Volume 10 Number 7 1982 Nucleic Acids Research

Nucleotide sequence of an external transcribed spacer in *Xenopus laevis* rDNA: sequences flanking the 5' and 3' ends of 18S rRNA are non-complementary

B.Edward H.Maden, Michael Moss and Mohammad Salim

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK

Received 15 February 1982; Accepted 17 March 1982

ABSTRACT

We have sequenced the external transcribed spacer (ETS) of a ribosomal transcription unit from Xenopus laevis, together with sections of the preceding non-transcribed spacer. Our analysis was carried out on the same cloned transcription unit as that from which the internal transcribed spacers (ITS) were previously sequenced. The ETS is approximately 712 nucleotides long and, like the ITS regions, is generally very rich in C plus G. Features of the sequence include an excess of oligo-C tracts over oligo-G tracts and a tract of 37 nucleotides consisting almost entirely of G and A residues. Parts of the sequence can give rise to stable internal secondary structures. However, in contrast to Escherichia coli, there is no potential for major base-pairing between the 18S flanking regions of the ETS and ITS. Further findings are that there are no initiation (ATG) codons in the ETS and that, as in other X.laevis rDNA cloned units, the sequence preceding the ETS is duplicated, with a few changes, in the "Bam island" sequence of the non-transcribed spacer.

INTRODUCTION

Xenopus laevis has provided a model system for studying the organisation and transcription products of ribosomal DNA (rDNA) in higher organisms. Nucleotide sequence data are now available covering much of the repeating unit structure of rDNA including the sites of initiation and termination of transcription (1), much of a non-transcribed spacer (NTS) and much of the adjoining external transcribed spacer (ETS) (2), a complete 18S gene (3) with adjacent short section of ETS (4), the 18S-28S intergene region (5) and parts of 28S genes (1,2,5,6).

However, data on the ETS are incomplete in two respects. First, there is a gap of unknown length between the end of the partial ETS sequence reported in (2) and the start of the ETS segment flanking the 18S 5' region reported in (4). Bridging this gap in the sequence data may be important for understanding the processing of ribosomal precursor RNA (pre rRNA). For example, in Escherichia coli the

0305-1048/82/1007-2387$2.00/0

sequences flanking the small subunit rRNA (16S rRNA) interact to form an extensive base-paired structure which constitutes a processing site for RNase III (7). Knowledge of the ETS sequence in X.laevis is prerequisite for searching for the presence of any comparable processing site in X.laevis pre rRNA. Secondly, the previously reported ETS sequence data (2,4) were derived from two different rDNA clones. Because of possible sequence heterogeneties between different rDNA transcription units the complete sequence of a single (cloned) ETS would be more generally useful than a composite sequence derived from different clones. Furthermore, the question of possible interaction between the ETS and ITS during pre rRNA processing (7) can be most confidently approached if all of the respective sequence data are derived from the same transcription unit and thus the sequences are transcribed into the same pre rRNA molecule. With these considerations in mind we have sequenced the ETS in the X.laevis rDNA clone pXIr101, from which our analysis of the 18S-28S intergene region (5) and the 18S gene (3) were previously carried out. This paper reports our findings.

METHODS

The X.laevis rDNA clone pXIr101 contains a complete ribosomal transcription unit with its preceding NTS, and was a gift from R.Reeder. For the present work we used the subclone pXIr101A (ref.4) which contains the complete ETS with part of the NTS and most of the 18S gene. The sequence was determined by the method of Maxam and Gilbert (8) with 5' labelling of restriction fragments (figure 1) followed by secondary restriction. The A plus G reaction was carried out with pyridine formate, pH2, not piperidine formate as stated in (8). Other base-specific reactions were for G, T plus C and C (8). Sequencing gels were 40cm x 20cm x 0.4mm containing 8% or 6% polyacrylamide and 7M urea. 6% gels gave better resolution than 8% gels for fragments in the size range 120-190 nucleotides. Gels were run at 25 ma (approximately 2KV) after first prerunning until warm to minimise "compression effects". Nevertheless, many such effects were encountered, as mentioned in the following section.

