1–1.3 kb (see also results) was prepared in a pUC18 vector essentially as previously described (21) and screened using the radiolabeled PCR-fragment described above. Five identical plasmids were isolated from colonies showing hybridization signals, and one was subjected to a detailed restriction analysis. Fragments of 150–200 bp were subcloned into the pBluescript IISK(-) vector and sequenced as described above. Complete sequence information was developed for both strands of the 1.2 kb *HindIII-SalI* DNA fragment containing the U14 coding region. The sequence data were analyzed with the program developed by the University of Wisconsin Genetics Computer Group (22).

Plasmid constructions used to test expression of *S.pombe* U14 RNA in *S.cerevisiae*

A first attempt to express *S.pombe* U14 RNA in *S.cerevisiae* used plasmid pRSPU14, which was constructed by inserting the 1.2 kb

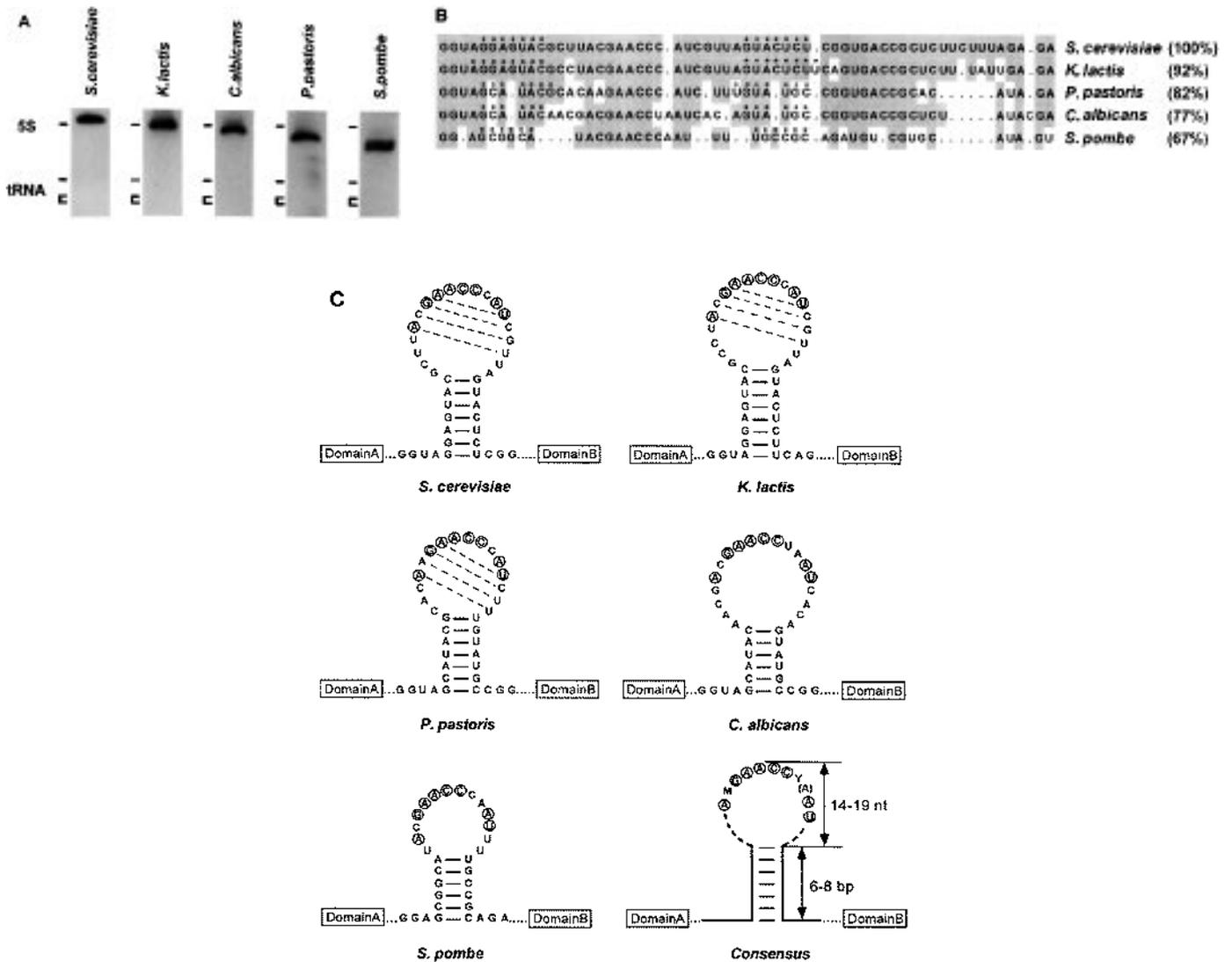


Figure 1. The U14 Y-domain is conserved in four other yeasts. Degenerate oligonucleotides corresponding to U14 domains A and B were used to amplify DNA segments from four yeast species. The PCR products were then used as probes for the homologous U14 RNAs and to test these RNAs for the presence of the Y-domain. **(A)** Northern hybridization. Blots of total RNA from each yeast were hybridized with the radiolabeled PCR fragments. As a control, *S. cerevisiae* RNA was fractionated on the same gel and hybridized with a labeled oligonucleotide specific for U14. 5S RNA and the large and small tRNA sub-classes served as size markers. The 5S and tRNA species from *S. cerevisiae* are 120 and 76–87 nt, respectively; the corresponding RNAs in the other yeasts are similar in size. **(B)** Comparison of the sequences of the amplified regions with the corresponding segment of *S. cerevisiae* U14 RNA. The 5' and 3' ends of the regions shown are flanked by the sequences encoded in the PCR primers. The left border is 3 nt downstream of domain A, and the right border occurs at the 5' end of domain B. Shaded regions indicate fully conserved nucleotides, and dots indicate gaps introduced into the sequence to obtain the best alignment. Inverted repeats are indicated with asterisks. The percentage of identity of the amplified regions compared with the *S. cerevisiae* sequence is given in parentheses. The additional nucleotides in the longer *S. cerevisiae* sequence were not included in the identity calculations. **(C)** Comparison of potential secondary structures of the new sequences with that of the Y-domain from *S. cerevisiae* U14. Fully conserved nucleotides are circled. Dashed lines indicate additional (non-conserved) base-pairing possibilities. In the consensus structure shown in the bottom right, partially conserved nucleotides are identified as Y (pyrimidine) or M (A or C). In all cases a pentanucleotide sequence GAACC is conserved in the mid-portion of the loop region.

