Sequence organization of the spacer in the ribosomal genes of Xenopus clivii and Xenopus borealis

René Bach[1]*, Bernard Allet[2]** and Marco Crippa[1]***

[1]Département de Biologie animale, Université de Genève, Ecole de Médecine, rue de l'Ecole de Médecine 20, and [2]Département de Biologie moleculaire, Université de Genève, 30, quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland

Received 6 July 1981

ABSTRACT

We have studied in X.clivii and X.borealis cloned EcoRI fragments containing the spacer located between the 28S and 18S ribosomal genes. We report for these two species the nucleotide sequences at both ends of the NTS region with special emphasis on the sequences around the transcription initiation site of the 40S rRNA precursor. In X.clivii the location of the 5' end of the precursor was mapped. In both species the sequences around the 40S origin are duplicated in the NTS. Nucleotide sequence comparison has revealed a stretch of 13 identical bases around the transcription initiation site of X.laevis and X.clivii. The same sequence is also present at the presumptive transcription initiation site of X.borealis rDNA.

INTRODUCTION

Ribosomal DNA (rDNA) in Xenopus is constituted of 400-800 tandemly repeated gene units localized at the nucleolar organizer (1,2). During oogenesis, this DNA is amplified to about one million gene copies (3). The X.laevis species is by far the best studied member of this genus. Its ribosomal gene unit has been shown to include sequences to specify one each of 28S, 18S (4,5) and 5.8S rRNA (6,7). In addition, each unit contains a non-transcribed spacer (NTS) located between the end of the 28S coding sequence and a transcribed spacer (ETS), which specifies RNA sequences present on the 5' extremity of the 40S rRNA precursor but absent in the mature rRNA molecules (8,9,10).

Studies by hybridization techniques and restriction enzyme analysis have revealed that sequences coding for the mature cytoplasmic RNA molecules are very conserved in the various Xenopus species, whereas the NTS and ETS have greatly diverged (11,12).

The structure of the transcribed and non transcribed spacers has been studied extensively in X.laevis. It has been shown that the NTS, but not the

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

ETS, is made of highly repetitive DNA (13,14,15,16). Recently, Boseley at al. (15) and Moss et al. (16) have sequenced almost completely the NTS and ETS regions of a cloned ribosomal unit of X.laevis. Their analysis confirmed the repetitive character of the NTS region and raised interesting questions regarding the mechanisms of spacer formation and evolution.

No comparative information is available on the primary DNA sequence of ribosomal genes of different Xenopus species. Such a comparison might reveal conserved DNA sequences which could be important for a better understanding of the regulation of rRNA synthesis. In other gene systems, like the genes transcribed by RNA polymerase II, sequence comparison and experiments on RNA synthesis using artificially modified genes have led to the establishment of interesting sequence-function relationships (17). In an attempt to detect DNA sequences which could play a role in rRNA transcription and might be conserved in different ribosomal genes, we have determined the nucleotide sequence around the initiation site of the 40S rRNA precursor of two additional Xenopus species, X.clivii and X.borealis.

We have established a partial restriction map of cloned EcoRl fragments which contain the spacer region between the 28S and 18S ribosomal genes of X.clivii and X.borealis. In addition, we report for these two species the nucleotide sequences at both ends of the NTS region with special emphasis on the sequences around the rRNA precursor initiation site. In X.clivii, the location of the 5' end of the precursor was mapped within these sequences.

MATERIALS AND METHODS

[γ-^{32}P]ATP (1000-3000 Ci/mmole) and 3'-[α-^{32}P]dATP (1000 Ci/mmole) (cordycepin) were purchased from New England Nuclear.

Polynucleotide kinase and restriction enzymes were from New England Biolabs, and terminal transferase was from P.L.Biochemicals. Ligase and S1 nuclease were purchased from Miles Company. Calf alkaline phosphatase was supplied by Boehringer Mannheim.

Autoradiography was by contact with X-Ray No-Screens films (NS-5T, Kodak) at -20°C or with intensifying screens (Ilford) and preflashed X-Omat XR5 films at -70°C.

X.clivii and X.borealis used in this study were a generous gift from Professor M. Fischberg.

