The EMBO Journal Vol.2 No.3 pp.443-448, 1983

Patterns of major divergence between the internal transcribed spacers of ribosomal DNA in *Xenopus borealis* and *Xenopus laevis*, and of minimal divergence within ribosomal coding regions

J.C.Furlong and B.E.H.Maden*

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

Communicated by R.A.Garrett
Received on 13 January 1983

We have determined the nucleotide sequences of the two internal transcribed spacers, the adjacent ribosomal coding sequences and the boundary between the external transcribed spacer and the 18S coding sequence in a cloned ribosomal transcription unit from *Xenopus borealis*. The transcribed spacers differ very extensively from those of *X. laevis*. Nevertheless, embedded in the internal transcribed spacers are several short sequence elements which are identical between the two species. These conserved elements are laterally displaced by substantial distances in the *X. borealis* sequence with respect to that of *X. laevis*. These relative displacements imply that insertions and deletions have played a major role in transcribed spacer divergence in *Xenopus*. This in turn implies that large regions of the transcribed spacers do not play a sequence-specific role in ribosome maturation. In contrast, the sequenced parts of the ribosomal coding regions, which encompass 670 nucleotides, differ at only three points from the corresponding sequences in *X. laevis*, each by a single substitution. These substitutions are readily accommodated by current models for rRNA higher order structure.

Key words: divergence/ribosomal/transcribed spacers/*Xenopus*

Introduction

There is considerable uncertainty as to the role of transcribed spacers during the maturation of eukaryotic ribosomes from ribosomal precursor RNA. These components of the primary ribosomal transcript (Figure 1) are eliminated during ribosome maturation and are much more variable in primary structure between different taxonomic groups than are the rRNA sequences. Whereas comparative sequence data on rRNA have revealed structural features that are conserved across an extremely broad phylogenetic range (see for example, Zwieb *et al.*, 1981), sequence data on transcribed spacers have revealed practically no homology between three eukaryotes which have been studied in detail: *Saccharomyces* (Skryabin *et al.*, 1979a, 1979b; Veldman *et al.*, 1980, 1981a), *Xenopus laevis* (Hall and Maden, 1980; Maden *et al.*, 1982a) and rat (Subrahmanyam *et al.*, 1982).

In the absence of sequence conservation between the transcribed spacers of distantly related eukaryotes it is important to seek among more closely related species for the processes that generate divergence. The transcribed spacers of the frogs, *X. laevis* and *X. borealis*, are known to lack major sequence homology by criteria of nucleic acid hybridisation (Brown *et al.*, 1972; Forsheit *et al.*, 1974). Here we have extended this work by carrying out comparative sequence analysis. Our findings indicate that large sections of the transcribed spacers are tolerant to major changes including insertions and deletions.

*To whom reprint requests should be sent.

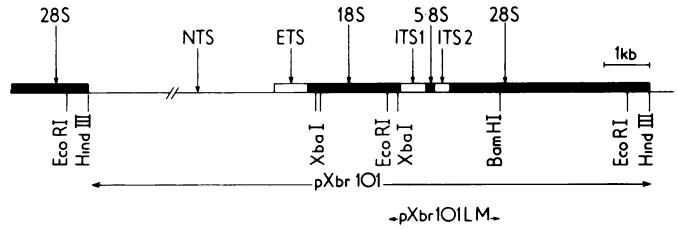

Fig. 1. Scheme of repeating structure of rDNA in *X. borealis*. NTS, ETS, ITS 1 and ITS 2 denote non-transcribed, external transcribed, first and second internal transcribed spacers, respectively. The locations of the restriction sites for *Eco*RI, *Hind*III, *Xba*I and *Bam*HI are shown. The clone pXbr101 and the subclone pXbr101LM were used in this work. For further details see Materials and methods.

Such changes are generally disruptive to functions that are closely coupled to sequence content; hence we infer that no such close coupling exists in large parts of the transcribed spacers of *Xenopus*. In contrast, the sequenced parts of the rRNA coding regions contain only three sites of variation from *X. laevis*; these differences can be readily accommodated in current models for rRNA higher order structure.

Results

The main sequence analysis was carried out on the *X. borealis* rDNA clone, pXbr101, and its subclone pXbr101LM (see Materials and methods). Some confirmatory data were obtained from other clones, as described later. Figure 2 shows the sequence data from pXbr101, together with corresponding data from the *X. laevis* clone pXlr101. The layout of the figure emphasises the main findings, which are as follows.