RESULTS AND DISCUSSION

Sequence determination

The ETS and neighbouring sequences in pXIr101 are shown in figure

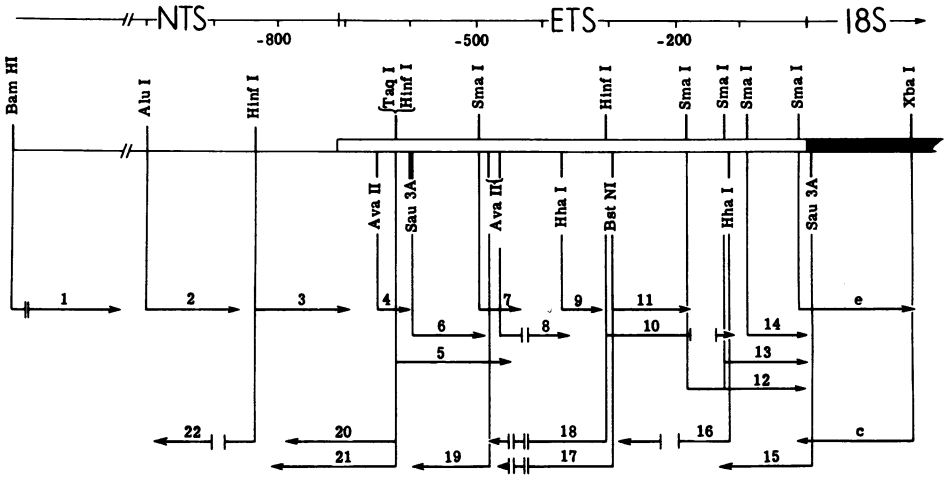

Figure 1 Summary of sequencing gels from which the sequence in figure 2 was established. Restriction fragments were obtained from the plasmid pXIr101A (ref. 4) after preliminary excision of the indicated Bam HI/Xba fragment. Only those restriction sites that were actually used for 5' labelling are shown in the figure; complete restriction data can be deduced from the sequence (figure 2). Note that the Alu repetitious regions to the right of the Bam HI site in the NTS (2) is omitted from this figure, as indicated by the interruption to the left of the Alu site near nucleotide -1000. (Distances from the start of the 18S coding region are shown on the top line). Sequencing gels are numbered sequentially for reference. (Gels e and c were from previous work (4)). Gels 12-14 were from fragments labelled after a partial Sma digest. Most of the sequence was covered on both strands except the "Bam island" which was only sequenced rightwards. The majority of gels were run on more than one occasion. Many gels showed unreadable sections at characteristic sites due to "compression effects". Where these extended over a substantial region or were otherwise troublesome a gap in the respective arrow is shown. For further discussion of this problem see the text and figure 3.

2. Nucleotides are numbered leftwards and negatively from the start of the 18S gene. This numbering scheme is suitable for discussing 18S rRNA maturation and is applicable for this purpose to systems where complete ETS sequence data may not be available.

Many spacing irregularities ("compression effects") were encountered in the sequencing gels and some of these were quite extensive. Nevertheless, by combining data from all the gels (figure 1) we derived an unambiguous sequence throughout the ETS except at one point. We were unable to determine accurately the number of consecutive G residues leftwards from G-419 but we estimated this number to be approximately six. (See the legend to figure 3). The explanation of several of

