A structural analysis of *P.polycephalum* U1 RNA at the RNA and gene levels. Are there differentially expressed U1 RNA genes in *P.polycephalum* ? U1 RNA evolution

Evelyne Myslinski, F.-Xavier Wilhelm¹ and Christiane Branlant*

Laboratoire d'Enzymologie et de Génie Génétique, Université de Nancy 1, U.A. CNRS 457, BP 239 54506, Vandoeuvre les Nancy cedex and ¹Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Rene Descartes, 67084 Strasbourg cedex, France

Received October 26, 1988; Revised and Accepted January 5, 1989

EMBL accession nos X13841, X13842

ABSTRACT

U1 RNAs were isolated from *P.polycephalum* microplasmodia nuclei and sequenced. A *P.polycephalum* gene coding for U1 RNA was also isolated. The coding region of this gene differs at 3 positions compared to the isolated U1 RNA species. Both isolated RNAs and the gene encoded RNA can be folded according to the secondary structure model previously proposed for U1 RNA. Putative regulatory elements very similar to those required for efficient transcription of U RNA genes from vertebrates, in particular, the -200 distal enhancer element, are present in the flanking regions of this gene. The presence of several U1 RNA genes in *P.polycephalum* was confirmed by Southern blot analysis of genomic DNA. In contrast to yeast *S.cerevisiae* U1 RNA, *P.polycephalum* U1 RNAs have a length similar to that of U1 RNAs from higher eukaryotes. Nevertheless, *P.polycephalum* U1 RNAs probably differ from these RNAs in the 5'-terminal segment supposed to base-pair with the 5'-end of introns. The results are discussed taking into account phylogenetic evolution and functional role of U1 RNA.

INTRODUCTION

The metabolically stable nuclear U1, U2, U4, U5 and U6 RNAs have been found to play a role in mRNA splicing (for review see (1)).

The primary structure of U1 RNA has been determined for a large variety of higher eukaryotes, either by sequencing the RNA (Hela cells, rat, chicken (2) ; *Drosophila* (3)) or by sequencing the genes (*Xenopus laevis* (4), sea urchin (5) *Phaseolus vulgaris* (6)). A high degree of sequence conservation was observed, especially for the 5'-terminal region and for the protein binding domain (7). As yet, only one U1 RNA from a lower eukaryote has been sequenced, that of the yeast *S.cerevisiae* (8,9). This RNA strongly differs from that of higher eukaryotes.

Except for U6 RNA (10,11), the snRNA genes are transcribed by RNA polymerase II and are characterized by the absence of the classical TATA and CAAT boxes (12,13). Sequence comparison of the flanking regions of U1 (4,14-18) and U2 RNA genes (19-22) from higher eucaryotes has revealed the presence of conserved nucleotide sequences in vertebrate U RNA genes. These

sequences correspond to regions essential for correct transcription initiation or for 3'-end formation (for a review see 23). Two distinct evolutionarily conserved regions have been observed in the 5'-flanking region. A proximal region centered between nucleotides -50 and -60 from the 5'-end of the coding region, which appears to be functionally analogous to a TATA box and a distal region near position -200, which has the properties of an enhancer (23). A conserved sequence is essential for correct 3'-end formation, it is found 8-19 nucleotides downstream from the coding region (23). Except for Drosophila, (16,22), there is little information about transcription signals in U RNA genes from invertebrate species. Some U1 RNA genes should have additional regulatory elements since they were found to be subjected to developmental control : while mice U1a RNA is synthesized in all tissues, mice U1b RNA is only synthesized in fetal tissues and in a few adult tissues (24). Another example of developmental regulation was observed in X.laevis, where some U1 RNA genes are transcribed during early oogenesis, while others are transcribed during early embryogenesis (25). Furthermore, in addition to the true expressed bona fide U RNA genes, a large number of non expressed U RNA pseudogenes were isolated (26).

We report here the characterization of two U1 RNAs and one U1 RNA gene from the slime mold P.polycephalum. The results obtained are discussed in the light of our present knowledge of U1 RNA gene expression and of U1 RNA function and evolution.

MATERIALS AND METHODS

1. Isolation of P.polycephalum U1 RNA

Nuclei were isolated from P.polycephalum strain M3C VIII micro-plasmodia (27), following the procedure described by Mohberg and Rusch (28). Nuclear RNAs were isolated by two successive phenolic extractions of the nuclei at 4°C in 100mM NaCl, 10mM Tris-HCl buffer pH 7,6 as previously described (29) and were fractionated on a 15-30 % sucrose gradient in the same buffer. U RNAs were immunopurified from the 4S-8S RNA fraction using antibodies specific for m3G as described by Krol et al.,(30). They were then purified by electrophoresis on 15 % polyacrylamide gel, in the presence of 8M urea.