HindIII–*SalI* genomic DNA fragment harboring the *S. pombe* U14 gene into yeast vector pRS316 (23) after digestion with *HindIII* and *SalI*. A subsequent experiment was carried out with plasmids in which the *S. pombe* U14 coding sequence was inserted in place of the *S. cerevisiae* U14 segment in the context of a fragment of *S. cerevisiae* DNA. To prepare these plasmids, a 1.2 kb *ClaI* fragment containing *S. cerevisiae* genomic DNA with coding sequences for snR190 and U14 RNAs was isolated from plasmid pJZ45 (8) and inserted into the *ClaI*-site of the vector pRS316 in both orientations (plasmids pCer1 and pCer2) and

separately into the *ClaI*-site of the vector pBluescript IISK(–) to produce plasmid pBCer. A *BclII*–*Bst1107I* fragment (130 bp) containing the coding region for *S. cerevisiae* U14 was replaced in plasmid pBCer with the genomic DNA fragment *BclII*–*PacI* (145 bp) containing *S. pombe* U14, to generate plasmid pBPom; the DNA fragment ends formed by *PacI* were blunted using T4 DNA polymerase prior to cloning. In the final step, the *ClaI* fragment containing the *S. pombe* DNA was cloned from the pBPom plasmid into the pRS316 vector in both orientations to produce the plasmids pPom1 and pPom2.



Figure 2. Structure of the *S.pombe* genomic locus encoding the U14 gene. **(A)** Restriction map of the locus containing the *S.pombe* U14 coding sequence developed from Southern analysis; a 1.2 kb *HindIII*–*SalI* fragment containing the U14 gene was cloned and sequenced. **(B)** Sequence of the *S.pombe* U14 coding and flanking regions. The larger 1.2 kb sequence has been deposited in GenBank, accession number U29583. The RNA coding sequence is indicated by bold italics. Inverted repeats at the ends of the coding region that might participate in formation of an extended terminal stem are marked (*). A putative TATA element is boxed.

RESULTS

The Y-domain is conserved in several yeast species

To determine if the Y-domain occurs in other yeasts, we designed oligonucleotides corresponding to the phylogenetically stable domains A and B and several adjacent conserved nucleotides and used these for PCR amplification of total DNAs from *K.lactis*, *P.pastoris*, *C.albicans* and *S.pombe*. A single band product of ~80–90 bp was obtained in each case, and a positive control reaction with *S.cerevisiae* DNA gave a product of similar size (data not shown). The PCR fragments were radiolabeled and used to probe northern blots of total RNA from the homologous yeast species. Strong signals were identified in the small RNA region in each case, ranging in size from ~110 to ~130 nt. This compares with ~130 nt for the *S.cerevisiae* RNA, detected with a homologous probe (Fig. 1A).

The PCR products were cloned into a pBluescript IISK(–) vector and sequenced. In the 45–60 nt region between conserved domains A and B, the sequences showed different levels of identity to the corresponding 60 nt segment of *S.cerevisiae* U14, ranging from 67% for *S.pombe* to 92% for *K.lactis* (Fig. 1B). This homology, together with the conserved length of the hybridizing RNA species, argues that the PCR fragments obtained were derived from the U14 homologues in the yeast species analyzed.

