Mapping of the Xenopus laevis 5.8S rDNA by restriction and DNA sequencing

Paul G. Boseley, Anne Tuyns and Max L. Birnstiel

Institut für Molekularbiologie II der Universität Zürich, Winterthurerstr. 266A, 8057 Zürich, Switzerland

Received 20 January 1978

ABSTRACT

The location of the 5.8S rDNA within the internal transcribed spacer has been found by restriction and sequence analysis. These analyses indicate the deletion of a dinucleotide from the known rRNA sequence. Regions to the 5' and 3' of the gene contain both uncommon sequences and palindromic structures which might provide potential control points. A secondary structure model is suggested for the 5.8S rRNA incorporating the flanking sequences.

INTRODUCTION

Xenopus laevis rDNA is located within the nucleolar organising locus and is arranged in several hundred tandemly repeated units. Each unit is known to code for a 40S rRNA precursor molecule which during maturation gives rise to 18S, 5.8S and 28S rRNA's respectively (1-3). The 5.8S rRNA whilst undergoing maturation becomes non-covalently bonded to the 28S rRNA (4). Whether this bonding is prior to the processing event or after, is presently not fully understood (5), but the 5.8S rRNA is found in the mature ribosome in this bound form (6). The approximate location of the 5.8S rRNA within the internal transcribed spacer (ITS) was first indicated by hybridization buoyancy shift experiments (7). This has recently been confirmed in Xenopus laevis by hybridization of purified 5.8S rRNA to cloned rDNA fragments (8), and hybridization to the BamH I-L fragment (Fig.1) (Boseley, unpublished results). Such a general location for the 5.8S rDNA has also been confirmed in the slime mould, Dictostelium discoideum (9) and in mouse DNA (10) in yeast (31) and in Drosophila (32).
Presented below is a detailed restriction map of a cloned rDNA fragment of Xenopus laevis, known to be homogeneous in length (11), and containing nearly 85% coding sequences for 28S rRNA. Using this restriction data the exact location and sequence of the 5.8S rDNA within this fragment has been found. Sequences of potential interest near to the 5' and 3' ends of the gene are discussed.

MATERIALS AND METHODS

Preparation of DNA

Ribosomal DNA was prepared from plasmid pXL212 (pCRI + 28S DNA (14)). Eight colonies were picked from a freshly streaked plate into 8 x 20 ml of L-broth (+ 100µg/ml kanamycin) and grown overnight. These cultures were used to inoculate 8 x 1 L of L-broth (+ antibiotic) and were grown to OD_650 = 1.3 and then amplified for 20 hrs by addition of 170µg/ml Chloramphenicol. Lysis was performed after the method of Katz et al.(18) and the supernatant deproteinised with phenol:chloroform (1:1). After dialysis, precipitation, RNasing (A and T), a phenol:chloroform step and another precipitation the DNA was dissolved in buffer and made up into CsCl/Ethidium bromide gradients (Nd = 1.3890), which contained a ^3H-pCRI marker. These were spun at 35,000 rpm, 23°C, 48 hrs, and the closed circular plasmids collected. After lysis of the Ethidium bromide with iso-amyl alcohol, and the CsCl by dialysis against TE buffer, the plasmids were EcoRI restricted and separated on CsCl/Actinomycin C gradients (Act C:DNA = 0.6:1W/W and Nd = 1.3890; spun at 33,000 rpm, 30°C, for 3 days). The rDNA fragment was easily recovered from the well separated bands. The majority of Act C was removed by shaking with isoamyl alcohol whilst the remainder with phenol:chloroform (1:1) after dialysis. The DNA fragments were stored in TE buffer at -20°C.

5'-^32P labelling of the rDNA fragments

The 5'-^32P-labelling of the rDNA fragments was based mainly on Maxam and Gilbert's sequencing procedure (19). The DNA was treated with bacterial alkaline phosphatase (1.5 units/
Nucleic Acids Research