2. RNA sequencing

i) Chemical method : the chemical degradation method of Peattie (31) was applied to 3'-end labeled molecules. Labeling was achieved using (5',³²P) pCp and T4 RNA ligase (32). ii) Reverse transcriptase method : a

31mer oligonucleotide complementary to the central region of U1 RNA was chemically synthesized (Pharmacia DNA-synthesizer), 5'-end labeled and used as a primer for RNA sequencing with reverse transcriptase according to Hu Qu et al., (33).

3. Molecular cloning

The P.polycephalum genomic DNA library was constructed by Wilhelm and Wilhelm (34) by insertion of EcoRI fragments generated by complete digestion of the genomic DNA into the λ gt WES arms. Transfer of λ plaques to nitrocellulose filters, was done according to Maniatis et al., (35). Filter Hybridization using 3'-end labeled U1 snRNA or 5'-end labeled oligonucleotide was performed overnight at 37°C, in 50 % formamide, 5 x SSC, the filters were then washed 3 times in 2 x SSC at 37°C. Dot blot selection of M13mp9 single-stranded recombinant phages containing fragments of the gene was performed using the same hybridization conditions.

4. Southern blot analysis

Southern transfer of DNA to nitrocellulose was achieved according to Maniatis et al., (35). M13mp9 recombinant DNA was labeled by nick translation with (α - 32 P) dATP using the B.R.L. kit. Pre-hybridization and hybridization conditions were those of Wahl et al., (36) modified by Leary et al., (37), except that filters were washed several times with 5 x SSC at 60°C and that dextran sulfat was omitted during hybridization.

5. Recombinant DNA technology and DNA sequencing

The general procedure used for recombinant DNA technology were as described by Maniatis et al., (35). DNA segments to be sequenced were subcloned into phage M13mp9 and sequenced with the chain termination dideoxy method of Sanger et al., (38).

RESULTS

1. Isolation and sequence analysis of P.polycephalum U1 RNAs

Nuclei from P.polycephalum microplasmodia were prepared and the nuclear snRNAs were isolated as described in Materials and Methods. They were then 3'-end labeled and fractionated by electrophoresis on a 15 % polyacrylamide gel, in parallel with rat U snRNAs. A clear pattern of discrete species was obtained (Fig.1), but due to differences in electrophoretic mobilities, it was not possible to assign the various P.polycephalum RNA species to their rat counterparts. This was made possible by 3'-terminal sequence determination and comparison with the nucleotide sequences of higher eukaryote U1 RNAs. RNA from all the bands

Figure 1 : Electrophoresis fractionation of *P.polycephalum* and rat U snRNAs on 15 % polyacrylamide gel in the presence of 8M urea. *P.polycephalum* U snRNAs (lane a) and rat U snRNAs (lane b) were obtained after affinity chromatography of nuclear RNAs with anti 2,2,7 trimethylguanosine antibodies, as described in Methods. They were 3'-end labeled using (5'-³²P) pCp and T4 RNA ligase. The material contained in all the bands which have been numbered in this Figure had been subjected to 3'-end nucleotide sequence determination. Bands 3 and 4 were found to contain U1-1 and U1-2 RNA species, respectively.

numbered in Figure 1 were analyzed, only bands denoted 3 and 4 were found to correspond to U1 RNA and the corresponding RNAs were denoted U1-1 and U1-2 RNAs. Their nucleotide sequences were analysed by combining the results of two approaches : i) determination of the structure of the 3'-half of the molecule using the chemical RNA sequencing method (31) applied to the 3'-end labeled molecule, ii) determination of the structure of the 5'-half of the molecule using both the chemical method and the reverse transcriptase method developed by Hu Qu et al., (33). The oligonucleotide used as a primer is indicated in Figure 2.