Major divergence between transcribed spacers

The transcribed spacers differ extensively between the two sources. The sequences flanking the 18S coding regions diverge within a few nucleotides of the gene-spacer boundaries. Then, on the 5′ flank of the 5.8S sequence there is only a single conserved nucleotide, while on the 3′ flank there is a larger conserved tract of 26 nucleotides. On the 5′ flank of 28S rRNA there are only four contiguous conserved nucleotides, but both sequences are rich in C for some distance upstream.

Conserved tracts in the internal transcribed spacers

Embedded in the largely divergent sequences of the two internal transcribed spacers are short tracts of completely or almost completely conserved sequences. There are three such tracts in the first internal transcribed spacer (ITS 1) and four in the second internal transcribed spacer (ITS 2). The tracts range in length from 13 to 28 nucleotides (Figure 2). The conserved tracts constitute a smaller amount of the total spacer in ITS 1 than in ITS 2, both in relative and in absolute terms. The conserved tracts in ITS 1 are at widely spaced sites (see below) but those in ITS 2 are concentrated in the 5′ part of the spacer.

Lateral displacement of conserved tracts

The distances between the conserved tracts differ in *X. borealis* and *X. laevis*. In ITS 1 some of these length dif-

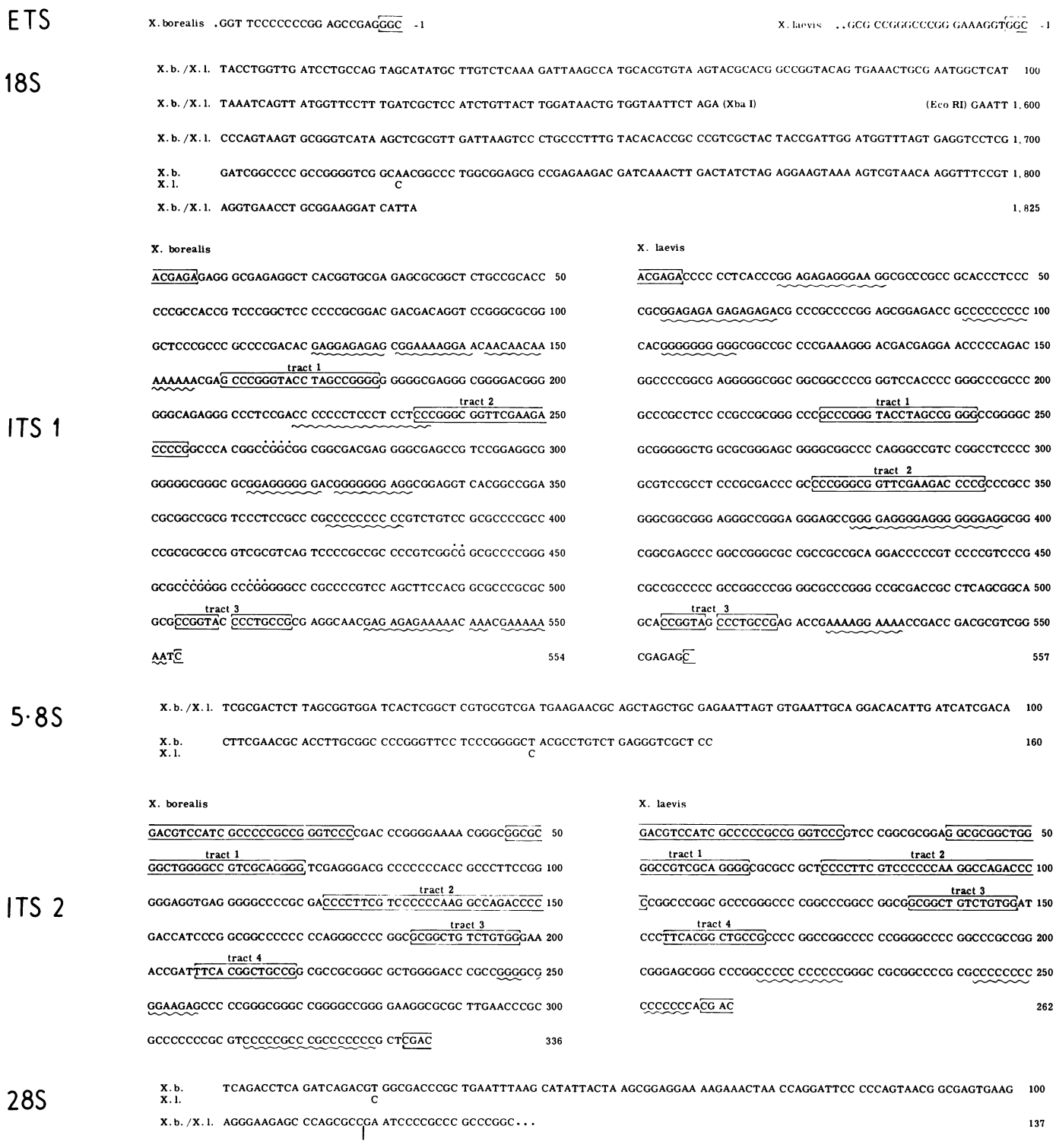

Fig. 2. Sequence data from *X. borealis* and *X. laevis* rDNA. The data for the ribosomal coding regions are very similar for the two species, and are shown in single panels. Data for the transcribed spacers are shown in separate panels for the two species. For the ETS, only the short section immediately preceding the 18S gene is shown. This is followed by the first section of the 18S gene. Then, from the *Eco*RI site in the 18S gene (position 1596 in the *X. laevis* 18S sequence) the rDNA sequence data are continuous through to the start of the 28S gene. Nucleotides are numbered from the first nucleotide in each region (ITS 1, etc.). This numbering system replaces the provisional system of Hall and Maden (1980). Sequences in the transcribed spacers that are conserved between the two species are boxed, and internally located conserved sequences in the spacers (i.e., those which are not