μl) for 30 min at 37° in 0.5M Tris-HCl pH 8. The enzyme was inactivated by bringing the volume to 500μl with 0.5M NaCl in TE buffer, adding 25μl 2M sodium-TCA, shaking twice with phenol/chloroform (1:1), and twice with chloroform/isoamyl alcohol (24:1). For labelling 5'-protruding ends (e.g. EcoRI, G\ATATTC) the DNA was precipitated, centrifuged and redissolved in 40μl with a final composition of 50mM Tris-HCl pH 7.6, 10mM MgCl₂, 125mM NaCl, 5mM DTT, 1mM spermidine up to 1μM ³²P-ATP and 4 units T4 polynucleotide kinase. For labelling flush ends (e.g. Alu I, ACCT) or for 3' overhangs (e.g. Hha I, GCG\C) the NaCl concentration was reduced to 25mM and the mixture was made 30% DMSO, whilst the remaining reactants were kept as for labelling 5' protrusions (Boseley, unpublished data). Both labelling reactions were carried out at 37° for 30 min and the enzyme inactivated by heating at 65° for 5 min. The volume was increased, bromphenol blue added and the mixture put on a G75 column to remove ³²P ATP. After precipitation and centrifugation the DNA was restricted and electrophoresed on various percent agarose gels (mainly slab but 0.9 x 25 cm tubes for the 28S/BamHI digest). The bands were stained in 1μg/ml ethidium bromide, visualized with long wave UV light, excised, chopped finely and syringed through a 22G needle into 10mM Tris-HCl pH 8 1mM EDTA 0.15M NaCl (at least 10 volumes) and left at 4° overnight. After centrifugation (39,000 rpm, 4°, 30 min) the supernatants were loaded onto a small pre-equilibrated Whatman DE52 column (0.3ml). After washing, the DNA was eluted with 1M sodium acetate pH 8, precipitated, and its purity checked by analytical gels.

Restriction Analysis

Restriction analyses employed a final 50μl volume from which time points of 10μl aliquots were taken at 1, 5, 10, 15 and 30 minutes. The 10μl aliquots were immediately mixed with 3μl of "stop solution" (glycerol, agarose suspension containing Bromphenol blue and enough EDTA to give a 25mM final concentration) and heated at 65° for 5 min. This mixture was loaded onto 1.5% agarose gels (0.2 x 17 x 40 cm) and run in 0.2M glycine, 0.015M NaOH buffer (pH 8.5) at initially 40V and then
at approximately 5V/cm. The gels were prepared for autoradiography as described by Smith and Birnstiel (20). The restriction reactions were optimised by either diluting the enzyme (with 50mM Tris-HCl, 10mM DTT, 100μg/ml BSA, 50% Glycerol) or by adding suitable amounts of sonicated calf thymus DNA to slow the reaction down for the enzyme. All reaction mixtures were preincubated for 5 min at the correct temperature before addition of the enzyme.

Isolation of fragments and their sequence analysis

The 28S fragment was cleaved either with Bam HI (see Fig. 3) or Alu I to obtain preparative amounts of DNA in the region of the 5.8S gene for further restriction and \(^{32}\text{P}\)-labelling. The required labelled fragment was then cleaved asymmetrically and electrophoresed on either 1.5% or 2% agarose gels (22 x 17 x 0.2 cm) in glycine/NaOH buffer. On completion of electrophoresis the gel was Saran-wrapped, \(^{32}\text{P}\)-radioactive spots taped at the corners of the gel and then autoradiographed. Using the resulting autoradiograph the fragments, labelled now only at one end, were cut out of the gel and the DNA eluted as described above.

The method of Maxam and Gilbert (19) was used for the sequence analysis of the labelled fragments. The chemical reactions chosen were G+A, A+C, T+C and C. 20% polyacrylamide 7M urea gels (40 x 17 x 0.1 cm) were used whose slots were filled with 7M deionised urea prior to a 12hr pre-electrophoresis. Electrophoresis was performed at 1000V, whilst loading was done at staggered time intervals to obtain an overlapping series for each fragment. Flash activation of the X-ray films and fast tungstate screens were employed whenever necessary for the enhancement of bands in the autoradiograph (33).

RESULTS AND DISCUSSION

Isolation of rDNA fragments

Isolated Xenopus laevis rDNA may be digested with endonuclease EcoRI to yield a homogeneous fragment of 4.8 kb, which contains about 85% of the 28S gene (called below, 28S fragment) and a second larger heterogeneous fragment (5.7 - 11.3 kb) con-
containing 80% of the 18S gene together with the complete non-transcribed spacer (18S fragment) (11,12). These restriction cuts were utilized by Morrow et al. (11) to clone the rDNA fragments within plasmid pSC101. Starting from one of their clones, pCD4 (pSC101 + complete rDNA repeat) we optimized conditions (13), (for details, (14)) for the transfer of the rDNA fragments into the EcoRI site of the plasmid pCRI (Col El/Kanamycin resistant plasmid) (15,16). As these plasmids can be amplified with chloramphenicol, large quantities (>3mg/L culture) could be made. After restriction of the plasmid pXL212 (containing 28S fragment) with EcoRI the rDNA fragment was separated on CsCl/Actinomycin C gradients (17) which gave excellent separation at high input and DNA in good condition for further enzyme reactions.