The sequence at the extreme 5'-end of the molecule was difficult to solve. Reverse transcriptase stopped completely after an adenine residue. In all other U1 RNAs, this adenine residue is located at position 7 as referred to the 5'-end of the molecule. Since this adenine residue is always preceded by pyrimidine residue, one possible explanation for the

```

1  CCTTTTATTT ATATTAAAGC AGGGTATGCA CATTGGTATC GATTGTTTGC ATTTAATCTT
61  TGCAGTGGCG GTAATCTTTG CCATTGATGT TTGGTAAGTT CAAATTTAAT TGTGGCTTAA
121 TAAAAATATG TATGTAAAG ACAAGTATGT ACGGGCTGTG CGCGGGGGTT GGGCATTACA
181 CAATGGCATC CAAACAAGCA TAATCTTACC AAGCAAAAT GTTGCAATCC ACAAGATCAG
241 CCACTCATAA CCTGCAACAT CCATCTCTAG CGGGGATTGA ATTATCTCTT TTGTGAACC
301 TTCCCGTAGC CCCGTCCAGT TTTTGCCACA TTCGCCA ATC CCACAAGCTC ATCTTCCACC
361 CAAGCACATA AAAACGCAGG GCGCGCAGAC AAAGCATACA TACCTGGCCG GGGGAAGTCG
421 GTGATCAAGA CGGCCGAGTC CCTAGTTGGG GTACCTCTCC TGCATTGAG AGGGGCCCTT
      A                                     U
481 TCTATGATTG CCTTCGGGTA ACTCAACGGC ATAATTGTG ATAGCGTGGG GTGCGCTTGT
      ~~~~~
541 CGCGCCCTGC ATCCCAATA GTTTCAAAG AATATAATC ACTCAACCT ATTGCATATA
      (U)
601 ATTCAACAT CCATCTTCAA TAATACAACA TTAATGTATT TATAAATCC CATTCTCTCT
661 AGCATGTTAC AATAAACGCC TATATTACAG CCTCTTCCTC AATAACTCTT CTGGCATTTT
721 AGCGCCACTC GGCCTTAACG TGAGAATTC

```

Figure 2 : Sequence of the isolated U1 RNA gene compared to those of the two isolated U1-1 and U1-2 RNAs. The sequence is that determined for the DNA segment. The DNA region corresponding to U1 RNA is indicated in heavy characters. Only the differences in the RNA sequences compared to the DNA sequence are indicated. The two nucleotides underlined with a broken line are present in the RNA but their identify was only determined in the DNA. The wavy line represents the binding site on U1 RNA of the synthetic oligonucleotide used for RNA sequencing with reverse transcriptase. The boxed regions in the flanking non-coding regions of the gene indicate potential transcriptional regulatory signals as this is discussed in the text.

reverse transcriptase arrest could be a strong cleavage of the Py-Ap phosphodiester bond during U1 RNA isolation. A second possible explanation could be the presence of a modified nucleotide at position 6 in P.polycephalum U1 RNA that would stop reverse transcriptase action. Although, the nucleotide sequence of the extreme 5'-end of the molecule was also difficult to determine using the chemical method applied to the 3'-end labeled molecule, the sequencing gel obtained strongly suggested that P.polycephalum U1 RNAs have an unusually shorter 5'-terminus. Reverse transcriptase may had stopped at the level of the cap 2'-O methylated nucleotides.

Therefore, combination of the two sequencing approaches allowed us to determine the primary structures of U1-1 and U1-2 RNAs, except for their extreme 5'-ends (Fig. 2). The sequence of U1-1 and U1-2 RNAs are almost identical and only differ by an additional uridylic acid at the 3'-end of U1-1 RNA. In order to complete the nucleotide sequence determination we decided to clone a U1 RNA gene.

Figure 3 : Representation of the various DNA fragments which have been used for U1 RNA gene sequence determination. The 7kb fragment isolated from the recombinant phage was digested by Sau3A and AluI nucleases and the resulting DNA fragments were inserted in M13mp9 DNA. A recurrent approach was used to select among the various phages obtained those having inserted segments corresponding to the coding region or flanking region. Five phages were selected in this way, the fragments they contained were subjected to sequence analysis. Arrows indicate the DNA strand which has been sequenced. The EcoRI 7Kb fragment is also represented on this Figure, with the U1 coding region in full line and the orientation of transcription. Fragment 5 was subcloned in pUC19.

2. Isolation and sequencing of a *P.polycephalum* gene for U1 RNA

A genomic library which had been constructed by total digestion of the genomic DNA with EcoRI and insertion into phage λ (34) was screened using the 3'-end labeled U1 RNA as a probe. Of the 80 000 plaques screened 3 hybridized to the probe, and were plaque purified. The three phages were all found to contain a 7Kb EcoRI fragment of genomic DNA and gave the same pattern of digestion with BamHI, and HindIII enzymes, suggesting that they carried identical genomic fragments. The 7Kb EcoRI fragment of one of the recombinant phages was fractionated on agarose gel, and subjected to partial digestion with Sau3A nuclease and total digestion with AluI nuclease. The mixture of Sau3A or AluI fragments was used for ligation in M13mp9 DNA cleaved with BamHI or HindII nuclease, respectively. JM103 competent cells were transformed and the phages containing U1 RNA complementary sequences were selected by hybridization with the synthetic oligonucleotide probe represented in Figure 2. Phages isolated in this way were then used to select by successive hybridization steps a series of recombinant phages containing fragments from the coding and flanking regions of the U1 RNA gene (Fig. 3). Nucleotide sequence of the fragments inserted in these phages was determined using the dideoxy-chain-termination method of Sanger et al., (38). The primer used was either the commercial AMERSHAM primer or the oligonucleotide used for RNA

sequencing. In this way, as shown in Figure 3, most of the sequence, of the coding region and of the 3' and 5' flanking regions were determined on both strands.