contiguous with the rRNA sequences) are denoted tract 1, etc. (see also Figure 3). Wavy lines denote tracts of 10 or more nucleotides in the spacers that consist of purines only or pyrimidines only, or which show strong bias towards a single type of base. These are denoted 'simple sequence' tracts in the text. Superscript dots over nucleotides in ITS 1 in *X. borealis* denote locations where the sequence is not completely certain (see Materials and methods). The *X. borealis* data are from the clone pXbr101 and its subclone pXbr101LM. The *X. laevis* data are from the clone pXlr101 (Hall and Maden, 1980; Salim and Maden, 1981; Maden *et al.*, 1982a). See the last paragraph of Results for comments on intraspecies variants. Also note that the 5.8S DNA sequences differ slightly from the versions that were originally derived from rRNA, as discussed by Hall and Maden (1980). 5.8S rRNA in *X. laevis* shows terminal heterogeneity (Ford and Mathieson, 1978; Khan and Maden, 1977); the sequence shown here is for the longest possible rRNA molecules. The 28S sequence in *X. laevis* has only been determined to nucleotide 118 (vertical line at bottom of figure).

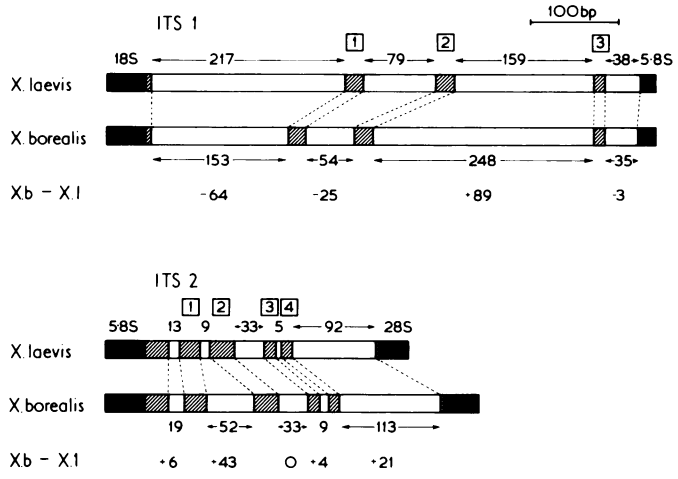

Fig. 3. Summary diagram of ITS 1 and ITS 2 in *X. laevis* and *X. borealis*, showing the relative locations of the sequence tracts that are conserved between the two species. The ends of the rRNA coding sequences are shown in black. Conserved tracts in the transcribed spacers are shaded; divergent tracts are unshaded. Conserved tracts which are internally located in the spacers (i.e., those which are non-contiguous with the rRNA coding regions) are designated by boxed numerals, which correspond to the numbers of the tracts in Figure 2. The other numbers on these lines designate the sequence lengths, in nucleotides, of the respective divergent tracts. X.b.-X.l. denotes the differences in lengths 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*.

ferences are considerable, as summarised in Figure 3. However, in ITS 1 the length differences also largely compensate for each other, so that this spacer is approximately equal in length in the two rDNA sources. In ITS 2 the length differences do not compensate, so that this spacer is considerably longer in *X. borealis* than in *X. laevis*.