Purification of amplified rDNA from Xenopus ovaries

Ovaries from 150 X.clivii and X.borealis females at stage 65 were dissected, washed two times in Barth's (18) solution and resuspended in 4ml of 50mM Tris pH 7.5, 50mM EDTA. Sarkosyl (Ciba-Geigy, NL30) was added to a final concentration of 1.5% and the suspension was left for 20 min. at room temperature. Pre-digested pronase was added (2 mg/ml) and the mixture was dialyzed overnight at 37°C against 50mM Tris pH 7.5, 50mM EDTA. Pancreatic RNAase (pre-boiled for 10 min.) was added at a concentration of 200µg/ml. After 2 hours at 37°C, the sample was homogenized slightly, brought to 2xSSC (1xSSC is 0.15M NaCl, 0.015M sodium citrate) in 0.1M Tris pH 9, and extracted with phenol equilibrated in 1xSSC. The aqueous phase was ethanol precipitated twice, and the final pellet was dissolved in 4.5ml of 10mM Tris pH 7.5. The amplified ribosomal DNA was then purified by centrifugation in CsCl gradient as described by Dawid et al. (3)

Construction of hybrid plasmids

For each species, three micrograms of rDNA were mixed with 0.5µg of plasmid DNA (pBR313 or pBR322), digested with EcoRl in a ligase buffer containing 30mM Tris pH 8, 4mM $MgCl_2$, 1mM EDTA, 0.2mM ATP, 10mM DTT and 50µg/ml BSA. After heat inactivation of EcoRl (10 min. at 63°C) ligase was added and the mixture was kept overnight at 14°C.

The E.coli strain HB101 was transformed with the ligated sample using the technique described by Mandel and Higa (19). Tetracycline-resistant clones were analyzed by the method of Grunstein and Hogness (20) with the use of ^{32}P-labelled ribosomal 18S and 28S RNA as probes (21). DNA from positive clones was further characterized by restriction analysis with EcoRl and HindIII. The plasmids were propagated and purified by the technique of Katz, Kingsburg and Helinski (22) in a P2 facility according to the NIH guidelines.

Purification from X.clivii of total ovarian RNA containing the 40S rRNA precursor

Ovaries from one X.clivii or X.borealis female were homogenized in 0.1 M NaOAC pH 5, 0.5% SDS and 40µg/ml PVS (polyvinyl-sulfate) and extracted 5 times with phenol, saturated with the same buffer. After phenol-chloroform (1:1) and chloroform extraction, RNA was precipitated with ethanol. The RNA was then passed over a G-75 Sephadex column and reprecipitated with ethanol.

DNA mapping

Mapping of restriction sites in the DNA inserts was performed by the technique of partials described by Smith and Birnstiel (23). The protocol was basically the same as that described in a previous study (24) except that the labelled sites used were either the HindIII site close to the left-hand end of the insert (Figure 1) or the EcoRl site at the righthand end.

DNA labelling and sequencing

Restriction fragments were purified as described previously (25). BglI sites were labelled by the terminal transferase technique as described by Maxam and Gilbert (26) except that $[\alpha-^{32}P]$ ATP was replaced by ^{32}P-labelled cordycepin (100µCi).

Labelled DNA fragments were either cut with a second restriction enzyme or strand-separated on neutral 6% and 8% acrylamide gels (1/60 crosslinked) and run at 200 V at 4°C as described by Sakano et al. (27).

The sequencing technique of Maxam and Gilbert (26) as modified by Allet (25) was used throughout this work.

Hybridization and S1 digestion

DNA fragments to be hybridized with 40S precursor RNA were labelled at both 5' ends and strand-separated as described above. Aliquots of both strands were sequenced and the coding strand was then mixed with about 200µg of total ovarian RNA, in 80% formamide, 0.4M NaCl, 40mM MOPS pH 6.7, 1mM EDTA, in a final volume of 50µl. After 10 min. at 60°C hybridization was allowed to proceed for 5 hrs at 49°C. The hybrids were then diluted 10 fold into S1 buffer (50mM NaOAc pH 4.5, 200mM NaCl, 1mM Zn_4Cl_2). Samples incubated either with 10^5 or $5x10^5$ units of S1 for one hour at 45°C were extracted with phenol-chloroform and chloroform-isoamylalcohol and ethanol precipitated. After lyophilization, the pellet was dissolved in denaturing loading buffer (26), heated and analyzed in a 8% acrylamide sequencing gel in parallel with the sequence ladder of the previously sequenced coding strand.