3. Comparison of the nucleotide sequences and predicted secondary structures of the isolated U1 RNAs and of the RNA encoded by the isolated gene

As shown in Figure 2, a few differences are observed between the nucleotide sequences of U1-1 and U1-2 RNAs and the DNA sequence. Two of these differences concern internal nucleotides : at position 78 as referred to the RNA putative 5'-end, a U residue present in both RNAs is replaced by a C residue in the DNA and at position 53 a A residue also present in both RNAs is replaced by a T residue in the DNA. The third difference concerns the extra U residue at the 3'-end of U1-1 RNA, a C residue is observed at this position in the DNA fragment. In otherwords, none of the two isolated RNAs is encoded by the isolated gene and P.polycephalum should contain at least two U1 RNA genes. This assumption was verified by Southern blot analysis of genomic DNA. The DNA was digested with various restriction enzymes (EcoRI, BamHI, HindIII), and fractionated on agarose gel. The fractionated products were transferred to nitrocellulose filter and hybridized with nick-translated fragment 5 (Fig 3). As can be seen on Figure 4, in addition to the 7 kb fragment, several other genomic EcoRI fragments hybridized to the probe confirming the presence of several U1 RNA genes in P.polycephalum genome. We would like to point out that hybridizations have been performed with fragment 5 containing both the coding region and 5' and 3' flanking regions required for expression. Therefore it might be that some of the weaker bands observed correspond to genes having regulatory elements similar to those of U1 RNA gene, for instance other U snRNA genes.

The secondary structure of rat U1 RNA has been previously studied by different approaches and a model verified by phylogenetic studies has been established for U1 RNA (2,3). As shown in Figure 5, this model fits well with the sequence determined for P.polycephalum U1-1 and U1-2 RNAs, as well as, with that of the RNA encoded by the gene. The base substitutions observed in the DNA do not disturb the potentiality for secondary structure formation of the encoded RNA. On the contrary, in hairpin II, a GU pair in the isolated RNA is replaced by a more stable GC pair in the RNA encoded by the gene.

Figure 4 : Genomic organisation of U1 RNA genes in *P.polycephalum*. The recombinant phage containing the 7 Kb EcoRI fragment was digested with EcoRI enzyme (lanes a, b and c) and *P.polycephalum* DNA (6µg) was digested with the restriction endonucleases indicated (lanes d, e, f and g). The digestion products were fractionated on a 1 % agarose gel, transferred to nitrocellulose and hybridized with the nick-translated fragment 5 (Fig.3). The amount of λ phage DNA loaded in lanes a, b and c have been chosen taking into account *P.polycephalum* genomic DNA molecular weight, the amount of DNA loaded in lanes d, e, f and g and λ recombinant molecular weight. In lanes a, b and c, amounts of the 7 Kb fragment correspond to 3 ; 1,5 and 0,3 copies in the genomic DNA, respectively.

4. Comparison of *P.polycephalum* U1 RNA with U1 RNAs from higher eukaryotes, U1 RNA evolution

Both the isolated RNAs and the RNA encoded by the gene display a high degree of sequence conservation compared to U1 RNAs from higher eukaryotes, for instance 58 % homology with rat or human RNA. This is very similar to the degree of homology found between *Tetrahymena pyriformis* and rat U5 RNAs: 57 % (39). On the basis of primary structure comparison, *P.polycephalum* U1 RNA has a slightly higher homology with bean U1 RNA (6) than with vertebrate U1 RNAs : 62 % against 58 %, respectively. The homology between mold and plant RNA is more pronounced when comparison is made at the level of the folded molecules. On the other hand, *P.polycephalum* and animal RNAs display a large block of homology in the stem of hairpin III (Fig.5).