Restriction Map

The site mapping procedure (20) is based on the 5'-32P-terminal labelling of a fragment, its asymmetric cleavage and separation by gel electrophoresis. These resulting DNA segments are digested with a chosen enzyme so as to produce a partial digest. A large spectrum of partial digestion products are produced but the labelled fragments form a simple overlapping series, all with a common labelled terminus. These are then fractionated according to molecular weight by gel electrophoresis and detected by autoradiography. The relative mobility of each labelled fragment is compared with that of molecular weight standards which locates the restriction sites' distance in base pairs from the labelled terminus. The order of the fragments and their lengths thus correspond directly to the order of restriction sites along the DNA molecule. This method was chosen because of its sensitivity in measuring DNA fragment lengths. It is also able to clearly pick out repetitive regions (14) giving both the length and number of these repeats. The method is capable of the mapping of several hundred restriction sites within a reasonable amount of time.

A critical difference in the terminal labelling used (see Materials and Methods section) over that of Smith and Birnstiel (20) was that inactivation of the alkaline phosphatase at low pH was avoided. This step was found to cause precipitation of
the DNA and so resulted in inefficient labelling. Syringing of the gel pieces containing rDNA fragments and subsequent retention on a Whatman DE52 column (Clarkson, unpublished results) was preferred over dissolving in potassium iodide and retention on hydroxypatite (20). The yield of DNA was greater and possible risk of phosphate contamination after passage over a desalting column was completely eliminated.

The position of a restriction cut in the DNA was determined from several gels using overlapping fragments, such as B and C (Fig.1). Examples of partial mapping of the B fragment are shown in Fig.2 which are discussed in more detail below. The number of base pairs of each partial fragment from the 5'-end was determined graphically from a semi-logarithmical plot of standard DNA fragments (\(\lambda\)-HindIII, \(\lambda\)-EcoRI/HindIII, PM2-HaeIII (21,22)), sized in kb units versus relative mobility in cm. The overall fragment size was determined by the addition of the B and S fragments (average of \(\gtrsim\)50 determinations) and found to be 4.8 kb.

The gene and spacer regions in Fig.1 have been obtained from values of \(0.22 \times 10^6\) daltons for the distance between ITS and EcoRI restriction cut in the 18S gene (23) and \(0.48 \times 10^6\) daltons for the width of the ITS (24). By deduction the remaining DNA length is concluded to be 28S gene.

The 28S fragment restriction map has many potential uses. It may be used to locate the 3' end of the 18S rDNA or the 5' end of the 28S rDNA by sequencing of the DNA. Also, owing to the conserved nature of rRNA's it may be of use to study evolution in the various eukaryotic rDNA's. In the context of this paper its most interesting application is the prediction of the location of the 5.8S rRNA gene within the ITS using the known rRNA sequence followed by verification by DNA sequencing techniques (19).

The 5.8S rRNA sequence

The 5.8S rDNA sequence of Xenopus laevis (25,26) contains the sequence GACUC at its 5' end, which, when transposed into DNA yields a HindIII restriction site. Examination of the partial HindIII digest of the B-fragment (Fig.2A) shows the presence of a
Fig. 1
Restriction map of the 28S fragment. (Restriction endonucleases Hpa I, HindIII, Fst I, Sal I, Xho I and Xba I do not cut this fragment). The 32P-terminally labelled 28S fragment was cleaved with Bgl II or Bam HI which gave easily separable, yet overlapping fragments (eg B and R) for partial mapping.
A. Partial restriction mapping of the B-fragment (Fig.1) by endonuclease Hinf I. The lowest band in each of the time course reactions is the first site, lying 810 bp from the 5' end, and located within the starting nucleotides of the 5.8S rDNA sequence.

B. Partial restriction mapping of the B-fragment (Fig.1) by endonuclease Hha I.

* indicates the 280 bp gap which contains the complete 5.8S gene whilst clearly indicating the absence of a Hha I cut predicted from the rRNA sequence (25), its complete sequence is shown in Fig.4.)
site 810 bp from the 5'-end and lying within the ITS. The next site, ~1230 bp from the 5'-end, lies within the 28S gene, and therefore cannot be a potential site for the 5.8S gene. Using this information (Fig. 3) a whole series of fragments covering this region were sequenced by the Maxam and Gilbert procedure.

