
The external transcribed spacer and preceding region of *Xenopus borealis* rDNA: comparison with the corresponding region of *Xenopus laevis* rDNA

Judith C.Furlong¹, John Forbes, Mary Robertson and B.Edward H.Maden²

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

Received 7 October 1983; Accepted 4 November 1983

ABSTRACT

We report sequence data from a cloned rDNA unit from *Xenopus borealis*, extending leftwards from the 18S gene to overlap a region previously sequenced by R.Bach, B.Allet and M.Crippa (Nucleic Acids Research **9**, 5311-5330). Comparison with data from other species of *Xenopus* leads to the inference that the transcription initiation site in *X.borealis* is in the newly sequenced region and not, as was previously thought, in the region sequenced earlier. The *X.borealis* external transcribed spacer thus defined is some 612 nucleotides long, about 100 nucleotides shorter than in *X.laevis*. The *X.borealis* and *X.laevis* external transcribed spacers show a pattern of extensive but interrupted sequence divergence, with a large conserved tract starting about 100 nucleotides downstream from the transcription initiation site and shorter conserved tracts elsewhere. The regions in between the conserved tracts differ in length between the respective external transcribed spacers indicating that insertions and deletions have contributed to their divergence, as previously inferred for the internal transcribed spacers. Much of the overall length difference is in the region flanking the 18S gene, where there are also length microheterogeneities in *X.laevis* rDNA. As in *X.laevis*, the transcribed spacer sequences flanking the 18S gene in *X.borealis* contain no major tracts of mutual complementarity. The accumulated data on transcribed spacers in *Xenopus* render it unlikely that processing of ribosomal precursor RNA involves interaction between the regions flanking 18S RNA.

INTRODUCTION

It is well known that the transcribed spacers of ribosomal DNA (rDNA) in eukaryotes are phylogenetically much more variable than the sequences encoding mature rRNA (refs.1-5 and references therein). Sequence comparison between the internal transcribed spacers (ITS) of two species of African frog of the same genus, *Xenopus borealis* and *X.laevis*, has revealed a pattern of major sequence divergence with small conserved tracts (5). The relative spacing of the conserved elements in the *X.borealis* sequence with respect to that of *X.laevis* implies that insertions and deletions have played a major role in ITS sequence divergence (5). A similar conclusion has been reached for the ITS

regions of rat and mouse (6).

Here we describe the external transcribed spacer (ETS) and preceding region of *X.borealis* rDNA and we compare this with the previously determined sequence in *X.laevis* (7,8). An important requirement for this comparison was knowledge of the locations of the respective transcription initiation sites. The transcription initiation site in *X.laevis* has been unambiguously identified (9). Bach et al (10) identified the transcription initiation site in *X.clivii* and, with less certainty, a putative transcription initiation site in *X.borealis*. Our sequence data reveal a region to the right of that examined by Bach et al (10) which appears by sequence criteria to be a more favourable candidate for the transcription initiation site. Experiments by McStay and Bird (11) have confirmed the new identification. As in the ITS regions, a pattern of extensive but interrupted sequence divergence is found between the ETS regions of *X.borealis* and *X.laevis*, with evidence of a history of insertions and deletions, and particularly large differences between the respective ETS sequences flanking the 18S gene.

MATERIALS AND METHODS

Sequence analysis was carried out on the rDNA clone pXbr101. This recombinant contains a complete rDNA unit from amplified rDNA from *X.borealis* oocytes, cloned into the Hind III site of pMB9. The same clone was also used for our previous analysis of the ITS regions (see ref.5). To simplify the purification of restriction fragments the subclone pXbr101A was constructed. This contains the region from the Hind III site at the 3' end of the 28S gene, through the NTS and ETS to the Eco RI site in the 18S gene (see figure 1 of ref.5) cloned into pAT153.

The rDNA region of interest (figure 1) was subjected to restriction mapping by the procedure of Smith and Birnstiel (12). Sequence analysis was carried out by the method of Maxam and Gilbert (13) using the same protocols as previously (7). Most restriction fragments were labelled at the 5' ends; a few were labelled at the 3' ends using a "fill-in" reaction with an appropriate α -³²P labelled deoxynucleotide triphosphate. One sequencing run was carried out by the M13 dideoxy method (see the legend to figure 1).