The Y-domains consist of similar stem–loop structures and a conserved loop sequence

All the new sequences contain inverted repeats of 6–8 nt which can be arranged into secondary structures analogous to the stem of the *S.cerevisiae* Y-domain (Fig. 1C). The stems vary in length and differ in their nucleotide content; nevertheless, the base pairing is always preserved. The content and sizes of the loops are also variable, however, within an 11 nt core region of the loop, there is considerable homology. Eight of these nucleotides are fully conserved, including the contiguous stretch GAACC. Two additional nucleotides in this segment are identical in four of the five yeast species, permitting a consensus sequence to be derived (Fig. 1C). This high degree of sequence relatedness suggests that Y-domain function is conserved in all cases and that the conserved nucleotides in particular are important for this function. Structure probing of protein-free *S.cerevisiae* U14 RNA supported the existence of an additional stem element in the Y-domain involving several nucleotides in the consensus loop region (14). The potential to form the smaller helix is also present in *K.lactis* and *P.pastoris* RNAs, but is difficult to formulate for the U14 species from *C.albicans* and *S.pombe*.

Structure of the S.pombe U14 gene

The amplified U14 segment from *S.pombe* had the least resemblance to the corresponding region in *S.cerevisiae* U14, containing the shortest stem (6 bp), the smallest loop (14 nt), and the shortest distance between the stem and domain B (16 nt). It was therefore of interest to characterize the entire U14 sequence from this yeast and to assess its function in *S.cerevisiae*. A detailed Southern analysis of *S.pombe* DNA, using the homologous PCR fragment as probe, showed the U14 gene to be present in a single copy in the genome (not shown), and allowed us to make a restriction map spanning almost 18 kb of this locus (see Materials and Methods, Fig. 2A). Using this map, a subgenomic library was constructed from *S.pombe* DNA containing *HindIII*–*SalI* fragments ranging in size from 1.1 to 1.3 kb. The library was screened with the radiolabeled *S.pombe* PCR fragment, and a 1.2 kb *HindIII*–*SalI* fragment was cloned. The complete sequence of this fragment was determined (GenBank accession number U29583). The *S.pombe* U14 coding region of ~110 nt was located based on its homology with *S.cerevisiae* U14. The approximate ends of the coding region were defined by comparison with other U14 RNAs including that of *S.cerevisiae*, and estimation of the RNA size (Fig. 2B).

Unlike the U14 locus in *S.cerevisiae* which contains a second snRNA (snR190) only 65–70 nt upstream of U14, the cloned 1.2 kb *S.pombe* fragment does not appear to encode any other small RNA. The sequence shows no homology with snR190, and the fragment detects only U14 in a northern analysis of *S.pombe* total RNA (data not shown).

Since all known vertebrate U14 RNAs, in contrast to *S.cerevisiae* U14, are encoded within introns of protein genes (1,24), we analyzed the 1.2 kb fragment containing the *S.pombe* U14 sequence for this possibility. Sequences matching the canonical intron junction sequences (5' splice site, G/GTAWGT, and 3' splice site TAG, where W = A or T; 25) could be found framing the U14 coding sequence, and sequences adhering to the branchpoint consensus sequence (TRCTAAC, where R = A or G; 25) were found between the junction-like sequences. However, no reasonably long open reading frames could be identified outside the candidate junction sites, arguing that *S.pombe* U14 is not encoded in an intron, but is derived from an independent transcription unit. This interpretation is supported by the presence of a TATA box consensus sequence ~65 bp upstream of the coding region.

Schizosaccharomyces pombe U14 contains, in addition to the Y-domain, all of the conserved features of the other U14 RNAs, i.e., domains A and B, boxes C and D and a putative terminal stem (Fig. 3A). Alignment of the *S.cerevisiae* and *S.pombe* U14 RNA sequences showed an overall identity level of 72% when the