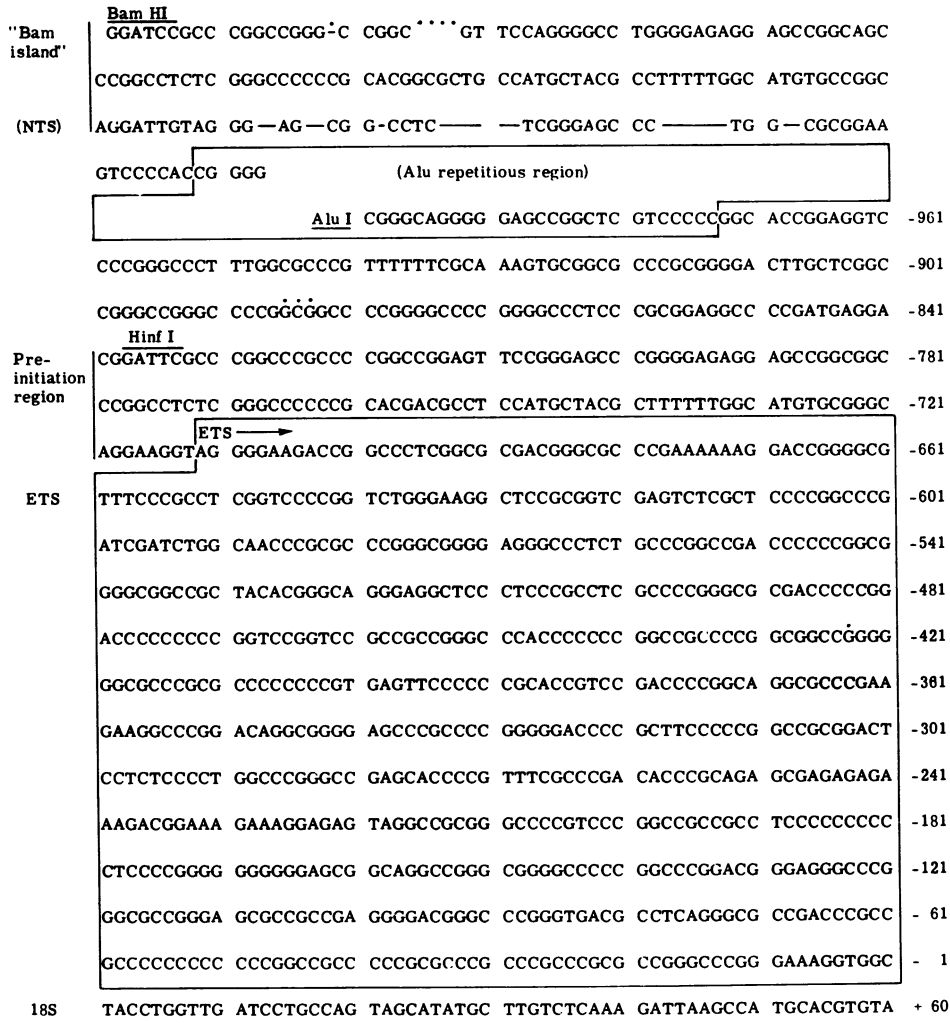

Figure 2 Sequence of the ETS in pXIr101A, with the start of the 18S coding sequence and parts of the NTS. Heavy dots represent uncertain nucleotides. Features of the sequence are described in the text.

the compression effects was revealed from the completed sequence as potential for DNA fragments to form local secondary structure in the sequencing gels. The two largest potential secondary structures are shown in figure 3. These structures would be highly stable in single stranded nucleic acid and would therefore be expected to form also in pre rRNA.