The clones pXbr104, 105 and 106 are of homologous construction to pXbr101 and were used for comparative restriction analysis as summarised at the end of the results section.

RESULTS

Figure 1 shows the rDNA region sequenced in this work, the sequencing strategy, and the regions sequenced by Bach *et al* (10) and McStay and Bird (11). Figure 2 shows our sequence with the corresponding sequence from *X.laevis* (7) for comparison.

Overlap with sequence of Bach *et al*

The sequence of Bach *et al* (10) was derived from a different rDNA clone to ours. Nevertheless, comparison of the data revealed a clear overlap of some 400 nucleotides with only a few minimal differences between the two sequences.

The overlap includes the last part of the repetitive region of the NTS (10). In the repetitive region as a whole a number of different

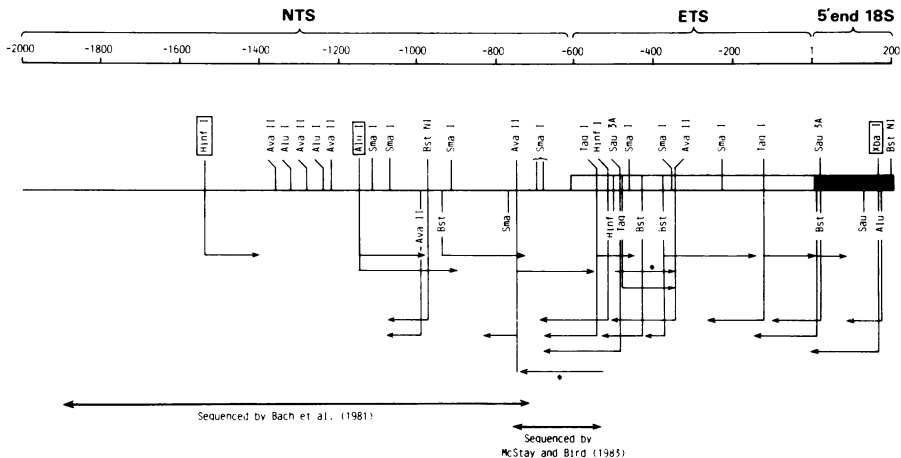


Figure 1 *X.borealis* NTS and ETS: summary of sequence analyses. The regions sequenced by Bach *et al* (10) and McStay and Bird (11) are shown. In the present work sequencing runs were carried out by the Maxam and Gilbert method on 5' labelled restriction fragments, with the following exceptions. The two runs marked by asterisks were carried out on 3' labelled material. The long run extending rightwards from the boxed Alu I site at nucleotide -1150 was obtained by the dideoxy terminator method after subcloning into bacteriophage M13 mp11. (This and other Alu I sites in the NTS are also Sst I sites: the fragment subcloned was the 1.3 kb Sst I/Xba I fragment). The general structure of the region to the left of the boxed Alu I site was examined by the restriction mapping procedure of Smith and Birnstiel (12). The pattern of interspersed Alu I and Ava II sites and the Hinf I site is reduplicated to the left but is not included in the diagram because the distances were not accurately calibrated. Although our sequence data agree closely with those of Bach *et al* (10) in the region of overlap it is possible that the respective rDNA clones differ in sequence organisation further to the left in the repetitive region of the NTS.



Figure 2 Sequence comparison between *X.borealis* and *X.laevis* rDNA in the ETS and adjacent region of the NTS. In both species the repetitive regions of the NTS contain Alu I/Sst I sites, but the sequence organisation surrounding these sites differs between the two species. The sequences shown here start at the last Alu I/Sst I site in the NTS and extend into the start of the 18S gene. The sequences are from the clones pXbr101 (*X.borealis*) and pXlr101 (*X.laevis*). A few uncertainties remain in the sequences (dots above nucleotides) concerning the exact number of nucleotides at the indicated positions; for example at nucleotide -10 in the ETS of *X.borealis* there was a spacing irregularity