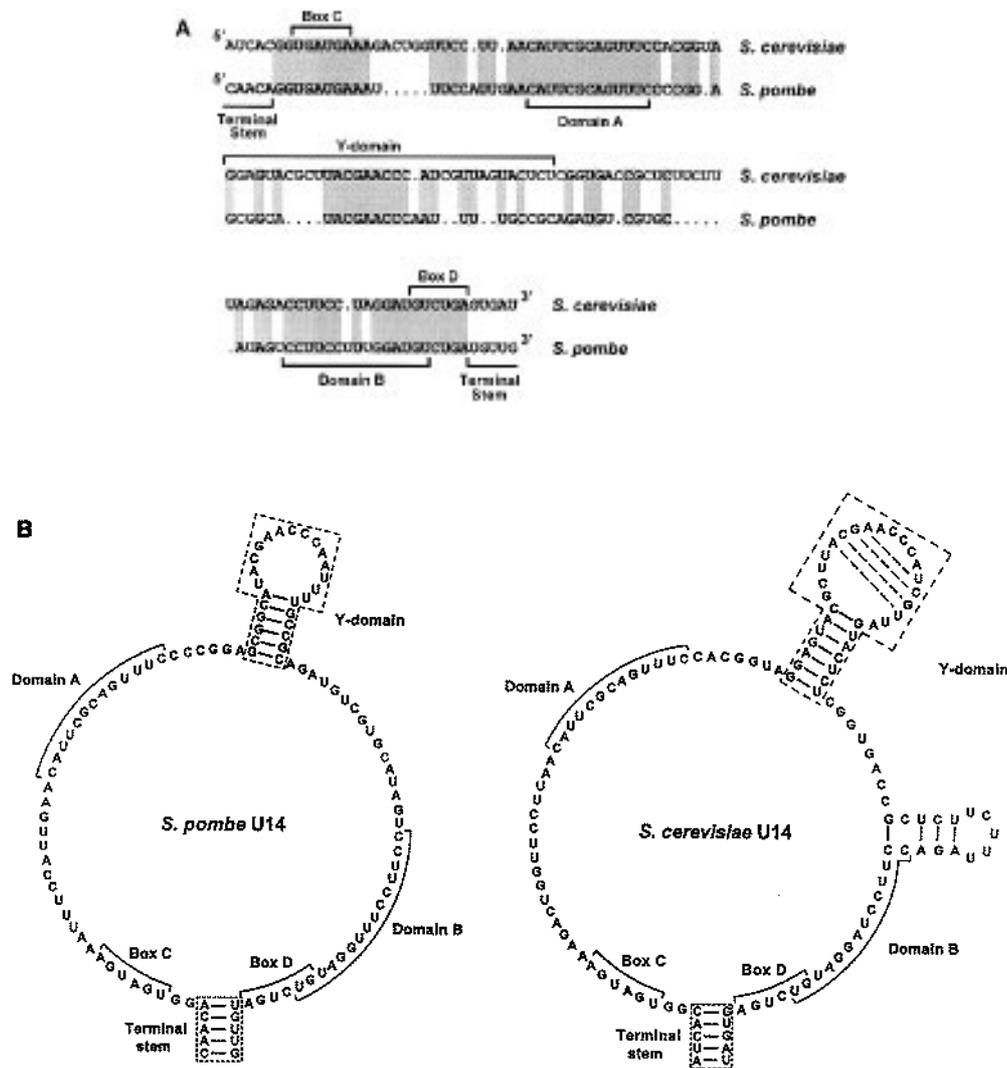


Figure 3. Comparison of *S.pombe* and *S.cerevisiae* U14 RNAs. (A) Alignment of the complete *S.pombe* and *S.cerevisiae* U14 coding sequences. Fully conserved nucleotides are shaded, and dots indicate gaps introduced into a sequence. The universal elements essential for U14 function, i.e. boxes C and D and domains A and B, are almost fully conserved. Sequences corresponding to the terminal stem, though not conserved, preserve base-pairing potential. The total sequences are 72% identical. (B) Proposed secondary structure for *S.pombe* U14. The folding scheme for *S.cerevisiae* U14 is based on the experimentally determined secondary structure (adapted from 14). The potential to form an upper helix in the Y-domain of *S.cerevisiae* U14, indicated by dashed connected lines, is not conserved in *S.pombe*. The terminal stem and Y-domain are contained within the boxed regions.

additional nucleotides in the *S.cerevisiae* homolog are excluded. Most of the signature elements, specifically boxes C and D and domain A, are identical. Domain B of *S.pombe* is one nucleotide longer and contains one nucleotide difference. The complementarity of domain A to *S.pombe* 18S rRNA is perfect. Domain B of *S.cerevisiae* has one mismatch with its complementary element in 18S rRNA. Complementarity is perfect for *S.pombe* at this particular position, with the sequence difference located in U14 rather than 18S rRNA. However, a mismatch also occurs because of the extra nucleotide contained within domain B of *S.pombe* U14. As for *S.cerevisiae* U14, the coding sequence in *S.pombe* is flanked by inverted repeats so that the potential to form an extended helix in U14 precursors is maintained, although with different sequences. The *S.pombe* U14 RNA is 20 nt shorter than the *S.cerevisiae* homologue. While the differences in length are clustered throughout the molecule, ~40% of this difference (8 nt)

occurs in the Y-domain. The remaining differences occur in the non-conserved regions between the various signature elements.

A secondary structure is proposed for the *S.pombe* RNA, based on that established experimentally for *S.cerevisiae* U14 (Fig. 3B). The structure proposed for *S.pombe* differs from that of *S.cerevisiae* in that it does not possess a second stem-loop structure between domains Y and B. This structure is present in *S.cerevisiae* though known to be non-essential (8). Nevertheless, the order, organization, and structure of the phylogenetically conserved elements are the same. As in the *S.cerevisiae* secondary structure, domains A and B of *S.pombe* U14 are predicted to occur in unpaired regions and would thus be open for interaction with the corresponding complementary elements in 18S rRNA, at least in free RNA. Comparison of the *S.cerevisiae* and *S.pombe* sequences and secondary structures raised the question of their interchangeability.