RESULTS

Highly enriched rDNA from X.clivii and X.borealis was digested with EcoRl and subsequently cloned in either pBR322 or pBR313 as described in Materials and Methods. Analysis of positive recombinants revealed two types

of inserts for both Xenopus species : one small fragment of about 5000 bp which contains the end of the 18S coding sequence, the ITS and the main portion of the 28S sequence and several large fragments ranging from 6000 bp to 9000 bp containing the end of the 28S coding region, the NTS and the major part of the 18S coding region. The size variation observed for the large fragment is due to length heterogeneity in the NTS (see Figure 1). Since our main interest was focused on the determination of possible regulatory sequences preceeding the transcription initiation site of rRNA, we analyzed in more detail two of the large EcoRl fragments from X.borealis (pXB8, pXB3), and one fragment from X.clivii (pXCl). Figure 1 shows the general structure of two tandem repeat units of X.laevis ribosomal genes and compares, to the same scale, the large EcoRl fragments from X.laevis, X.borealis and X.clivii.

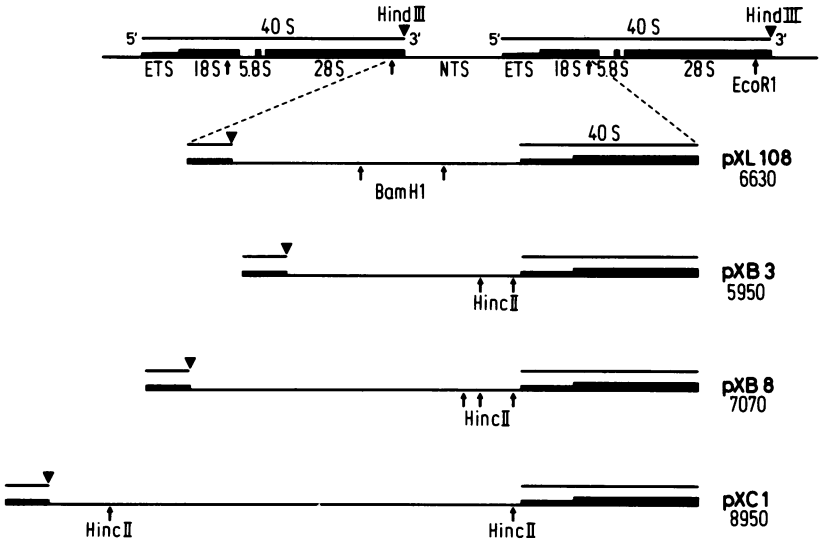

Figure 1 - Structural organization of the Xenopus rDNA

Two repeating units of X.laevis rDNA are shown. The lower part of the Figure represents an expansion of the cloned EcoRl fragments used in this study. pXL108 is the X.laevis plasmid used by Boseley at al. (15). pXB3 and pXB8 are two plasmids of X.borealis and pXCl is a plasmid of X.clivii.The number under the plasmid symbols indicates the size in basepairs.

A) X.CLIVII

General structure of the NTS

The rDNA fragment from X.clivii has a size of 8950 bp as measured by electron microscopy. The overall restriction pattern (Figure 2) does not show any similarities to the corresponding fragment of X.laevis (12). However, when subfragments of the NTS were analyzed in more detail by partial digestion with additional restriction enzymes, a distinct pattern of repetitiveness which has also been found in the NTS of X.laevis (15) could be observed (lower part of Figure 2).