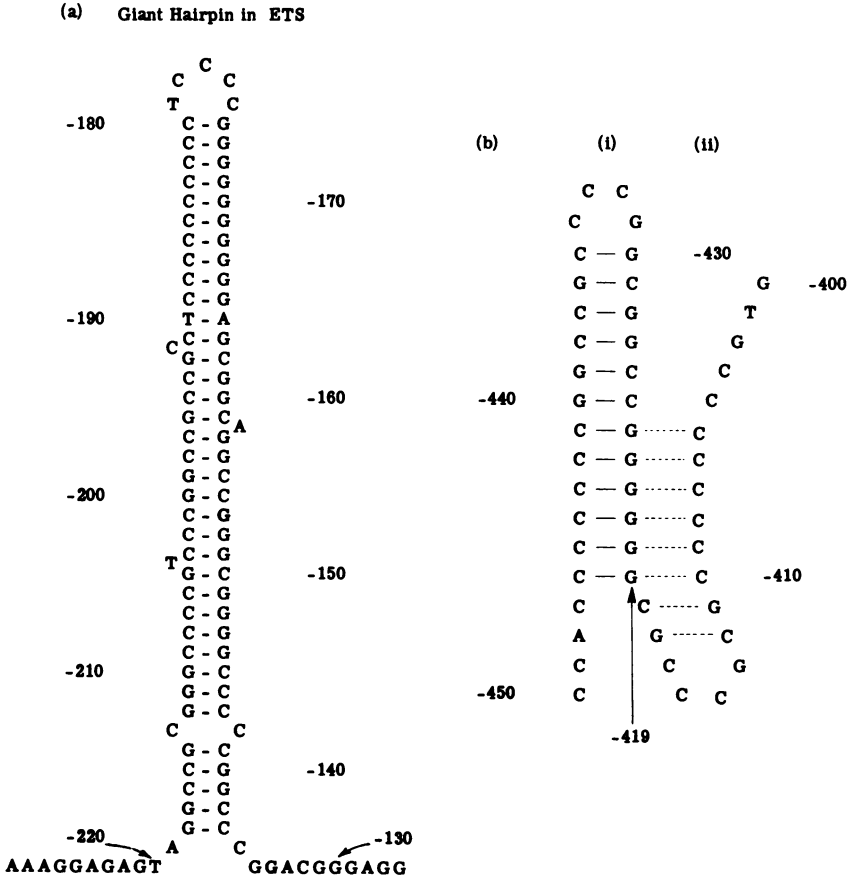

Figure 3 (a) Potential secondary structure of the rightwards (RNA-like) strand between nucleotides -220 and -135. Gels 10 and 16 (figure 1) showed substantial unreadable sections corresponding to the descending limb of the sequence shown, and its complementary sequence, respectively. The sequence was unambiguously established through this region from the combined data from gels 10, 12 and 16. (b) Alternative potential secondary structures in the region -450 to -405. In gels of the rightwards strand nucleotides -430 to -419 were completely obscured in a single intense band, consistent with a very stable structure of type (i). The sequence was deduced from the leftwards strand, except that the number of C residues corresponding to the indicated G tract at nucleotides -419 could not be counted exactly due to another secondary structure interaction of type (ii) (involving the complementary sequence to that shown). In one gel sufficient denaturation was obtained to indicate that this tract contains approximately 6 C residues on the complementary strand, with probable uncertainty limits of ±1. Moss et al (2) reported 6 G residues in this region of clone pXI108, but found uncertainty in neighbouring nucleotides (figure 4 of ref. 2). We have assumed 6 G residues here; the general form of the structure shown is not critically dependent upon the precise length of the G tract.

Features of the ETS sequence

The ETS is approximately 712 nucleotides long from the site of initiation of transcription (1) to the start of the 18S gene, with a very minor uncertainty in length near nucleotide -420 as discussed above. The sequence is generally very rich in C plus G, in which it resembles the internal transcribed spacers (table 1). There is a considerable excess of C over G and of A over T in the sequence. There are several tracts consisting of a single type of nucleotide, or of purines only or pyrimidines only (table 2). C-rich tracts outnumber G-rich tracts. Some of the above tracts contribute to the major secondary structure features already mentioned (figure 3) but others occur in regions without obvious potential for local secondary structure, as listed in table 2. In particular the C-rich region near the 18S gene (-70 to -30) and the A-rich tract from -253 to -217 lack complementary features within the ETS. There is no major repetitious motif to the ETS, in contrast to the NTS (2) (see below) although there are some local features where sequences of a few nucleotides are repeated two or three times. Curiously, two of these give rise to paired restriction sites (table 2).

Sequences flanking the 5' and 3' ends of 18S rRNA are non-complementary

As already mentioned, in the ribosomal transcription unit of E.coli the sequences flanking 16S rRNA interact to form an extensive base-paired structure which constitutes a site for cleavage by RNase III during ribosome maturation. Does 18S rRNA maturation in X.laevis involve a comparable interaction ?

The sites of pre rRNA cleavage during 18S rRNA maturation in X.laevis

Table 1 Base composition of ETS compared with ITS regions and 18S and 5.8S genes. Base composition data are expressed as percentage values, with numbers of nucleotides in parentheses for the ETS. ITS and 5.8S values are from ref. 5 and 18S values from ref. 3. All values are from the transcription unit of pXIr101.