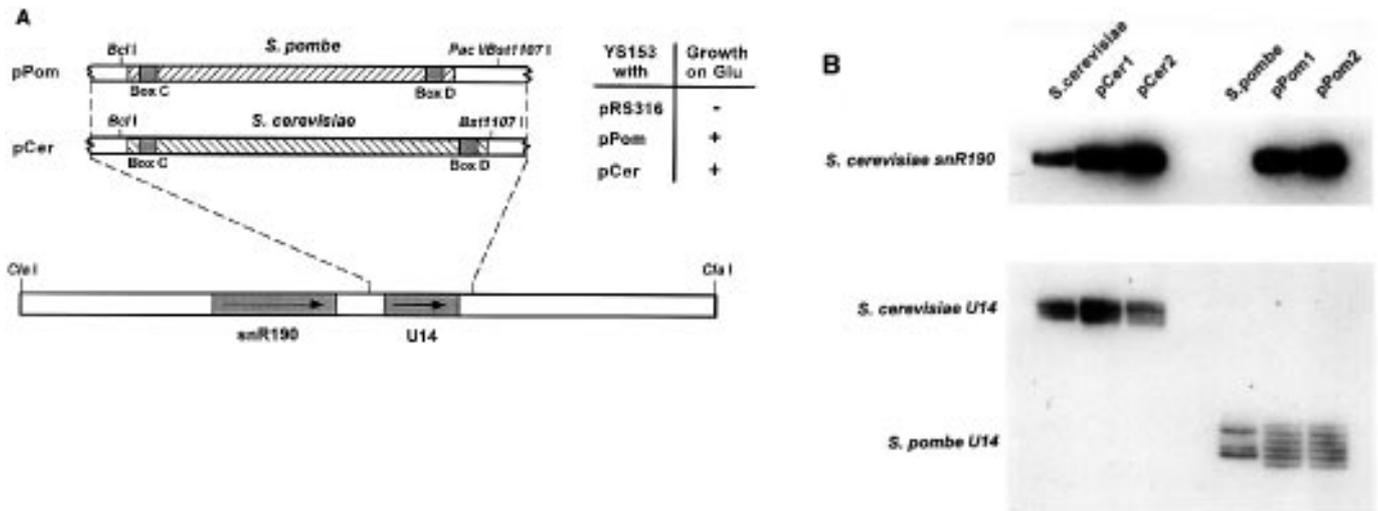


Figure 4. *Schizosaccharomyces pombe* U14 RNA is functional in *S. cerevisiae*. (A) DNA constructions used to test the activity of the *S. pombe* U14 gene in *S. cerevisiae*. Function was examined in a test strain (YS153) whose chromosomal U14 gene is under control of a galactose-dependent promoter and is strongly repressed during growth on glucose (8). The coding region of the *S. pombe* U14 was precisely inserted in place of *S. cerevisiae* U14 DNA and the resulting *Clal* fragment was inserted in both orientations into a yeast shuttle vector (pRS316); plasmids containing the *S. cerevisiae* *Clal* fragment or no insert served as controls. (B) Northern analysis of the *S. cerevisiae* and *S. pombe* U14 RNAs in *S. cerevisiae*. A mixture of radiolabeled *Clal* DNA fragments containing the *S. cerevisiae* snR190 gene and either *S. pombe* or *S. cerevisiae* U14 sequence was used to probe the blot. RNA was prepared from transformants containing plasmids with *S. pombe* U14 DNA (pPom) or *S. cerevisiae* U14 DNA (pCer). Plasmid numbers pPom1, 2 and pCer1, 2 refer to different orientations of the *Clal* insert in the vector. RNA from untransformed *S. cerevisiae* and *S. pombe* strains served as positive controls.

Schizosaccharomyces pombe U14 is functional in *S. cerevisiae*

To determine if *S. pombe* U14 can functionally substitute for *S. cerevisiae* U14, we first tried to express the *S. pombe* U14 gene in *S. cerevisiae* in the context of the original *S. pombe* genomic fragment (*Hind*III–*Sal*I, Fig. 2A) cloned into a *S. cerevisiae* shuttle vector. This construct supported very limited growth of a galactose-dependent *S. cerevisiae* test strain on glucose medium, and a northern analysis with a *S. pombe* U14 probe did not give a readily detectable RNA signal (not shown). We note that growth was screened on glucose medium, whereas RNA was prepared from cells cultured on galactose in order that both positive and negative controls could be obtained. It is possible that *S. pombe* RNA was expressed at a low level in glucose, as the cells would be under greater selective pressure in this condition.