Fig. 3 32P-labelleed fragments used to obtain the 5.8S rDNA sequence using the Maxam and Gilbert procedure (19). The upper portion of the diagram shows the positions of the restriction cuts (as in Fig.1). The fragments asterisked in the lower portion of the diagram indicate the series of overlapping fragments which were sequenced.
Nucleic Acids Research

resulting in the sequence shown in Fig.4. The rDNA sequence is identical to that of Khan and Maden's 5.8S rRNA sequence (25) apart from a GC deletion at position 51 and 52. The presence of this GC would give rise to a Hha I/Hae II restriction site which has not been found (see Fig.2B). This partial mapping gel clearly shows a gap (between the asterisks) of 280 bp which has been used in the sequencing (its actual sequence being -83 to +32 in Fig.4). This GC deletion (\(\triangleleft\)) found in our rDNA sequence comes within a loop in proposed secondary structures for the 5.8S rRNA (25,26) and occurs within a palindrome which is subsequently reduced to one of 14 base pairs (Fig.4). The absence of the GC within our DNA sequence could mean that the 5.8S rRNA coding sequences are polymorphic and that the sequenced cistron represents a rare variant. Alternatively it might be argued that the deletion is an artefact of gene cloning.

The alternative CUC(C)\(_{\text{OH}}\) sequence at the 3' end of the 5.8S RNA (25) can be found in the rDNA, as can the alternative 5' starts pCG and pG (25,26). These alternative start and stop points are thought to reflect non-specific processing or aging events controlled by the degree of base pairing between the 5' and 3' ends of the molecule (26).

The precise nature of rRNA processing is still unknown although three probable methods (or combinations thereof) may exist. There might be a specific cleavage of the precursor RNA in a manner analogous to DNA restriction endonucleases, or a cleavage resulting from an enzyme recognising a special RNA secondary structure, or possibly the cleavage of the RNA after attachment of specific proteins to form RNP particles. On examination of the rDNA sequence (Fig.4) short sequences up to seven nucleotides in length are found to be repeated both within the 5.8S gene, between the gene and flanking sequences and also within the flanking sequences themselves. Their locations however, do not emphasize any strong interrelationships. Furthermore, no extensive homologies are found between sequences at the 5' and 3' ends of the 5.8S rDNA sequence. However, palindromic sequences are found transversing the borders of the gene and flanking se-
Fig. 4
a) The 5.8S rDNA sequence showing the rRNA sequence (25) above
b) palindromes found between the 5.8S gene and flanking sequences and within the flanking sequences. (\( V = \) deletion of base(s)).

sequences (Fig.4). Transcripts of such sequences would allow folding back of the leading and trailing spacer sequences onto the 5.8S rRNA.

Of potentially greater interest are the palindromic sequences which are found in the flanking sequences. The longest of these containing 12 bases follows closely after an unusual \( \text{A}_4\text{G}_2\text{A}_4 \) sequence (see Fig.5, discussed below). As the 5.8S rRNA is transcribed prior to 28S rRNA it might be expected that the pre-5.8S rRNA would assume a secondary structure before processing.
Fig. 5
Autoradiogram of a 20% polyacrylamide 7M urea gel used for the sequence analysis of a 60 bp fragment derived from a HinII digest of fragment C (see Fig. 3). The lettered sequence shows nucleotides -53 to -27 (see Fig. 4) which includes the unusual A<sub>4</sub>G<sub>2</sub>A<sub>4</sub> polynucleotide discussed in the text.
and attachment to the 28S rRNA. Such a likely structure could be drawn (Fig.6) using the large hyphenated palindrome occurring nearly equidistant from both the 5' and 3'-ends of the 5.8S gene (GACC...GCG..GCGA 5' (5.8S rDNA) 3' TCGCC..CGC..GGTC (Fig.6)). According to this model the 5' processing of the 5.8S rRNA occurs within a stem whilst the 3' processing occurs within a loop. Experiments have indicated that the 3'-terminal 20-21 nucleotides, occurring within a stem region in Fig.6, interact with the 28S rRNA (34). As the 5.8S rRNA is known to be transcribed before the 28S rRNA the precursor 5.8S rRNA is very likely to assume a metastable structure prior to processing (such as in Fig.6A). However, it is not yet known whether the 5.8S rRNA is processed before or after its association with the 28S rRNA but the structures indicated in Fig.6 could accommodate either possibility. Thus the 3' end would be situated in a loop in either situation whereas the 5'-end would lie at the bottom of a stem in the metastable structure whilst appearing in a loop when the 5.8S rRNA is associated with 28S rRNA (Fig.6B).