X.laevis
NTS

Set I / A1u I

.....GAGCT -1001

end of repetitive region →

CGGGCAGGG GAGCCGGCTC GTCCCCGGC ACCGGAGGTC CCCGGGCCCT -951

TTGGCCCGG TTTTTTCGCA AAGTCCGGG CCCGCGGGA CTTGCTCGGC -901

CGGGCCGGG CCCGGGGCC CCCGGGGCCC GGGGCCCT CC CGCGGAGGC -851

CCGATGAGGA CGGATTCC CGGCCGCC CGGCCGGA GT TCCGGAGCC -801

CGGGGAGAG AGCCGGCGC CCGGCTCTC GGGCCC CCG CACGAAGCCT -751

CCATGCTACG CTTTTGGC A TTGCGGC C AGGAAGGTAG GGGAAGACCG -701

40S start →

GCCTCGGG CGACGGGCG CCGAAAAAG GACCGGGCG TTTCCCGCCT -651

tract 1

CGTCCCCG TCTGGGAAG CTCGGGT C GAGTCTCT CT CCCGGGCC -601

tract 2

ATCGATCTGG CAACCCGGC CCGGCCGGG AGGGCCCTCT GCCCGGCCA -551

tract 2

CCCCCGGG GGCGGGCCG TACACGGCA GGGAGGCTC CTCCGCCTC -501

GGCCGGGG CGACCCCGG ACCCCCCCCC GGTCCGGTC GCCGCGGGC -451

CCACCCCCCC GGCCGCCCG GGGCCGGG GGGCCCGG CCCCCCCGT -401

tract 3

GAGTTC CCC CGA CCGTCC GACCCCGCA GGGCCCGAA GAAGGCCCG -351

tract 4

ACAGCGGGG AGCCCGCCC GGGGACCCC GTTCCCCCG CCCGGGA CT -301

tract 4

CTCTCCCT GGCCGGGCC GAGCACCC CG TTTCGCCG GA CACCCGAGA -251

tract 5

GCGAGAGA AAGACGAAA GAAAGGAG TAGGCCGG GCCCCGTCC -201

GGCCGCCGC TCCCCCCC CTCCCGGG GGGGGAGC GCAGGCCGG -151

CGGGCCCCCC GGCCGGACG GGAGGGCCG GGGCCGGGA GCCGCGCGA -101

GGGACGGG CCGGTGAC CCTCAGGG CCGACCCGC GCCCCCCCC -51

CCCGCCCGC CCCGGCCCG CCCGCCCG CCGGGCCCG GAAAGGTGGC -1

Xb/XI -51 TGCACGTGTA AGTACGCAG GCCGGTACAG TGAAACTCG AATGGCTCAT -100
18S

which may possibly conceal a second A residue. The numbering system extends leftwards from the start of the 18S gene and is based on the best interpretation of sequence data at ambiguous points. Minor variants in the sequences including length microvariants occur in other X.laevis clones outside the 18S coding region (15), and possibly in X.borealis also, and this fact renders any numbering system somewhat arbitrary. Boxes surround regions of strong homology between X.borealis and X.laevis. In the ETS these homologous regions are denoted tracts 1-5. At the top of the X.borealis panel the boxing over the central regions of the 44 bp repeats denotes the short zones of strong homology with the 40S initiation sequence. For further details see the text.

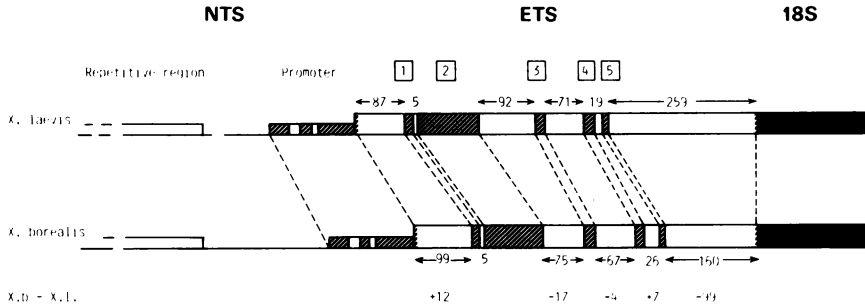


Figure 4 Summary diagram of ETS and preceding region in *X.laevis* and *X.borealis*, showing the relative locations of the tracts that are conserved between the two species. The 5' end of the 18S coding sequence is shown in black and is fully conserved (5). Conserved tracts in the ETS and promoter region are shaded. Conserved tracts which are internally located in the ETS (i.e. tracts which are non-contiguous with the transcription initiation site or 18S gene) are designated by boxed numerals which correspond to the numbers of the tracts in figure 2. Numbers in between these boxes, and shown within arrows where space permits, designate the lengths, in nucleotides, of the respective divergent tracts. X.b - X.l denotes the length differences between the divergent tracts in *X.borealis* and *X.laevis*: a minus sign signifies that the respective tract is shorter in *X.borealis* than in *X.laevis*, a plus sign signifies that the tract is longer in *X.borealis*.

figure 2 contains two of these features (underlined). The second of these corresponds to the last such feature in the region sequenced by Bach et al (10), and is followed by a transition from repetitive to unique sequence DNA. This non-repetitive DNA was thought by Bach et al to represent the ETS, and hence the last 13bp tract was thought to be the initiation site for transcription of 40S RNA. However, their S_1 nuclease protection experiments did not confirm this identification (10).