To avoid the possibility that the *S. pombe* transcription signals are poorly utilized in *S. cerevisiae*, we expressed the *S. pombe* U14 gene under control of the *S. cerevisiae* U14 promoter. U14 in *S. cerevisiae* is believed to be co-transcribed with the upstream snR190 and both snR190 and U14 are produced by post-transcriptional processing (1,6,14).

To achieve expression, the *S. cerevisiae* U14 DNA was precisely excised and replaced with the corresponding region of *S. pombe* U14 DNA (Fig. 4A). *Clal* restriction fragments containing homologous and heterologous U14 segments were inserted into yeast shuttle vector pRS316 in both orientations, to assess the potential contributions of plasmid promoters. The resulting plasmids were transformed along with a pRS316 vector into *S. cerevisiae* test strain YS153. This haploid strain contains a chromosomal copy of the U14 gene under GAL1 promoter control. The activity of the plasmid-encoded U14 RNAs was then tested by culturing on glucose-containing medium, which represses expression of the wild-type chromosomal U14 gene (8).

As expected, the positive control cells containing *S. cerevisiae* U14 DNA in either orientation grew well on glucose-containing plates, while the negative control cells transformed with the parent vector failed to grow. Plasmids with *S. pombe* U14 DNA were able to support normal growth in the absence of *S. cerevisiae* U14 RNA (Fig. 4A). This result indicates that *S. pombe* U14 contains all the structural information required for function, despite differences in length and sequence. Because the sequences of the universal U14 elements are identical or nearly so, it is reasonable to conclude that the Y-domains are functionally interchangeable.

The content and quality of the *S. pombe* U14 produced in *S. cerevisiae* were examined by northern analysis (Fig. 4B). The patterns showed the heterologous RNA was produced at levels similar to that of the homologous RNA, with *S. cerevisiae* snR190 snoRNA serving as an internal control. As seen previously, *S. cerevisiae* U14 consists of several subspecies that exhibit length heterogeneity at both ends, most likely due to imprecise processing (14). This heterogeneity is also evident for the *S. pombe* RNA produced in either organism, although with quantitative differences. The major subspecies of *S. pombe* U14 range in size from 108 to 112 nt and the patterns indicate that synthesis of *S. pombe* U14 also involves processing, most likely by machinery closely related to that in *S. cerevisiae*.

DISCUSSION

The Y-domain defines a novel class of phylogenetic elements

Characterization of conserved snoRNA elements is still at an early stage. However, it is already clear that these elements fall into different classes, based on phylogenetic conservation and presence in different snoRNA types. Prior to this work two

distinct classes were evident. One consisted of elements conserved in several different snoRNAs in a wide range of organisms, from lower to higher eukaryotes. The box C and box D elements are in this class, as they are found not only in the U14 RNAs, but many other snoRNAs as well (1). The second class is defined by elements conserved among homologues of a particular snoRNA, but absent thus far from other snoRNAs. The domain A and B elements in U14 are in this class, as are the functionally undefined box A and B elements common to all U3 snoRNAs.

A few elements, such as the stem and loop-IV domain of *S.cerevisiae* U3, which is absent from *S.pombe* U3 and from U3 homologues from non-yeast sources (26,27), fall into a class of organism-specific structures; the evolutionary importance of such elements is unknown. Prior to this study the Y-domain also fell into this category, as it had only been observed in *S.cerevisiae*. Its presence in four other yeasts demonstrates that this element is not organism-specific, and defines a novel category of elements, those that are specific for a group of organisms.

Y-domain consensus structure

Comparison of the Y-domain sequences from the five disparate yeasts yielded a simple consensus structure that will be valuable for future structure–function studies. This structure consists of a stem–loop domain with a conserved sequence in the loop region (Fig. 1C). The individual stems vary in length and sequence but sequence complementarity is preserved, arguing that base pairing is essential for function, perhaps for structural stability. The size and composition of the loop segments also vary; however, there is considerable homology within an 11 nt region of the loop, permitting a consensus sequence to be derived. Eight of these nucleotides are identical in all yeasts, including the contiguous stretch GAACC.

U14 function does not appear to depend on non-conserved sequences

Our finding that *S.pombe* U14 can function in *S.cerevisiae* indicates that this homologue contains all the structural features required for activity, despite size, sequence and folding differences in the Y-domain and different flanking sequences surrounding the other essential elements. *Schizosaccharomyces pombe* U14 is 20 nt shorter than its *S.cerevisiae* counterpart. The fact that it is active in *S.cerevisiae* indicates that the additional 20 nt in *S.cerevisiae* U14 are not essential for function. Eight of the 20 extra nucleotides are within the Y-domain region, indicating that neither the size of the loop nor length of the stem are critical for function, as implied by the consensus structure. The remaining 12 nt fall in other non-conserved regions, providing additional evidence that function is determined by the universal elements.