As can be seen in Figure 5, nucleotides conserved in all U1 RNA species are concentrated in different areas of the molecule. The primary

Figure 5 : Secondary structure of the isolated *P.polycephalum* U1 RNAs : The structure was derived from that previously established from rat and *Drosophila* RNAs (2,3). The sequence is that of the isolated U1-1 RNA, nucleotide differences in the DNA sequence are indicated by circled nucleotides, N represent undetermined nucleotides. Invariant nucleotides in human, rat, chicken, *X.laavis*, sea urchin, *Drosophila*, common bean and *P.polycephalum* RNAs are boxed with full line, the structure boxed with dashed lines is common to all these RNA species except common bean. Hairpin structures and single-stranded sequences have been numbered according to 47.

structure conservation is essentially located in single-stranded regions rather than in double-stranded regions. Only the 3 base-pairs at the top of hairpin IV are conserved in all sequenced species, the bottom of the stem of hairpin III being highly conserved except in plant RNA. Whereas, as previously observed for higher eukaryotes (2,3), the primary structures of the top loops of hairpin I and II and of the single-stranded region denoted (i) are conserved.

4.1. Evolution of the U1 RNA 3'-domain (domain A) : Only two regions of the Sm binding domain A have primary structures highly conserved from slime mold to man : the block of three base-pairs at the top of hairpin IV and the internal single-stranded region containing sequence (i) (sequence (i) was shown to be essential for binding of Sm proteins (40,41)). Sequence

(g) in the top loop of hairpin III, previously defined on the basis of homologies between vertebrate U1, U2 and U4 RNAs (7), is poorly conserved in P.polycephalum U1 RNA. The segment (j) : Y-N-Y-Gp present in the top loop of hairpin IV of vertebrate U1, U2, U4 and U5 RNAs, had the sequence U-U-C-Gp in all U1 RNAs sequenced up to now, P.polycephalum U1 RNA slightly diverges at this level since its sequence is C-U-U-Gp.

4.2. Evolution of the U1 RNA 5'-domain : In the 5'-domain, the high degree of nucleotide sequence conservation of the single-stranded regions contrasts with the absence of nucleotide sequence conservation in the double stranded regions (Fig 5). The conservation of loops I and II reveals that they should be essential for U1 RNA function, and this is in agreement with the recent finding of Patton and Pederson (40) and Hamm et al., (41) of a binding site for the 70 K and A proteins in loop I. The role of the 5'-terminal single-stranded region in U1 RNA function is now confirmed by several experimental data (42-44). Surprisingly, P.polycephalum U1 RNA gene differs from all U1 RNAs in this region. All U1 RNAs sequenced have as 5'-terminal sequence : $m_3GpppAmUmAC^U/\psi^U/\psi ACCUGp$, in the P.polycephalum U1 RNA gene this sequence is replaced by ATACATACCTGp where the U residue normally observed at position 5 is replaced by an A residue. Furthermore, according to RNA sequencing this A residue may constitute the capping site of P.polycephalum U1 RNA.

5. Flanking regions of the P.polycephalum U1 RNA gene : Comparison with U1 RNA genes from vertebrates

5.1. 5'-flanking region : No characteristic TATA box is observed, but a putative -55 element is present. The homology observed between this element and the consensus sequence established for the vertebrate proximal element is not very strong (Fig 6). This is not very surprising since the degree of sequence conservation of this -55 element is not high in vertebrates (for review, see 23). The C-Cp dinucleotide universal in vertebrate proximal element, is present in the P.polycephalum sequence. The distal enhancer element located between positions -200 and -250 of the U RNA genes from vertebrates has on the contrary a very conserved primary structure (23). A sequence T-A-T-G-T-A-A-A-Cp found between nucleotides -257 and -265 in the P.polycephalum U1 RNA gene fits perfectly with the consensus sequence Y-A-T-G-Y-A-R-A-Tp established for vertebrates (23).

5.2. 3'-flanking region. Genes transcribed by RNA polymerase II, bear a characteristic sequence in their 3' non-coding region (T-T-T-N-A-A-A-G-A/T-Ap), which is found at variable positions from the

Figure 6 : Comparison of P.polycephalum U1 RNA (pU1) gene putative transcriptional regulatory elements with the consensus sequences of vertebrate U1 RNA gene regulatory elements. The P.polycephalum sequences have been aligned with : the consensus sequences for (A) the proximal sequence element of vertebrate U1 RNA genes, (B) the distal sequence element of vertebrate U1 RNA genes, (C) the 3' non coding element of vertebrate U1 RNA genes and of polymerase II transcripts (21). The consensus sequences were established using nucleotide sequences gathered by Dahlberg and Lund in (23). The size of capital letters represents frequencies of appearance of nucleotides in the various U1 RNA gene sequences taken into consideration, large capital letters : frequencies above 92 %, small capital letters : frequencies between 50 % and 92 %.

A-A-T-A-A-Ap sequence (21). Such a sequence was also found in the 3' non-coding region of U RNA genes very close to the 3'-end of the coding region (23). As can be seen in Figure 6, a similar sequence is observed at position + 10 in the P.polycephalum U1 RNA gene. Interestingly, there is a better fit of the P.polycephalum sequence with the consensus sequence found for RNA polymerase II transcripts in general, than with that obtained by restricting the comparison to U RNA genes from higher eukaryotes.