A region showing extensive homology with the *X.laevis* promoter

A search of the newly sequenced region in figure 2 revealed a more likely candidate sequence for the transcription initiation site. There is an extensive region of incomplete homology between the *X.borealis* sequence and a section of the *X.laevis* sequence extending from about 150 nucleotides upstream to a few nucleotides downstream of the transcription initiation site in *X.laevis* (figure 2). The region of homology corresponds to the *X.laevis* promoter defined by functional tests (14) (see also the discussion) and also to a region of strong sequence homology preceding and including the transcription initiation site in *X.clivii* (10). We infer that the *X.borealis* transcription initiation

site lies in this region as shown in figure 2, and not where it was previously thought to be. This inference has been substantiated by combined sequencing and S_1 mapping data obtained by McStay and Bird (11). The sequence homologies in the promoter regions of *X.laevis*, *X.clivii* and *X.borealis* are summarised in figure 3.

The ETS in *X.borealis*: major divergence from *X.laevis*

The ETS in pXbrl01 is 612 nucleotides long, 100 nucleotides shorter than in the *X.laevis* clone pXlrl01 (figure 2), from which it shows major sequence divergence. The sequences diverge within a few nucleotides of the transcription initiation site, and next to the 18S gene only three nucleotides are identical. However a number of homologous tracts occur internally within the ETS.

Homologous tracts in the ETS

The main tracts of homology are indicated by boxing in figure 2. The tracts thus identified contain a minimum of ten perfectly homologous nucleotides or a larger number with almost perfect homology. The longest tract starts about 120 nucleotides downstream from the transcription initiation site and is just over 100 nucleotides long. The homologous tracts occur in the same linear order in the two sequences.

Divergent tracts and length differences

In between the homologous tracts the sequences differ in content and length. The length differences are summarised in figure 4. These divergent regions contain several traces of partial homology, such as the C rich tract a short distance upstream from the start of the 18S gene. However, none of these is as clearcut as the tracts identified in figures 2 and 4.

The region flanking the 18S gene is the most extensive divergent tract in the ETS. This region differs in length by 100 nucleotides between *X.borealis* and *X.laevis* (figure 4).

Other *X.borealis* and *X.laevis* rDNA sources

We have carried out restriction tests on three further *X.borealis* rDNA clones, pXbrl04, 105 and 106, in parallel with pXbrl01. The tests afforded length calibration of the four clones in the region between the Alu I and Xba I restriction sites denoted by boxes in figure 1, and some data on the internal sequence organisation in this region. No differences were detected between the clones. We conclude that pXbrl01 is typical of *X.borealis* in its overall sequence organisation in this region. This is confirmed in part by the substantial agreement between

our sequence data and those of Bach *et al* (10) and McStay and Bird (11).

Comparative data on *X.laevis* rDNA clones show that pXlr101 is typical in its sequence organisation throughout the region discussed in this paper (7,8,9,15 and our unpublished data). However microheterogeneities occur in the transcribed spacers of *X.laevis* rDNA (15) as discussed later, below. We have not as yet examined *X.borealis* rDNA for microheterogeneities at the nucleotide sequence level in the ETS.

DISCUSSION

The promoter region

The present findings and those of McStay and Bird (11) lead to a revised assignment of the promoter region in *X.borealis*, which is of interest in relation to the assay for promoter function developed by Moss (14). The assay involved microinjection of *X.laevis* rDNA into *X.borealis* oocytes and detection of *X.laevis* transcripts. On the basis of the previously reported very limited homology between the *X.borealis* and *X.laevis* transcription initiation regions it was perhaps surprising that endogenous *X.borealis* RNA polymerase I functioned efficiently on an exogenous *X.laevis* rDNA template. The relatively high degree of sequence homology which is now evident is presumably a key factor in successful heterologous transcription.