Extent of variation in divergent sequences

The regions between the conserved tracts generally differ extensively in their sequence patterns as well as in length. For example, at position ~100 in *X. laevis* ITS 1 there is a distinctive tract of 10 cytosines followed, after two nucleotides, by 10 guanines. There is no obvious trace of this feature in *X. borealis*. Then, shortly before the first conserved tract in ITS 1, the *X. borealis* sequence is A-rich whereas the *X. laevis* sequence consists almost entirely of G + C. However, there are some short regions of partial homology outside the main conserved tracts. For example, in the region of ITS 2 preceding the 28S sequence, both species possess C-rich tracts as already mentioned, but the *X. laevis* sequence is even richer in C than that of *X. borealis*. In the region of ITS 1 preceding the 5.8S gene both species possess A-rich motifs. That in *X. borealis* is more extensive than that in *X. laevis*.

Simple sequence tendencies

There are no large scale repetitious sequence patterns in the transcribed spacers such as those which occur in the non-transcribed spacers (Moss *et al.*, 1980). However, there are tendencies towards 'simple sequence' patterns such as homopolymeric tracts, pyrimidine tracts and purine tracts. The locations of some of these tracts have been mentioned and several are indicated in Figure 2. Whereas traces of some of these features occur in homologous locations in both rDNA sources, as mentioned above, others are characteristic of one or other rDNA source.

Ribosomal coding sequences

The sequenced parts of the ribosomal coding regions are extremely similar between the two sources. There is one difference between the 3′ regions of the respective 18S sequences, one between the 5.8S sequences and one near the 5′ end of the 28S sequence. Each of these differences is a base substitution. The effects of these substitutions in models for rRNA secondary structure and interaction are considered at the end of the Discussion.

Other X. borealis and X. laevis rDNA sources

Other *X. borealis* rDNA clones were analysed with various restriction enzymes in the regions encompassed by the present study, and yielded data that were indistinguishable from pXbr101. However, partial sequence analysis on one clone, pXbr106, showed evidence for a few sites of minor variation from pXbr101 in the internal transcribed spacers; the variant sites remain to be fully characterised. In *X. laevis* rDNA, the transcribed spacers also show several sites of microheterogeneity (Stewart *et al.*, 1983), as summarised in the Discussion. In the rRNA coding regions, at the three sites of variation between *X. borealis* and *X. laevis* (Figure 2) the respective sequences appear to be the predominant ones in the two species as revealed by analysis of one or more further clones (this work and Maden *et al.*, 1982b), of uncloned *X. laevis* rDNA (Maden *et al.*, 1982b) and direct analysis of 5.8S rRNA (Ford and Mathieson, 1978).

Discussion

Transcribed spacers

Does detailed sequence information in the transcribed spacers of *Xenopus* determine their accurate excision during ribosome maturation? In *Escherichia coli* the sequences flanking 16S and 23S rRNA, respectively, form extensive base-paired structures which are recognised and cleaved by ribonuclease III during ribosome maturation (Young and Steitz, 1978; Bram *et al.*, 1980). In *Saccharomyces carlsbergensis* the possibility of long range interactions of very limited scale involving RNA processing points has also been noted but it is clear that large parts of the transcribed spacers do not interact in this way (Veldman *et al.*, 1981a). In *X. laevis* the sequences flanking 18S rRNA show no significant potential for interaction (Maden *et al.*, 1982a). No clear cut molecular model for the function of transcribed spacers during ribosome maturation has been formulated for *X. laevis* or other higher organisms.

The nature of the processes that have generated divergence between the internal transcribed spacers of *X. laevis* and *X. borealis* can be inferred from the sequence data. First, the precision of homology in the conserved tracts (Figure 2) and their occurrence in the same linear order in the transcribed spacers of the two species indicate that these tracts are indeed the unchanged remnants of common ancestral ITS sequences. Secondly, the considerable interspecies differences in spacing between some of the conserved tracts (Figure 3) indicate a history of insertions and/or deletions in the intervening, divergent regions. Thirdly, the general lack of residual homology in the divergent regions implies that multiple changes have occurred in these regions. Fourthly, and perhaps relatedly, there is a tendency for simple sequence tracts (Figure 2) to be generated apparently at random in the divergent regions. In summary, it may be inferred that the sequence contents of the divergent regions have undergone ex-

tensive change through the cumulative effects of many individual occurrences including insertions, deletions and (presumably) point mutations.