In conclusion, both 5' and 3' flanking regions from the isolated P.polycephalum gene bear characteristic structures homologous to those required for efficient transcription of U RNA genes of vertebrates. U1 RNA complementary sequences have been isolated for two other invertebrates : Drosophila (16) and sea urchin (45), comparison of the 5' and 3' flanking sequences with those of vertebrate U RNAs reveals little, if any, homology. But there is no evidence that these sequences correspond to true genes.

DISCUSSION

We have isolated and sequenced two U1 RNA species from microplasmodia of P.polycephalum. They were found to differ by one additional nucleotide at the 3'-end of the largest one. Such 3'-terminal heterogeneity due to imprecision in the maturation process, was frequently observed in U4 and U5 RNAs from higher eukaryotes (46,47) and was also observed in U1 RNA from Drosophila (48).

We have also cloned and sequenced a P.polycephalum U1 RNA gene. The sequence of this gene differs at 3 positions from those of the isolated RNAs. Since no U1 RNA corresponding to this gene could be isolated from microplasmodia nuclei. Several possible explanations for this result must be discussed : i) differences observed in the RNA reflect post-transcriptional modifications ii) the isolated gene is expressed at a very low level and its transcription product could not be detected, iii) we have isolated a pseudogene, iv) this gene is not expressed in microplasmodia but is expressed at another stage of P.polycephalum life cycle.

Concerning the first hypothesis, misinterpretation of RNA sequencing gels due to the presence of base-modifications cannot be completely excluded. Nevertheless, the positions concerned were never described as post-transcriptionally modified in U1 RNA. Concerning the second hypothesis, we have used the 3'-terminal labeling method to analyze the U RNA. This method is very sensitive and should allow detection of very low amount of RNA. Therefore if the isolated gene is transcribed in microplasmodia this should be in very low amount.

The third hypothesis must be considered in view of our sequence data : the nucleotide sequence differences observed in the gene affect neither the RNA secondary structure nor the characteristic features of single-stranded region. It even increases the stability of hairpin II. Furthermore, the flanking 5' and 3' regions of the gene display structural features homologous to those required for efficient transcription of U RNA genes from vertebrates. We examined a series of published sequences of U RNA pseudogenes from vertebrates belonging to the class I of pseudogene defined by Denison and Weiner (26), none of the examined sequence was found to display all the characteristics we observed for the P.polycephalum gene, either the -250 element was absent or the coding region was severely mutated without conservation of the secondary structure of the putative transcript. These observations suggest that the isolated P.polycephalum U1

RNA gene may be a bona fide gene coding for a variant RNA.

Coexpression of several variants in a given tissue is particularly frequent for U5 RNA (39), in the case of U1 RNA the expression of variants was found to be developmentally regulated (24,25). The possibility that the isolated gene is expressed differentially during P.polycephalum life cycle should be considered in connection with the fact that a developmental switch occurs during P.polycephalum life cycle, when amoeba are converted to plasmodia (49). There are characteristic morphological and cytological differences between amoebae and plasmodia. These differences are associated with changes in gene expression as evidenced by a comparison of the proteins of the two cell types (49). Maturation of messenger RNA specific for amoebae state may require a specific U1 RNA. Furthermore, we observed the presence of a peculiar structure at the level of the putative enhancer region of the P.polycephalum gene : the sequence T-A-T-Gp occurs three times, twice on the 5'-side of the enhancer, once on its 3'-side. A similar feature with a repetition of an alternated pyrimidine-purine tetranucleotide sequence was observed within a SV40 enhancer element whose sequence fits the consensus sequence of the U RNA enhancer (Fig. 7) (50), as well as in the gene for X.laevis Ulb RNA whose expression is developmentally regulated : a series of 4 repeats of the sequence T-G-C-Ap is located upstream of the beginning of the enhancer (4) (Fig. 7). In these three cases the repeated sequence creates a stretch of alternating purine-pyrimidine nucleotides which may form Z-DNA (51). Observation of such a peculiar structure at the level of the enhancer of the P.polycephalum U1 RNA gene may reflect a complex regulation of the expression of this gene.

Attempts to express this gene by microinjection of a recombinant pUC19 plasmid containing fragment 5 (Fig.3) in X.laevis oocytes have been unsuccessful. This is not an argument against the fact that the gene could be a true gene. Indeed, there is no information about the possibility to express P.polycephalum genes in X.laevis, secondly if this gene is developmentally regulated its expression may depend upon the presence of protein factors absent in X.laevis. Altogether these observations suggest that we probably have isolated a true U1 RNA gene not expressed in micro-plasmodia but expressed at another stage of P.polycephalum life cycle.