Cleavage of pXCl with BglI generates several small restriction fragments in

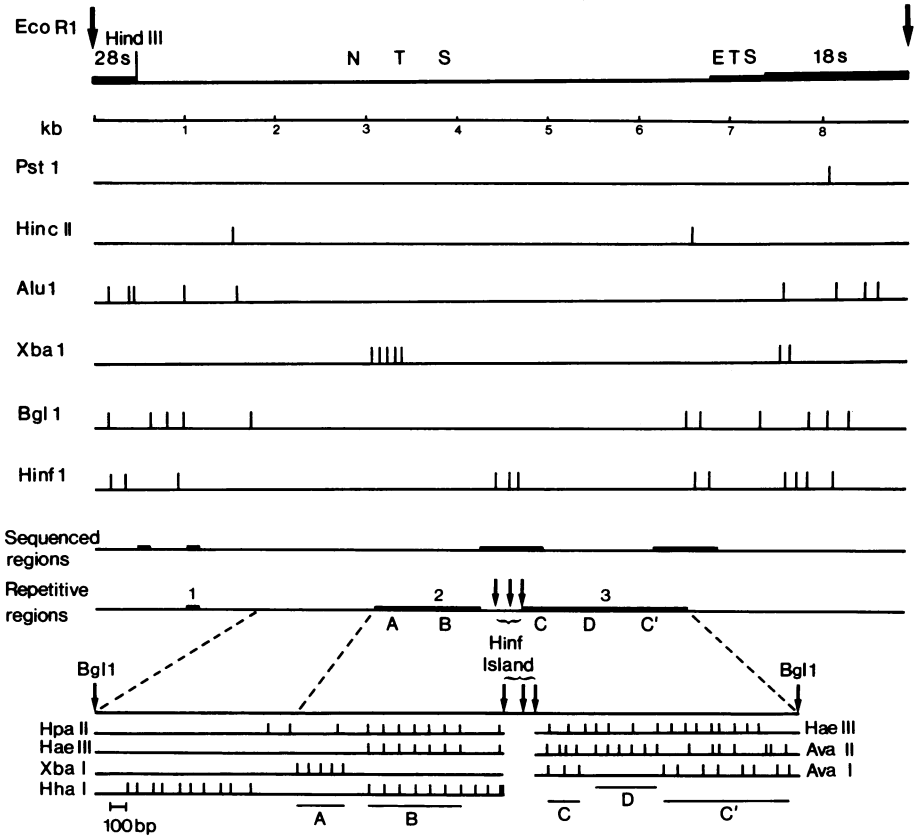

Figure 2 - Restriction map of the fragment of X.clivii rDNA cloned in pXCl

The position and the detailed restriction map of the repetitive regions are indicated.

addition to one large fragment of approximately 5 Kb which covers most of the NTS and was analyzed in more detail (see bottom of Figure 2). Roughly in the middle of the fragment, there are three Hinf restriction sites which give rise to two fragments of 135 bp and 86 bp. We will refer to this region containing these two fragments as Hinf Island in analogy to the Bam Islands described by Boseley et al. (15) (see below). The region located to the left of the Hinf Island is composed of at least two repetitive canons. Repetitive region 2A is regularly cleaved five times by XbaI and repetitive region 2B is cut seven times by HpaII, HhaI and HaeIII, approximately every 100 nucleotides. Whether or not an additional repetitive region is located to the left of canons 2A and 2B could not be determined unambiguously. The observed HhaI cutting pattern is not as regular when compared to region 2A and 2B and might just reflect the presence of GC clusters in this particular DNA segment.

The DNA region flanking the right of the Hinf Island is composed of three kinds of repetitive regions : 3C, 3D and 3C'. Region 3C and 3C' show a high degree of sequence homology (shown in Figure 3) and are composed of four and eleven repeat units, respectively. The observed irregularity in the AvaI, AvaII and HaeIII restriction patterns are obviously due to minor sequence variations at the enzyme recognition sites. Region 3D was deduced from the regular AvaII cutting pattern; preliminary sequence data indicate that the repeat units in 3D are different from those in 3C and 3C'. An additional repetitive canon (region 1) was found during sequence analysis of the DNA segment flanking the end of the 28S coding region. It contains four repeat units of 27 bp which are almost homologous. The primary sequence of some of the repetitive canons are shown in Figure 3.

```
          GANTCGGGTGCCGGCCCCGGGGCCACTTGCACTGTCCCGCGCCCTGCCCCGGGGGCAAGAGGGCCTCGCCGGG
                                CGCCCCAGC-TCCT--CGGGCTAAGTCC
1                               CGCCC-ACCCTCTCCCCGGCTTAAGTCC
                                CGGCCCACCCTCTCCC-GGCTTAAGTCC
                                CGGCCCACCCTCTCCCCGGCTTAAGTCC

    CCGCCC-GTGCCCGACCGGGAGTTCCATGAGCCTGGGTGTGGGAAGCCGGCTTTCCCCCTGTCCCTCCGGGGGACCGCGG
3C  CCGCCCTCTGCCCGACCGAGAGTTCCAGGAGCCTGGGTGTGGGAAGCCGGCTTTCCCCCTGTCCCTCCGGGGGACCCGCTG
    CCGCCC-CTGCCCGACCGAGAGTTC-AGGA-CCCGGGTGTGG--AGCAGG-TTTCCCGCCTC---
                                                                ----GGACCCGGCCG
3C' CCGCCC-CTGCCCGACCGGGAGTTC-AGCA-CCCGGGTGTGGGAAGCAGGTTTTCCCCCTGTCCCTCCGGGGGGACCAGGCGT
    CCGCCC-GTGCCCGACCGAGCGTTCCAGGAGCCCGGGCAAGCGAAGCAGGCTTTCCCCCTGTCCCTCCGGGGGGGACGAGGCGTCCGCCC
```

Figure 3 - Nucleotide sequence of different repetitive regions
The nucleotide sequences of the canons found in repetitve regions 1, 3C and 3C' are presented. Gaps, indicated by (-), have been introduced to maximize homology.