Evidence suggesting incipient sequence instability in the transcribed spacers has also been found within *X. laevis*. Detailed analysis of several rDNA clones and of uncloned rDNA has revealed several sites of heterogeneity in all three transcribed spacers (Stewart *et al.*, 1983). The heterogeneities comprise base substitutions and insertions and deletions of one to several nucleotides. Some of the latter occur in simple sequence oligo(C) and oligopurine tracts. All of the heterogeneities occur in regions showing major divergence between *X. laevis* and *X. borealis*. Partial sequence data on another *X. borealis* clone, pXbr106, have also revealed sites of minor variation from the *X. borealis* sequence in Figure 2, as mentioned at the end of the Results section. These heterogeneities imply a state of sequence flux which, given sufficient time and the condition of genetic isolation, could provide the basis for large scale phylogenetic divergence in the transcribed spacers.

Changes of the nature and extent that have occurred during divergence of the transcribed spacers in *X. borealis* and *X. laevis* would almost certainly be disruptive to functions that require the recognition of, or interaction between, particular features of nucleotide sequence. A plausible interpretation of the data is that large parts of the transcribed spacers in *Xenopus*, including regions located only a few nucleotides from the boundaries of the ribosomal sequences, do not function in a sequence-specific manner during ribosome maturation. Thus, the transcribed spacers might resemble some introns with respect to evolution under the relative absence of selection for sequence content (Konkel *et al.*, 1979), while differing from introns in that splicing is not involved in processing.

It is not clear whether the conserved sequences in the ITS have remained unchanged due to local functional constraints or by chance. One interesting feature of the conserved tracts in ITS 1 is that, in DNA, each is centred upon a site of dyad symmetry containing all four types of base. The DNA tracts contain cleavage sites for the enzymes *Kpn*I (tract 1), *Taq*I (tract 2) and *Kpn*I again (tract 3 in *X. borealis*, mutated by one base in *X. laevis*). Moreover, a 10-bp region containing the first *Kpn*I site is also present in ITS 1 in rat (Subrahmanyam *et al.*, 1982). However, none of the other conserved tracts noted here occurs in rat. Thus, any selective constraints, even in these short regions, must be relative rather than absolute. Moreover, the dyad symmetry correlation does not apply to the conserved tracts in ITS 2 in *Xenopus*. The question of function, if any, of the conserved tracts, either in DNA or in the transcript, may be resolved by further comparative sequence analysis and possibly by a search for initial cleavage sites in RNA at points remote from the boundaries of the ribosomal sequences. The results of a previous such search in *Xenopus* were negative (Wellauer and David, 1974).

Thus, the main findings described in this paper contribute to and highlight a growing body of evidence that is consistent with the following conclusions and inferences. Large parts of the transcribed spacers in eukaryotes undergo rapid phylogenetic change including insertions and deletions. Change of this nature and extent would almost certainly be disruptive to any function that is closely related to sequence content. It seems unlikely that such functions are mediated by the bulk of the transcribed spacers. Whether the short conserved tracts are under some degree of functional constraint or have remained unchanged by chance is at present unknown.

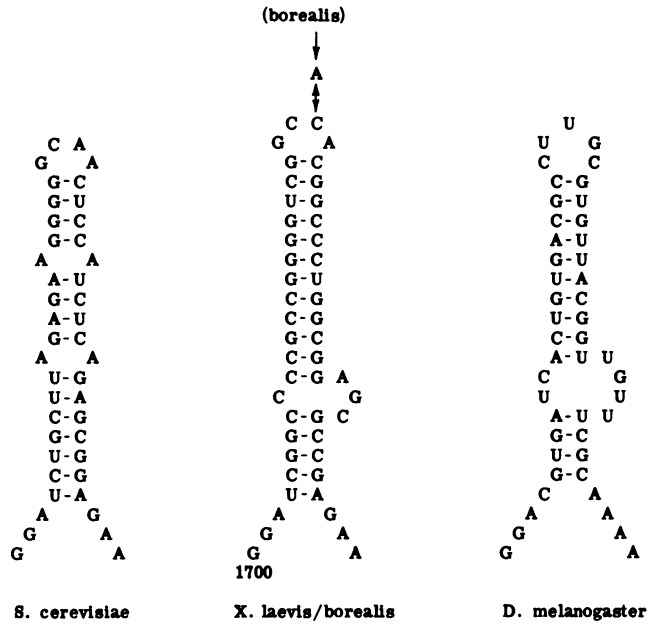

Fig. 4. The distal part of the penultimate arm in the secondary structure model for 18S rRNA, showing the inferred structures for *S. cerevisiae*, *Xenopus* and *Drosophila melanogaster*. Sequence data for yeast are from Rubstov *et al.* (1980) and for *D. melanogaster* are from Jordan *et al.* (1980). The secondary structures for yeast and *Xenopus* are as shown in the complete secondary structure model of 18S rRNA (Zweib *et al.*, 1981). The *Drosophila* sequence was fitted to the relevant part of the model in the present work.