In this gene, the U residue found at position 5 in the segment of U1 RNA supposed to base-pair with the 5'-extremity of introns, is replaced by

Figure 7 : Comparison of the distal enhancer element of P.polycephalum U1 RNA gene with that of X.laevis Ulb RNA gene (4) and a SV40 enhancer (50). Repeated segments with alternated purine pyrimidine sequences are indicated by arrows.

a A residue. Only a limited number P.polycephalum genes have been sequenced (52-54). A strong conservation of the dinucleotide GT at the 5' end of introns, is observed. The downstream region of introns that should base-pair with the U1 RNA region which is mutated in the P.polycephalum gene displays a high degree of sequence variation. Actually, this is also the case for introns of other species. The difference observed in the P.polycephalum gene could be the result of a special role of the encoded RNA. Another alternative is that all P.polycephalum U1 RNA genes have an adenine residue at position 5, creating a new potential site for RNA capping and explaining the apparent shorter size of the isolated RNAs. Additional experiments are presently carried out to clarify this important point.

CONCLUSION

The present results open very interesting questions and it will now be interesting to clone and sequence other P.polycephalum U1 RNA genes and to study the expression of these different genes at different stages of the life cycle. On the other hand, the present results demonstrate that U1 RNA which is an essential element of the splicing machinery has been highly conserved from lower eukaryotes to man. This observation and the fact that the classical GT-AG intron boundaries are also conserved in P.polycephalum genes (52-54) strongly suggest that splicing machinery is very ancient and has been strongly conserved in the course of eukaryotic evolution. The yeast S.cerevisiae which has a completely different U1 RNA is probably a special case among lower eukaryotes.

Finally the present results bring another important information concerning U RNA genes. Sequence conservation are observed at the level of coding region but also at the level of the peculiar regulatory elements of these genes.

ACKNOWLEDGMENTS

We thank Dr. R. Lührmann for its generous gift of anti 2,2,7 trimethylguanosine antibodies and Dr. P. Remy for micro-injecting the P.polycephalum gene into X.laevis oocyte nuclei. We are grateful to Dr. P. CARBON for helpful discussion. We thank M.F. Demouveau (Strasbourg) and J. BAYEUL (Nancy) for excellent technical assistance.

*To whom correspondence should be addressed

REFERENCES

1. Maniatis, T. and Reed, R. (1987) *Nature* **325**, 673-678.
 2. Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. and Jeanteur, P. (1980) *Nucl. Acids Res.* **8**, 4143-4254.
 3. Mount, S.M. and Steitz, J.A. (1981) *Nucl. Acids Res.* **9**, 6351-6368.
 4. Krol, A., Lund, E. and Dahlberg, J.E. (1985) *EMBO J.* **4**, 1529-1535.
 5. Brown, D.T., Morris, G.F., Chodchoy, N., Spreche, C. and Marzluff, W.F. (1985) *Nucl. Acids Res.* **13**, 537-556.
 6. Van Santen, V.L. and Spritz, R.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9094-9098.
 7. Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. and Jacob, M. (1982) *EMBO J.* **1**, 1259-1265.
 8. Kretzner, L., Rymond, B.C. and Rosbash, M. (1987) *Cell* **50**, 593-602.
 9. Siliciano, P.G., Jones, M.H. and Guthrie, C. (1987) *Science* **237**, 1484-1487.
 10. Kunkel, G., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8575-8579.
 11. Carbon, P., Murgo, S., Ebel, J.P., Krol, A., Tebb, G. and Mattaj, I. (1987) *Cell* **51**, 71-79.
 12. Chandrasekharappa, S.C., Smith, J.H. and Eliceiri, G.L. (1983) *J. Cell. Phys.* **117**, 169-174.
 13. Hoffman, M.L., Korf, G.M., Mc Namara, K.J. and Stumph, W.E. (1986) *Mol. Cell. Biol.* **6**, 3910-3919.
 14. Zeller, R., Carri, M.T., Mattaj, I.W. and De Robertis, E.M. (1984) *EMBO J.* **3**, 1075-1081.
 15. Manser, T. and Gesteland, R.F. (1982) *Cell* **29**, 257-264.
 16. Alonso, A., Jorcano, J.L., Beck, E. and Hovemann, B. (1984) *J. Mol. Biol.* **180**, 825-836.
 17. Watanabe, N., Itok, X., Tani, T., Okana, K., Koga, N., Okada, N. and Ohshima, Y. (1983) *Proc. Natl. Acad. Sci. USA* **81**, 3811-3815.
 18. Roop, D.R., Kristo, P., Stumph, W.E., Tsai, M.J. and O'Malley, B.W. (1981) *Cell* **23**, 671-680.
 19. Westin, G., Monstein, H.J., Zabielski, J., Philipson, L. and Pettersson, U. (1981) *Nucl. Acids Res.* **9**, 6323-6338.
 20. Tani, T., Watanabe-Nagasu, N., Okada, N. and Ohshima, Y. (1983) *J. Mol. Biol.* **168**, 579-594.
 21. Mattaj, I.W. and Zeller, R. (1983) *EMBO J.* **2**, 1883-1891.
 22. Alonso, A., Jorcano, J.L., Beck, E. and Spiess, E. (1983) *J. Mol. Biol.* **169**, 691-705.
-