Nevertheless, sequence homology between the respective promoter regions is not complete. One feature of the *X.laevis* promoter which has elicited comment is a tract of six T residues some 25 nucleotides upstream from the transcription start site (14,9). Only four of the six T residues occur in the *X.borealis* sequence: hence no absolute functional requirement can be attributed to the six T motif.

The region thought by Bach *et al* (10) to be the *X.borealis* 40S start site is the last of a series of very similar tracts in the repetitive NTS region, as already mentioned. It is possible that these partly promoter-like tracts may serve as pseudopromoters or RNA polymerase I binding sites, as is the case for the Bam islands and repetitive regions of the NTS in *X.laevis* (16).

Downstream from the promoter/transcription initiation region the *X.borealis* and *X.laevis* ETS sequences are fairly widely divergent for almost 100 nucleotides. This lack of homology enabled Moss (14) to distinguish *X.laevis* transcripts in the presence of excess *X.borealis* oocyte RNA by using an *X.laevis* rDNA probe terminating at the Taq I site

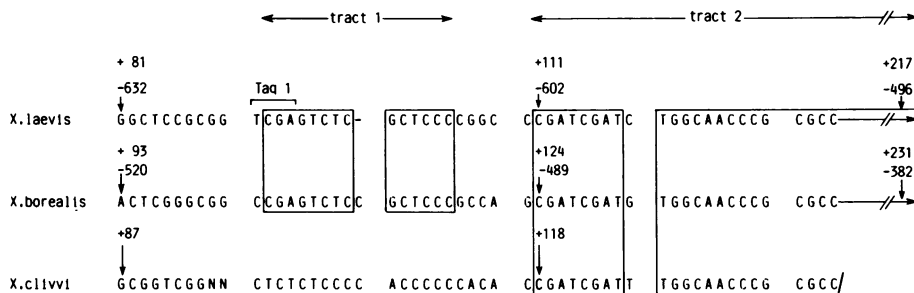


Figure 5 Sequence comparison through the left hand boundary of the major homology region in the *Xenopus* ETS. The *X.laevis* and *X.borealis* ETS sequences show no major homology tracts leftwards from the region shown, although there are traces of homology in the form of purine-rich or pyrimidine-rich blocks at similar locations in the sequences. The Taq I site in *X.laevis* borders the transition into a region of high homology with *X.borealis*. This region encompasses the relatively short tract 1 and then tract 2 whose full extent is shown in figure 2. The *X.clivii* sequence shows only a trace of homology with tract 1 in the form of a C-rich sequence (interrupted underlining) but shows extensive homology in tract 2 up to the end of the sequenced region (10), delineated in the figure by a slanting line. Nucleotides are numbered both positively in relation to the respective transcription initiation sites and negatively (in *X.laevis* and *X.borealis*) in relation to the 18S gene.

91 nucleotides downstream from the transcription initiation site. The present sequence data (figures 2 and 5) show that this Taq I site may well be the best possible restriction site for this particular purpose since immediately further downstream is a region of extensive sequence conservation.

The ETS: pattern of conserved and divergent tracts

The criteria adopted for definition of the homology tracts in figures 2 and 4 are the presence of ten perfectly homologous nucleotides or a longer tract of nucleotides showing at least 90% homology, as mentioned in the results section. The tracts so defined occur in the same linear order in the two sequences, and it is reasonable to conclude that they are the surviving, conserved remnants of a common ancestral sequence. It is not known whether any of these tracts plays a sequence-specific role in ribosome biosynthesis. However, all or part of the long tract 2 may be a possible candidate for such a role: *X.clivii* also contains this homology block starting at the same point (figure 5) and extending downstream to the end of the sequenced region. By contrast the shorter tract I has diverged in the *X.clivii* sequence

(figure 5). This finding implies that tract 1 has only survived unchanged by chance between *X.borealis* and *X.laevis*. Further comparative analyses should reveal which tracts, if any, have been consistently conserved.

As in the ITS regions (5), there are substantial length differences between the respective divergent tracts in the ETS regions of the two species (figure 4). This implies a history of insertions and deletions during divergence, which in turn suggests a lack of sequence-specific function in these regions.