An alternative view of ribosome formation in higher eukaryotes, including *Xenopus*, is that excision of rRNA from the precursor is signalled by the completion of steps in the assembly of nascent ribosomes themselves. This would be consistent with two early observations on ribosome formation in animal cells; (i) precursor rRNA cleavage occurs within nascent ribosomal particles in the nucleolus (Warner and Soeiro, 1967), (ii) metabolic perturbations which affect protein synthesis, and which might therefore be expected to affect ribosome assembly, correspondingly affect the kinetics of cleavage of ribosomal precursor RNA (Willems *et al.*, 1969; Maden *et al.*, 1969).

Variants in ribosomal coding sequences

The variants in the regions of the ribosomal coding sequences that have been examined are minimal in extent but are of interest in relation to models of rRNA secondary structure and interaction. The one site of variation in the 18S sequence data in Figure 2 is a C→A substitution ~100 nucleotides from the 3′ end of the sequence. In models of 18S rRNA secondary structure (Zwieb *et al.*, 1981; Stiegler *et al.*, 1981) this site is located at the extreme tip of a helical arm whose primary structure is among the most variable parts of the 18S sequence when distantly related eukaryotes are compared (Figure 4). Hence this site of minimal variation between *X. borealis* and *X. laevis* is in accordance with larger scale phylogenetic trends.

There is a single site of variation between the respective 5.8S sequences and another near the 5′ end of the respective 28S sequences. A working model for the interaction between 5.8S and 28S rRNA has recently been proposed by several groups of workers (Pace *et al.*, 1977; Veldman *et al.*, 1981b; Walker *et al.*, 1982; Michot *et al.*, 1982). The *Xenopus* se-

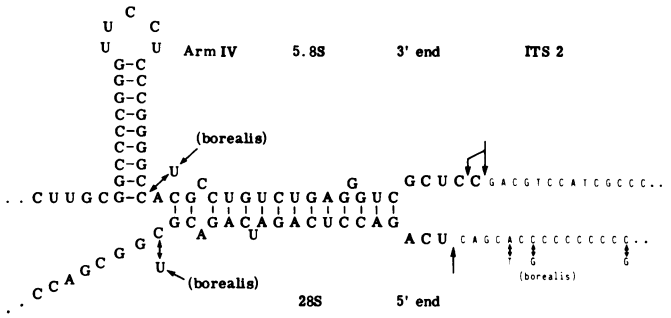

Fig. 5. Probable interaction between the 3′ region of 5.8S rRNA and 5′ end of 28S rRNA. Evidence in support of this general model includes sequence data from various eukaryotes, homology with the 5′ region of *E. coli* 23S rRNA and direct chemical evidence on 5.8S rRNA complexed with 28S rRNA (Veldman *et al.*, 1981b; Michot *et al.*, 1982; Nazar, 1980; Pace *et al.*, 1977, Walker *et al.*, 1982). The diagram is adapted from Walker *et al.* (1982). Arm IV of 5.8S rRNA is the internal helix referred to in the text. Also shown are the adjacent parts of ITS 2. There is no apparent potential for interaction between these regions of ITS 2 in either species of *Xenopus*. The double arrow at the 3′ end of 5.8S rRNA denotes probable terminal heterogeneity (Khan and Maden, 1977).

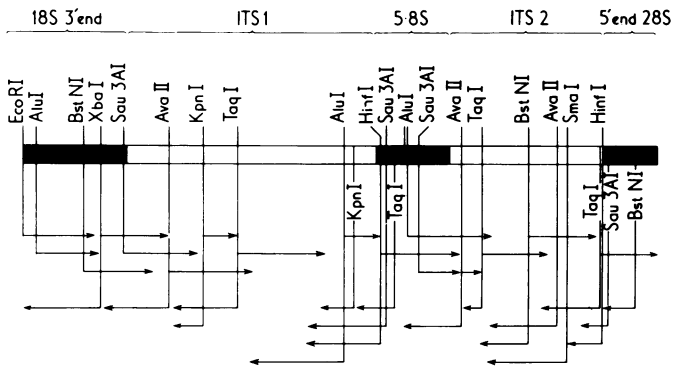

Fig. 6. Sequencing strategy for the region of rDNA from the *Eco*RI site at the 3′ end of 18S to the 5′ end of 28S as described in Figure 2. Only those sites used for labelling following primary restriction are noted in the figure and so this is not a complete restriction map. The beginning of each arrow denotes the restriction site at which the fragment was labelled at the 5′ end. The arrow tip denotes the longest reading of all the gels starting from the indicated restriction site.