Our earlier functional analysis of mouse U14 in *S.cerevisiae* showed that the mammalian snoRNA can be expressed and is stable in yeast, but does not provide the function required for rRNA processing (15). Activity was rescued, however, by insertion of the *S.cerevisiae* Y-domain in the same relative position it occupies in its normal host snoRNA. Taken together, the results obtained with the *S.pombe* and the hybrid mouse U14 RNAs argue that the essential elements are able to mediate their activities in an otherwise flexible molecule that varies in size and sequence. The unrelated linking segments might only provide connector function.

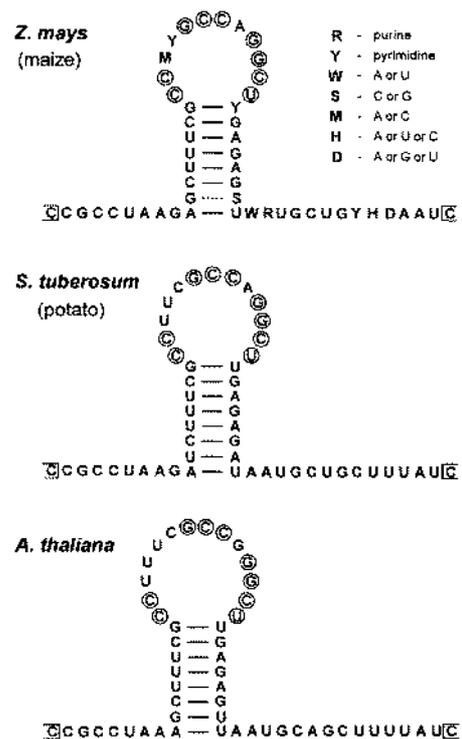


Figure 5. Plant U14 RNAs contain structures similar to the Y-domain. Analysis of recently identified U14 RNA sequences from maize, potato (28) and a partial sequence from *A.thaliana* (29) showed that the regions located between domains A and B can be arranged into secondary structures similar to that of the yeast Y-domain. Fully conserved nucleotides are circled. However, the consensus sequence in the loop region appears to be unrelated to that present in yeast (see text). The maize sequence is a consensus sequence developed from four genes, with non-consensus nucleotides indicated in the adjacent code.

Plant U14 RNAs contain a Y-domain-like structural element

Since this work was completed sequences have become available for U14 RNA from three plants, i.e., from maize (four species), potato and *Arabidopsis thaliana*. The maize and potato sequences came from direct analyses (28), whereas that for *Arabidopsis*, which is partial, was identified by computer screening of the GenBank genomic database (28,29). The identity of the *Arabidopsis* sequence is based on the presence of the signature elements and >95% identity with an 80 nt segment of maize DNA. The plant sequences (~120 nt) are longer than the vertebrate homologues, but similar in size to the yeast RNAs. Each contains the universal U14 sequence elements and the box C–box D–terminal stem motif (28). As with the yeast and vertebrate U14 homologues, considerable variation occurs in the ‘non-conserved’ sequences of the plant RNAs.

We have examined the folding potential of the plant sequences and, to our surprise, determined that each contains a segment located between conserved domains A and B that has the potential to form a stem-loop structure (Fig. 5). Like the Y-domain, the stem elements of the plant domains (all 8 bp) vary in sequence, but maintain complementarity. The loop segments contain 12–14 nt that display considerable homology, with nine of the nucleotides being fully conserved in all three sequences. However, the

consensus sequence that can be derived, C C - - Y G C C R G G C U, exhibits no obvious relationship to that derived for the five yeast species (A M G A A C C Y - A U).

Role of the Y-domain

The vital nature of the Y-domain was shown in previous deletion analyses of *S.cerevisiae* U14 (8). Our present phylogenetic comparison establishes the importance of the Y-domain structure and further implicates conserved nucleotides in the loop sequence in carrying out its function. The conserved nucleotides can be imagined to function in RNA base pairing or recognition by a protein. Base pairing could be intramolecular, involving interaction with another region of U14, or intermolecular, involving interaction with a different RNA, for example, pre-rRNA or another snoRNA. Intramolecular base pairs can be predicted in some cases, but these are short (<5 nt) and are not conserved in *S.pombe* (not shown). Interactions with other RNAs or protein are attractive possibilities. This could be important for many aspects of U14 function, including assembly or transport of the U14 snoRNP, or interaction of this complex with components of the pre-rRNA folding and processing machinery.

The dissimilarity between the loop consensus sequences for yeast and plants argues that the two domains have different recognition and interaction properties, and perhaps even different functions. No functional information is yet available for the Y-like domains in plant U14 RNAs. In time it will be interesting to test the activity of this and the other plant U14 segments in *S.cerevisiae*, but this will be done in the context of higher resolution structure and function studies of the *S.cerevisiae* elements.