Major divergence in the ETS region flanking the 18S gene

Sequence divergence by insertions and/or deletions appears to have been particularly active in the region flanking the 18S gene. If this region is defined for descriptive purposes as extending rightwards from homology tracts 4 and 5, then the *X.borealis* sequence is about 100 nucleotides shorter than that of *X.laevis* in this region.

The following findings from *X.laevis* may afford some insight into the apparent instability of this region. As already mentioned, the

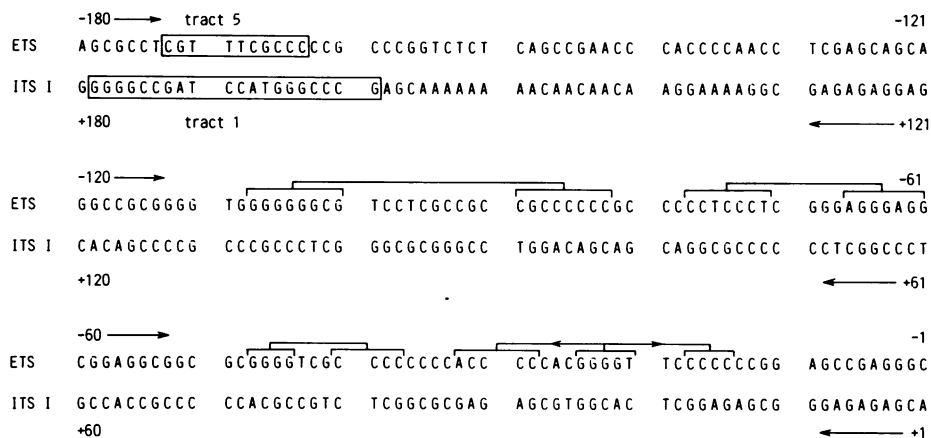


Figure 6 Alignment of the 18S flanking regions of the ETS and ITS I of *X.borealis* in antiparallel orientation. Although several short regions of potential complementarity can be found by sliding the alignments, there are no long regions of complementarity. Brackets in ETS show some blocks of nucleotides which would be expected to interact locally. Tract 5 is the respective ETS homology tract in figures 2 and 4; tract 1 is the respective ITS 1 homology tract in ref.5. The occurrence of these homology tracts at equal distances from the 18S gene in *X.borealis* seems to be fortuitous; in *X.laevis* the distances are unequal. For further discussion see the text.

transcribed spacers in *X.laevis* are known to contain several sites of microheterogeneity (15). In the 250 nucleotide region of the *X.laevis* ETS next to the 18S gene there are five sites of length microheterogeneity where individual variants differ in length by 1-3 nucleotides (figure 5 of ref.15). Thus the shortest and longest *X.laevis* sequences so far analysed differ in length by eight nucleotides in this region. Sequence comparisons in the remainder of the ETS have been less detailed, but two clones which show several differences between the 18S flanking regions (pXlr101 and pXl108, ref.15) show no further differences throughout the upstream region except at one or two sites where there are possible ambiguities in the sequence data (7,8). It was suggested that microheterogeneities in the transcribed spacers are indicative of a state of sequence flux which may underly larger-scale phylogenetic variability (15). The good correlation in the ETS between the specific distribution of length microheterogeneities in *X.laevis* and the region of greatest length difference between *X.laevis* and *X.borealis* is consistent with this suggestion.

One point of detail in this region is also of interest. In the *X.laevis* sequence there is a highly CG-rich tract which would generate a stable hairpin in RNA (nucleotides -218 to -137: see figure 3 of ref.7). In the *X.borealis* sequence this tract is absent or extremely changed: the G-rich tract at -110 may be a remnant. Hence it is unlikely that this secondary structure feature is of great functional significance.

Lack of complementarity between the regions of the ETS and ITS I flanking the 18S gene

In *E.coli*, the sequences flanking 16S rRNA in the primary ribosomal transcript interact to form an extensive base paired structure which contains a recognition site for the processing enzyme RNase III (17) (A similar interaction occurs between the sequences flanking 23S rRNA (18)).