	Length	T	A	G	C	G + C
ETS	712	6.2 (44)	10.7 (76)	37.2 (265)	45.9 (327)	83
ITS 1	557	3.5	12.5	41	43	84
ITS 2	262	7	5	35	53	88
18S	1825	22.5	23.7	28.2	25.6	54
5.8S	162	21.5	18.5	30	30	60

Table 2 Some distinctive tracts in the ETS of pXIr101. The table shows homopolymeric tracts of 6 or more residues, purine or pyrimidine tracts of 10 or more residues and two duplicated sites for restriction endonucleases (Sau 3A and Ava II respectively). The right hand column shows which of the purine and pyrimidine tracts occur in regions with clear potential for local secondary structure.

Nucleotides	Sequence	Potential for local secondary structure
Purine tracts		
-678 to -669	GAAAAAAGGA	No
-424 -419	GGGGGG(±1)	Yes (fig. 3b)
-253 -217	(A18,G16,C2,T)	No
-174 -163	GGGGGGGGGAG	Yes (fig. 3a)
Pyrimidine tracts		
-550 to -545	CCCCCC	No
-479 -471	CCCCCCCCC	No
-451 -440	CCCACCCCCCC	Yes (fig. 3b)
-410 -403	CCCCCCCC	Yes (fig. 3b)
-397 -390	TTCCCCCC	No
-302 -291	CTCCTCTCCCCT	No
-192 -175	(C2)T(C10)T(C4)	Yes (fig. 3a)
- 59 - 47	CCCCCCCCCCCC	No
Duplicated restriction sites		
-601 to -594	GATCGATC	
-470 to -461	GGTCCGGTCC	

have not been identified at the nucleotide sequence level. However, electron microscopic data (9) indicate that these sites are at or near the boundaries of the 18S sequence. Electron micrographs of 40S pre rRNA, the primary transcript, show a series of secondary structure features which can be identified with the ETS, ITS and 28S sequences. (See ref. 9 for details). The 18S sequence is identifiable by lack of secondary structure in the micrographs. The first intermediate in rRNA maturation, 38S RNA, differs from 40S RNA by loss of a segment which contains all of the secondary structure features of the ETS. The next intermediate, 34S RNA, lacks the 18S region but retains all the secondary structure features of the ITS. The length difference between 38S and 34S RNA is equal within experimental error to the length of 18S RNA. (38S and 34S RNA were assigned lengths of 2.375±0.113 μm and 1.810±0.073 μm respectively, giving a length difference of

0.565±0.186 μm, comparable with the value of 0.627±0.049 μm for 18S rRNA (9).) Furthermore, no "20S" precursor to 18S RNA (10) has been found in Xenopus cells (9,11). Taken together, these findings indicate that the cleavages which generate 38S and 34S RNA occur at or close to the 5' and 3' ends of the 18S sequence. We therefore searched the sequences flanking 18S rRNA for ability to base-pair with each other. The search was extended outwards to the boundaries of large, stable internal secondary structure features in the ETS (figure 3a) and the ITS (ref. 5). These secondary structure features may coincide with those nearest the 18S sequences that were seen by electron microscopy. Up to these boundaries the 18S-flanking sequences show only limited potential for internal secondary structure (see below) and are therefore theoretically free to make long range interactions.

Figure 4 shows the respective segments of the ETS and ITS aligned in antiparallel orientation. There is no potential for extensive base pairing between the ETS and ITS in the alignment shown, nor when the relative alignment is altered to search for better complementarity. Even the base compositions of the two seqeunces are largely non-complementary. Both sequences show quite extensive C-rich tracts, as already mentioned for the ETS. Other parts of the ITS sequence are rich in A plus G (figure 4) and there are no complementary CT-rich tracts in the 18S-flanking region of the ETS. There are a few regions in which a small number of consecutive base pairs can be drawn between the ETS and ITS. Some of these are indicated by vertical or sloping lines between the two sequences in figure 4. However, several of the respective nucleotides can be assigned equally convincingly to short internal loops within the ETS, as indicated by horizontal bridges in figure 4. Finally, there are two short regions in figure 4 where the potential for internal secondary structure is strong, as inferred experimentally from sequencing gels. These are shown as internal loops in the figure.