Regardless of whether plants contain a functional homologue of the Y-domain, the question remains as to why this element is absent in the U14 RNAs of vertebrates. It is possible that this function is provided by other elements in U14, but the inability of mouse U14 to function in yeast argues that this is not the case, or that the yeast and animal elements are not interchangeable. Folding models of mouse U14 do not reveal any candidate sequence and structure that might specify this function (data not shown). It is more likely that the function of the Y-domain in vertebrates is carried out by some other *trans*-acting molecule or is simply absent altogether. Identifying the role of the Y-domain and characterizing its mechanism of action are important goals for the future.

ACKNOWLEDGEMENTS

We thank Molly Fitzgerald-Hayes and Thomas Mason for providing yeast strains, and John Brown for sharing plant U14 sequence data and helpful discussion. We appreciate the assist-

ance with computer analyses provided by Jingwei Ni and discussions about Y-domain functions with Wen-Qing Liang, both in our laboratory. We also thank Elizabeth Furter-Graves for expert editorial help. This research was supported by NIH grant GM-19351.

REFERENCES

- Maxwell,E.S. and Fournier,M.J. (1995) *Annu. Rev. Biochem.*, **35**, 897–934.
- Sollner-Webb,B., Tycowski,K.T. and Steitz,J.A. (1995) In Zimmermann,R.A. and Dahlberg,A.E. (eds), *Ribosomal RNA: Structure, Evolution, Gene Expression and Function in Protein Synthesis*. CRC Press, Boca Raton, FL, pp. 469–490.
- Gerbi,S.A. (1996) *Biochem. Cell. Biol.*, in press.
- Bachelierie,J.-P., Michot,B., Nicoloso,M., Balakin,A., Ni,J. and Fournier,M.J. (1995) *Trends Biochem. Sci.*, **20**, 261–264.
- Steitz,J.A. and Tycowski,K.T. (1995) *Science*, **270**, 1626–1627.
- Zagorski,J., Tollervey,D. and Fournier,M.J. (1988) *Mol. Cell. Biol.*, **8**, 3282–3290.
- Li,H.V., Zagorski,J. and Fournier,M.J. (1990) *Mol. Cell. Biol.*, **10**, 1145–1152.
- Jarmolowski,A., Zagorski,J., Li,H.V. and Fournier,M.J. (1990) *EMBO J.*, **9**, 4503–4509.
- Huang,G.M., Jarmolowski,A., Struck,J. and Fournier,M.J. (1992) *Mol. Cell. Biol.*, **12**, 4456–4463.
- Terns,M.P., Grimm,C., Lund,E. and Dahlberg,J.E. (1995) *EMBO J.*, **14**, 4860–4871.
- Peculis,B.A. and Steitz,J.A. (1994) *Genes Dev.*, **8**, 2241–2255.
- Baserga,S.J., Yang,X.W. and Steitz,J.A. (1991) *EMBO J.*, **10**, 2645–2651.
- Liang,W.-Q. and Fournier,M.J. (1995) *Genes Dev.*, **9**, 2433–2443.
- Balakin,A.G., Lempicki,R.A., Huang,G.M. and Fournier,M.J. (1994) *J. Biol. Chem.*, **269**, 739–746.
- Li,H.V. and Fournier,M.J. (1992) *EMBO J.*, **11**, 683–689.
- Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557.
- Rose,M.D., Winston,F. and Heiter,P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor University Press, Cold Spring Harbor, NY.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor, NY.
- Kohrer,K. and Domdey,H. (1991) *Methods Enzymol.*, **194**, 398–405.
- Balakin,A.G., Schneider,G.S., Corbett,M.S., Ni,J. and Fournier,M.J. (1993) *Nucleic Acids Res.*, **21**, 5391–5397.
- Samarsky,D.A., Balakin,A.G. and Fournier,M.J. (1995) *Nucleic Acids Res.*, **23**, 2548–2554.
- Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, USA 53711.
- Sikorski,R.S. and Hieter,P. (1989) *Genetics*, **122**, 19–27.
- Sollner-Webb,B. (1993) *Cell*, **75**, 403–405.
- Rymond,B.C. and Rosbash,M. (1992) In Jones,E.W., Pringle,J.R. and Broach,J.R. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces*. Vol. 2, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Hartshorne,T. and Agabian,N. (1994) *Nucleic Acids Res.* **22**, 3354–3364.
- Porter,G.L., Brennwald,P.J., Holm,K.A. and Wise,J.A. (1988) *Nucleic Acids Res.*, **16**, 10131–10152.
- Leader,D.J., Sanders,J.F., Waugh,R., Shaw,P. and Brown,J.W.S. (1994) *Nucleic Acids Res.*, **22**, 5196–5203.
- Berthomieu,R., Guerrier,D. and Giraudat,J. (unpublished) Z26461.