There is no convincing evidence for such an interaction in *Xenopus*, and the available evidence is unfavourable. This evidence may be summarised as follows. In *X.laevis* the 18S-flanking regions lack the complementarity required for such an interaction, as discussed in detail in ref.7. In *X.borealis*, a similar search of the 180 nucleotides of the ETS flanking the 18S gene and the corresponding region of ITS I in antiparallel orientation to the ETS again reveal no complementary tracts

of more than a few nucleotides (figure 6). Moreover the potential for local secondary structure in parts of the ETS would compete against any long range interactions, and a highly A-rich tract in ITS I between nucleotides 120 and 160 lacks any T-rich counterpart in the ETS. The region of ITS I flanking the 18S gene, like the corresponding region of the ETS, appears to be a region of sequence instability. The distance from the 18S gene to the first homology tract in ITS I is some 64 nucleotides longer in *X.laevis* than in *X.borealis* (5); there are microheterogeneities within this region in *X.laevis* (15) and simple sequence tendencies in this region in both species, as discussed (5). The conserved tracts bounding the respective regions of the ETS and ITS I also do not appear to be designed for mutual interaction (figure 6, compare also figure 2 and ref.5). Finally the known intermediates in rRNA processing in both *Xenopus* and mammals indicate that kinetically distinct cleavages occur on the 5' and 3' flanks of 18S rRNA (19 and references therein, 20).

In conclusion, the findings summarised here imply that the transcribed spacer regions flanking 18S rRNA in *Xenopus* are phylogenetically highly variable and do not perform a concerted or other sequence-specific role in ribosome maturation. What is not clear is whether these regions are functionless or whether they perform some general function which is not sequence-dependent.

ACKNOWLEDGEMENTS

We thank R.H.Reeder for the pXbr series of clones, and B.McStay and A.Bird for communicating their findings to us in advance of publication. This work was supported by the Medical Research Council.

FOOTNOTES:

1. J.F.Furlong's present address is: Department of Biology, Brown University, Providence, Rhode Island, U.S.A.
2. Reprint requests should be sent to B.E.H.Maden, whose new address is: Department of Biochemistry, University of Liverpool, P.O.Box 147, Liverpool, U.K.

REFERENCES

1. Schibler, U., Wyler, T. and Hagenbuchle, O. (1975) *J. Mol. Biol.* **94**, 503-517.
 2. Veldman, G.M., Brand, R.C., Klootwijk, J. and Planta, R.J. (1980) *Nucleic Acids Research* **8**, 2907-2920.
 3. Hall, L.M.C. and Maden, B.E.H. (1980) *Nucleic Acids Research* **8**, 5993-6005.
-

4. Coen, E., Strachan, T. and Dover, G. (1982) *J. Mol. Biol.* 158, 17-35.
5. Furlong, J.C. and Maden, B.E.H. (1983) *The EMBO Journal* 2, 443-448.
6. Michot, B., Bachelierie, J-P. and Raynal, F. (1983) *Nucleic Acids Research* 11, 3375-3391.
7. Maden, B.E.H., Moss, M. and Salim, M. (1982) *Nucleic Acids Research* 10, 2387-2398.
8. Moss, T., Boseley, P.G. and Birnstiel, M.L. (1980) *Nucleic Acids Research* 8, 467-485.
9. Sollner-Webb, B. and Reeder, R.H. (1979) *Cell* 18, 485-499.
10. Bach, R. Allet, B. and Crippa, M. (1981) *Nucleic Acids Research* 9, 5311-5330.
11. McStay, B. and Bird, A. (1983) *Nucleic Acids Research*, submitted.
12. Smith, H.O. and Birnstiel, M.L. (1976) *Nucleic Acids Research* 3, 2387-2398.
13. Maxam, A.M. and Gilbert, W. (1980) in Grossman, L. and Moldave, K. (eds), *Nucleic Acids, Part I, Methods in Enzymology*, 65, Academic Press, N.Y. pp 499-560.
14. Moss, T. (1982) *Cell* 30, 835-842.
15. Stewart, M.A., Hall, L.M.C. and Maden, B.E.H. (1983) *Nucleic Acids Research* 11, 629-646.
16. Moss, T. (1983) *Nature* 302, 223-228.
17. Young, R.A. and Steitz, J.A. (1978) *Proc. Nat. Acad. Sci. U.S.A.* 75, 3593-3597.
18. Bram, R.J., Young, R.A. and Steitz, J.A. (1980) *Cell* 19, 393-401.
19. Wellauer, P.K., Dawid, I.B., Kelley, D.E. and Perry, R.P. (1974) *J. Mol. Biol.* 89, 397-407.
20. Weinberg, R.A. and Penman, S. (1970) *J. Mol. Biol.* 47, 169-178.