The data shown in figure 4 rule out the possibility of extensive base-pairing between the 18S-flanking regions of the transcription unit in pXr101. We have also examined the 18S-flanking sequences in other X.laevis rDNA clones. This was part of a larger study whose aims and findings will be reported in detail elsewhere (Stewart et al., manuscript in preparation). We note here that only minor differences were found between the 18S-flanking sequences in the different clones.

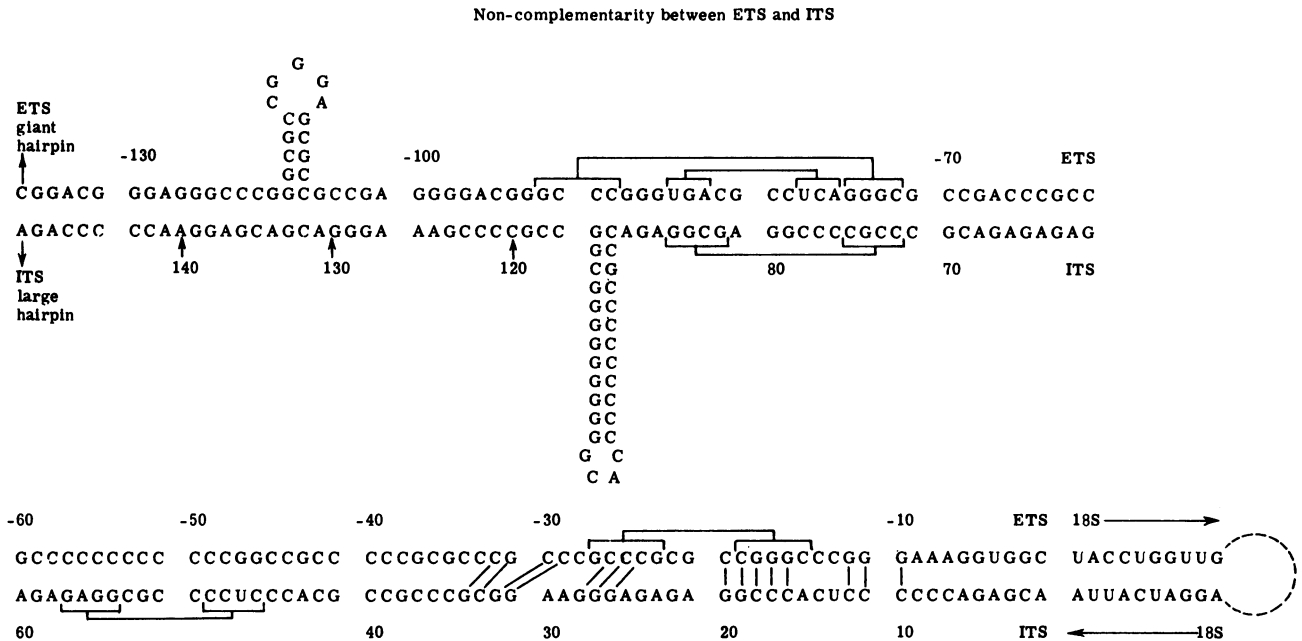

Figure 4 Alignment of the regions flanking the 18S gene in pXIr101 showing lack of potential for major base-pairing between the indicated regions of the ETS and ITS 1. The sequences are shown as for RNA and in antiparallel orientation. For further explanation see the text.

We draw the general conclusion that the 18S-flanking sequences of the ETS and ITS in X.laevis are non-complementary.

The occurrence of 38S and 34S RNA as distinct intermediates in ribosome maturation in X.laevis indicates that the cleavages on the 5' and 3' sides of 18S rRNA are kinetically separate events. This is consistent with the absence of physical pairing between the two cleavage sites as inferred here. Thus the combined data now available suggests that the molecular basis of rRNA processing differs substantially between X.laevis and E.coli. An intermediate situation has been reported in yeast (Saccharomyces carlsbergensis) (12). The sequence at the 5' end of the small subunit rRNA (17 rRNA) is cleaved simultaneously with a sequence downstream from the 3' end, but the potential for base-pairing between the cleavage sites (12) is very much smaller than in E.coli.