In eukaryotic DNA's the dinucleotide CpG is found at low frequency and is thought to represent some (unknown) signalling function (27). We note that the trinucleotide CCG occurs frequently in the flanking sequence (11 times) whilst only twice in the gene sequence (in the Sma I sites). Equally interesting are the trinucleotides CGC whose locations are almost exclusively within base paired structures when the presumed folding of the RNA is drawn from the DNA sequence (Fig.6).

Recent experiments, aimed at identifying the point of initiation of transcription within Xenopus laevis rDNA (35), employed the vaccinia capping enzyme complex. The results indicated that although 85% of the hybridization was associated with the 5' end of the 40S precursor rRNA the other 15% hybridised to the 28S fragment (see Fig.1). This could mean that a second initiation event is occurring within the transcriptional unit itself. A favourable position for such an initiation event would be within the ITS. One sequence that stands out within the GC-rich ITS spacer is that of A₄G₂A₄ (Fig.5). Comparison of this AG
Fig. 6 Possible models for the pre-5.8S rRNA derived from the DNA sequence (Fig. 3). The 5.8S rRNA secondary structure map of Khan and Maden (25) have been extended to include the extracistronic transcripts. Processing points of the 5.8S rRNA are shown together with the polarity of the molecule.
A) Transcription of the 5.8S rRNA is known to precede the 28S rRNA and such a structure may be found prior to its association with 28S rRNA.
B) Model of the 5.8S rRNA found in association with the 28S rRNA (Models drawn with reference to (36), ▽ = site of GC deletion)
cluster (together with its adjacent sequences) with other eukaryotic prelude sequences has revealed a resemblance to the VA RNA I gene in adenovirus DNA (28). This AG sequence is also found in tDNA\textsuperscript{met} of Xenopus laevis (29) and such sequences have been postulated to be the potential recognition site of RNA polymerase III (F. Müller & S. Clarkson; S. Weissmann, personal communications).

The placement of the 3'-end of the 18S RNA and the 5'-terminus of the 28S RNA within the cloned DNA fragment is not as yet known from sequence data. Thus, the exact location of the 5.8S rDNA relative to the amount of spacer on either side of it is still not clear. Also the significance, if any, of the sequence homology with the presumed RNA polymerase III recognition sequence and the palindromic sequence arrangement in the flanking sequences is yet to be understood. An insight into these unusual features of the 5.8S rDNA sequence may possibly come from experiments involving the techniques of Surrogate Genetics within the germinal vesicle of Xenopus laevis oocytes (30).

ACKNOWLEDGEMENTS

We thank Ham Smith, Stuart Clarkson, Rudolf Portmann and Heinz Daetwyler for helpful advice and Arlette Binkert for technical assistance. Especial thanks to Peter Ford for unpublished results and advice and Walter Schaffner and K. Ruegg for computer programming. E. coli HB101 was donated by M. Grunstein and plasmids pCD4, pCD30 and pCD42 (Morrow et al. (11)) made available to us by J. Hedgpeth. T\textsubscript{4} ligase was a gift from F. Rougeon. Many restriction enzymes were gifts from Ham Smith, Stuart Clarkson, Brigitte Bächli, Vincent Pirrotta, Charles Weissmann, H. Hofstetter and Susanne Stoffel, whilst others were purchased from Miles Labs Inc. and Bio-Labs. T\textsubscript{4} polynucleotide kinase was obtained from PL Biochemicals and bacterial alkaline phosphatase from Worthington. The containment conditions used for the construction of the recombinant plasmids were NIH Guidelines P2 + EK1 and British Guidelines Category II. This work
REFERENCES

3 Reeder, R.H., Higashinakagawa, T. & Miller, O., Jr. (1976) Cell 8, 449-454
26 Ford, P.J. & Matheson, T. (in preparation)
28 Pan, J., Celma, M.L. & Weissman, S.M. (in press)
29 Muller, F. & Clarkson, S.G. (in preparation)