quences are shown fitted to this model in Figure 5. Interestingly the single base changes between *X. borealis* and *X. laevis* occur in close proximity to each other in the model. Pace *et al.* (1977) and Walker *et al.* (1982) have discussed the possibility that the internal 5.8S helix IV (Figure 5) may contribute in some as yet undefined way to the intermolecular interaction with 28S. It may be noted here that the slightly destabilising change in the internal 5.8S helix in *X. borealis* is matched by a stabilizing change in the intermolecular interaction (Figure 5). Thus the two changes might possibly be compensating in terms of overall stability of the region. Comparative data between more distantly related organisms show further variations in the details of intramolecular and intermolecular interactions in these regions, while conserving the overall scheme as shown in Figure 5 (Veldman *et al.*, 1981b; Michot *et al.*, 1982).

Materials and methods

The rDNA clone pXbr101 (see Figure 1) is one of a series of recombinants that were obtained by digesting amplified rDNA from *X. borealis* oocytes with *Hind*III and cloning into the *Hind*III site of the plasmid pMB9 (R.Reeder, personal communication). [In early work mentioned in the Introduction (Brown *et al.*, 1972; Forsheit *et al.*, 1974) this species was mistakenly identified as *X. mulleri*; see Brown *et al.*, 1977.] Other recombinants in this series were designated pXbr102–106. Each contains a complete unit of the rDNA repeating structure from the start of the non-transcribed spacer to the end of the 28S gene (Figure 1). The clones correspond to the pXlr101 series of *X. laevis* rDNA clones, also constructed by R.Reeder and used in previous studies in this laboratory. Most of the sequence data described in this paper were obtained after first subcloning the indicated *Eco*RI-*Bam*HI fragment (Figure 1) into pBR322, this subclone being designated pXbr101LM. Sequence data encompassing the 5′ end of the 18S gene were obtained directly from pXbr101, utilizing the *Xba*I site near the left hand end of the gene (Figure 1). (Among the differences between *X. laevis* and *X. borealis* rDNA is the presence in the former and absence from the latter of a *Bam*HI site in ITS 2. Thus the *Eco*RI-*Bam*HI region 'LM' in *X. borealis* is represented by two smaller, separately subcloned regions, L and M in *X. laevis*; Hall and Maden, 1980.)

A restriction map of the 18S–28S intergene region in pXbr101LM was obtained by the method of Smith and Birnstiel (1976) and this provided the basis for sequencing (Figure 6) by the method of Maxam and Gilbert (1980). Restriction fragments were subjected to 5′ labelling followed by secondary restriction and chemical cleavage reactions as in Maden *et al.* (1982a). Most of the sequence was unambiguously established from data on both strands. A few short regions were sequenced on one strand only, due to lack of suitable restriction sites for sequencing the other strand. In individual sequencing gels the main uncertainties were caused by compression artefacts in GC-rich parts of the sequence, as was also found in the *X. laevis* transcribed spacers (Hall and Maden 1980; Maden *et al.*, 1982a). Most of these uncertainties were resolved from the combined data from several gels. In ITS 1 there remained a few uncertainties in the exact number and/or order of C and G residues at a few sites (indicated by superscript dots in Figure 2) due to persistent secondary structure. We believe that the sequences shown at these points are approximately correct; any minor errors at these sites would not affect our main conclusions.

Acknowledgements

We thank R.Reeder for the gift of rDNA clones, and J.Forbes and M.Robertson for technical assistance. This work was supported by the Medical Research Council.