The ETS in pXIr101 is untranslatable

There are 44 T residues in the ETS but no ATG codons. This formally rules out any possibility that the ETS contains information for protein synthesis. The published data for pXI108 (2) lead to the same conclusion. Thus the ETS cannot play any unexpected role mediated through translation upon ribosomes, either before or after cleavage from the rest of the pre rRNA molecule. The absence of ATG codons in the X.laevis ETS suggests that lack of protein coding information is a general feature of the ETS in higher animals, and probably also in other eukaryotes, regardless of the details of primary structure, since such a property is unlikely to be phylogenetically variable.

The NTS in pXIr101

The general features of sequence organisation in the non transcribed spacers of X.laevis rDNA are known (1,2,13). A major part of the NTS comprises a pattern in which sequences called "Bam islands", which closely resemble the region preceding the site of initiation of transcription, are followed by repetitious regions with multiple Alu I sites (see ref. 2). Our findings for pXIr101 confirm this general pattern and are summarised here for completeness. The NTS of this clone contains two Bam islands about 1 Kb distant from each other (data not shown). The sequence of the second Bam island is shown in alignment with the pre-initiation region (figure 2). The two sequences show extensive (though imperfect) homology up to the site of initiation of transcription in the latter sequence. At precisely this point the homology begins

to break down. A few further blocks of partial homology can be recognised by postulating deletions in the Bam island (shown as lines) relative to the ETS. Then the Bam island merges into an "Alu repetitious region", which was not sequenced in this clone.

Concluding comments

The sequence data in figure 2 can be combined with data in refs. 3 and 5 to give a continuous sequence of 3924 nucleotides in the X.laevis rDNA clone pXIr101. The sequence extends from the end of the Alu repetitious region in the NTS through the various elements of the transcription unit to the start of the 28S gene. Completion of the ETS sequence has enabled various aspects of this region to be examined in detail, as described above. Availability of extensive sequence data from the single rDNA transcription unit and its preceding sequences in pXIr101 should also prove useful for a variety of comparative and evolutionary studies and in providing a basis for further functional studies. We have noticed some minor differences between the sequences reported here and the overlapping sequences reported previously from different clones (1,2). A full account of sequence variations in the transcribed regions of X.laevis rDNA will be reported elsewhere (Stewart et al., manuscript in preparation).

ACKNOWLEDGEMENTS

We thank John Forbes and Mary Robertson for skilled technical assistance. Mohammad Salim was on study leave from the Department of Biology, Quaid-i-Azam University,Islamabad, Pakistan. This work was supported by a grant from the Medical Research Council.

REFERENCES

1. Sollner-Webb, P. and Reeder, R.H. (1979). Cell 18, 485-499
2. Moss, T., Boseley, P.G. and Birnstiel, M.L. (1980). Nucleic Acids Research 8, 467-485
3. Salim, M. and Maden, B.E.H. (1981). Nature 291, 205-208
4. Salim, M. and Maden, B.E.H. (1980). Nucleic Acids Research 8, 2871-2884
5. Hall, L.M.C. and Maden, B.E.H. (1980). Nucleic Acids Research 8, 5993-6005
6. Gourse, R.L. and Gerbi, S.A. (1980). Nucleic Acids Research 8, 3263-3637
7. Young, R.A. and Steitz, J.A. (1978). Proc.Nat.Acad.Sci. U.S.A. 75, 3593-3597
8. Maxam, A.M. and Gilbert, W. (1980). Meth.Enzym. 65, 499-560
9. Wellauer, P.K. and Dawid, I.B. (1974). J.Mol.Biol. 89, 379-395
10. Weinberg, R.A. and Penman, S. (1970). J.Mol.Biol. 47, 169-178
11. Loening, U.K., Jones, K.W. and Birnstiel, M.L. (1969) J.Mol.Biol. 45, 353-366

12. Veldman, G.M., Klootwijk, J., Heerikhuizen, H. van, and Planta, R.J. (1981). Nucleic Acids Rsearch 9, 4847-4862
13. Botcham, P., Reeder, R.H. and Dawid, I.B. (1977) Cell 11, 599-607