23. Dahlberg, J.E. and Lund, E. (1988) in structure and function of major and minor small nuclear ribonucleoproteins particles, Birnstiel, M.L. Ed., Springer Verlag, Berlin.
24. Howard, E.F., Michael, S.K., Dahlberg, J.E. and Lund, E. (1986) Nucl. Acids Res. **14**, 9811-8925.
25. Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984) Cell **38**, 681-689.
26. Denison, R.A. and Weiner, A.M. (1982) Mol. Cell. Biol. **2**, 815-828.
27. Daniel, J.W. and Rusch, H.P. (1961) J. Gen. Microbiol. **25**, 47-59.
28. Mohberg, J. and Rusch, H.P. (1971) Exptl. Cell. Res. **66**, 305-316.
29. Stevenin, J., Gallinaro, H., Gattoni, R. and Jacob, M. (1977) Eur. J. Biochem. **74**, 589-602.
30. Krol, A., Ebel, J.P., Rinke, J. and Lührmann, R. (1983) Nucl. Acids Res. **11**, 8583-8594.
31. Peattie, D.A. (1979) Proc. Natl. Acad. Sci. USA **76**, 1760-1764.
32. England, E. and Uhlenbeck, O.C. (1978) Nature **275**, 560-561.
33. Hu Qu, L., Michot, B. and Bachellerie, J.P. (1983) Nucl. Acids Res. **11**, 5903-5920.
34. Wilhelm, M.L. and Wilhelm, F.X. (1984) FEBS Lett. **168**, 249-255.
35. Maniatis, T., Fritsch, E. and Sambrook, J. (1982). Editors of Molecular Cloning. A laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
36. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA **76**, 3683-3687.
37. Leary, J.J., Brigati, D.J. and Ward, D.C. (1983) Proc. Natl. Acad. Sci. USA **88**, 4045-4049.
38. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Eur. J. Biochem. **74**, 589-602.
39. Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Dias, L. and Pousada, L. (1983) Nucl. Acids Res. **11**, 8359-8367.
40. Patton, J.R. and Pederson, T. (1988) Proc. Natl. Acad. Sci. USA **85**, 747-751.
41. Hamm, J., Kazmaier, M. and Mattaj, I.W. (1987) EMBO J. **6**, 3479-3485.
42. Mount, S.M., Pettersson, I., Hinterberger, M., Karvas, A. and Steitz, J.A. (1983) Cell **33**, 509-518.
43. Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) Cell **38**, 299-307.
44. Zhuang, Y. and Weiner, A.M. (1986) Cell **46**, 827-835.
45. Yu, J.C., Nash, M.A., Santiago, C., Marzluff, W.F. (1986) Nucl. Acids Res., **14**, 9977-9988.
46. Krol, A., Branlant, C., Lazar, E., Gallinaro, H. and Jacob, M. (1981) Nucl. Acids Res. **9**, 2699-2716.
47. Krol, A., Gallinaro, H., Lazar, E., Jacob, M. and Branlant, C. (1981) Nucl. Acids Res. **9**, 769-787.
48. Myslinski, E., Branlant, C., Wieben, E. and Pederson, T. (1984) J. Mol. Biol. **180**, 927-945.
49. Blindt, A.B., Chainey, A.M., Dee, J. and Gull, K. (1986) Protoplasma **132**, 149-159.
50. Benoist, C. and Chambon, P. (1981) Nature **295**, 622-624.
51. Nordheim, A. and Rich, A. (1983) Nature **303**, 674-679.
52. Nader, W.F., Isenberg, G. and Sauer, H.W. (1986) Gene **48**, 133-144.
53. Monteiro, M.J. and Cox, R.A. (1987) J. Mol. Biol. **193**, 427-438.
54. Wilhelm, M.L. and Wilhelm, F.X. (1987) Nucl. Acids Res. **15**, 5478