References

Bram,R.J., Young,R.A. and Steitz,J.A. (1980) *Cell*, **19**, 393-401.
Brown,D.D., Wensink,P.C. and Jordan,E. (1972) *J. Mol. Biol.*, **63**, 57-73.
Brown,D.D., Dawid,I.B. and Reeder,R.H. (1977) *Dev. Biol.*, **59**, 266-267.
Ford,P.J. and Mathieson,T. (1978) *Eur. J. Biochem.*, **87**, 199-214.
Forsheit,A.B., Davidson,N. and Brown D.D. (1974) *J. Mol. Biol.*, **90**, 301-314.
Hall,L.M.C. and Maden,B.E.H. (1980) *Nucleic Acids Res.*, **8**, 5993-6005.
Jordan,B.R., Latil-Damotte,M. and Jourdan,R. (1980) *FEBS Lett.*, **117**, 227-231.
Khan,M.S.N. and Maden,B.E.H. (1977) *Nucleic Acids Res.*, **4**, 2495-2505.
Konkel,D.A., Maizel,J.V. and Leder,P. (1979) *Cell*, **18**, 865-873.
Maden,B.E.H., Vaughan,M.H., Warner,J.R. and Darnell,J.E. (1969) *J.Mol. Biol.*, **45**, 265-275.
Maden,B.E.H., Moss,M. and Salim,M. (1982a) *Nucleic Acids Res.*, **10**, 2387-2398.
Maden,B.E.H., Forbes,J.M., Stewart,M.A. and Eason,R. (1982b) *EMBO J.*, **1**, 597-601.
Maxam,A.M. and Gilbert,W. (1980) in Grossman,L. and Modave,K. (eds.), *Nucleic Acids, Part I, Methods in Enzymology*, **65**, Academic Press, NY, pp. 499-560.
Michot,B., Bachellerie,J.-P., Raynal,F. and Renalier,M.-H. (1982) *FEBS Lett.*, **140**, 193-197.
Moss,T., Boseley,P. and Birnstiel,M. (1980) *Nucleic Acids Res.*, **8**, 467-485.
Nazar,R.N. (1980) *FEBS Lett.*, **119**, 212-214.
Pace,N.R., Walker,T.A. and Schroeder,E. (1977) *Biochemistry (Wash.)*, **16**, 5321-5328.
Rubstov,P.M., Musakhanov,M.M., Zakharyev,V.M., Krayev,A.S., Skryabin,K.G. and Bayev,A.A. (1980) *Nucleic Acids Res.*, **8**, 5779-5794.
Salim,M. and Maden,B.E.H. (1981) *Nature*, **291**, 205-208.
Skryabin,K.G., Krayev,A.S., Rubstov,P.M. and Bayev,A.A. (1979a) *Dokl. Akad. Nauk. SSSR*, **247**, 761-765.
Skryabin,K.G., Zakharyev,V.M., Rubstov,P.M. and Bayev,A.A. (1979b) *Dokl. Akad. Nauk. SSSR*, **247**, 1275-1277.

Smith,H.O. and Birnstiel,M.L. (1976) *Nucleic Acids Res.*, **3**, 2387-2398.
Stewart,M.A., Hall,L.M.C. and Maden,B.E.H. (1983) *Nucleic Acids Res.*, **11**, 629-646.
Stiegler,P., Carbon,P., Ebel,J.-P. and Ehresmann,C. (1981) *Eur. J. Biochem.*, **120**, 487-495.
Subrahmanyam,C.S., Cassidy,B., Busch,H. and Rothblum,L.I. (1982) *Nucleic Acids Res.*, **10**, 3667-3680.
Veldman,G.M., Brand,R.C., Klootwijk,J. and Planta,R.J. (1980) *Nucleic Acids Res.*, **8**, 2907-2920.
Veldman,G.M., Klootwijk,J., van Heerikhuizen,H. and Planta,R.J. (1981a) *Nucleic Acids Res.*, **9**, 4847-4862.
Veldman,G.M., Klootwijk,J., de Regt,V.C.H.F., Planta,R.J., Branlant,C., Krol,A. and Ebel,J.-P. (1981b) *Nucleic Acids Res.*, **9**, 6935-6952.
Walker,T.A., Johnson,K.D., Olsen,G.T., Peters,M.A. and Pace,N.R. (1982) *Biochemistry (Wash.)*, **21**, 2320-2329.
Warner,J.R. and Soeiro,R. (1967) *Proc. Natl. Acad. Sci. USA*, **58**, 1984-1990.
Wellauer,P.K. and Dawid,I.B. (1974) *J. Mol. Biol.*, **89**, 379-395.
Willems,M., Penman,M. and Penman,S. (1969) *J. Cell Biol.*, **41**, 177-187.
Young,R.A. and Steitz,J.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3593-3597.
Zwieb,C., Glotz,C. and Brimacombe,R. (1981) *Nucleic Acids Res.*, **9**, 3621-3640.