Location of the 40S precursor RNA starting site

Moss and Birnstiel (28) have mapped the 5' end of 40S rRNA of X.laevis 2200 nucleotides upstream from EcoRl site located in the 18S rRNA gene. If one assumes a similar transcription pattern in X.clivii, the NTS-ETS boundary should be located close to the HincII site which is present in the rightward portion of pXCl. We began therefore to sequence on both sides of this restriction site. The result is shown in Figure 4. The sequence represents the noncoding strand and is shown in the 5' to 3' direction. To the right of the HincII site, the DNA consists of non-repetitive sequences. 79 bp to the left starts repetitive region 3C, (Figure 2) of which two complete repeating units have been sequenced. Approximately 200 nucleotides downstream from the HincII site the sequence 5'-CAGGAAGGTAGGG-3' is found. This sequence is identical to the sequence around the 40S initiation site of X.laevis (28,29).

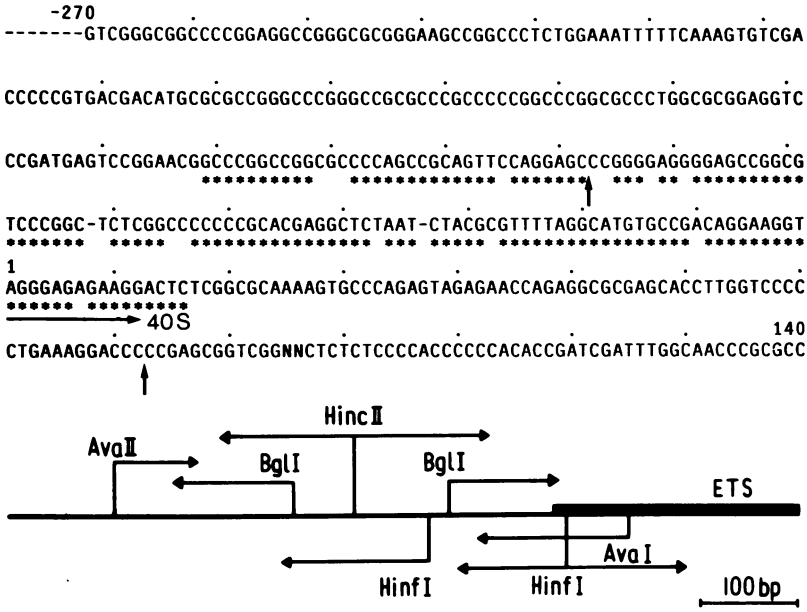

Figure 4 - Nucleotide sequence of the region surrounding the 40S transcription initiation site of X.clivii

The strategy employed to determine the nucleotide sequence is shown. The arrows indicate the origin, direction and length of each sequence obtained. The vertical arrows (↑) delimitate the AvaI fragment used for S1 protection experiments. Nucleotides identical to those found in the Hinf Island are marked by*.Gaps, indicated by (-), have been introduced to maximize homology.

To demonstrate that this sequence also represents the rRNA transcription initiation site in X.clivii, we designed an S1 protection experiment using total homologous ovarian RNA and an AvaI fragment which overlaps the presumed 40S rRNA starting site and is indicated by the two arrows in Figure 4. The DNA fragment was end-labelled, strand separated and the coding strand hybridized to RNA. After S1 digestion, the size of the protected fragment was determined on a sequencing gel in parallel with the corresponding sequence ladder. Figure 5 shows that the 3' end of the protected fragment extends (from the right) to the T of the 3'CATCCC-5' sequence (i.e. to the A of the 5'GTAGGG-3'). This result as well as the apparent sequence homology with the region surrounding the transcription initiation site of X.laevis (28,29) led us to conclude that in X.clivii transcription starts indeed at the determined sequence, even though we did not formally prove that the RNA used in the S1 protection experiment represents the primary transcript.

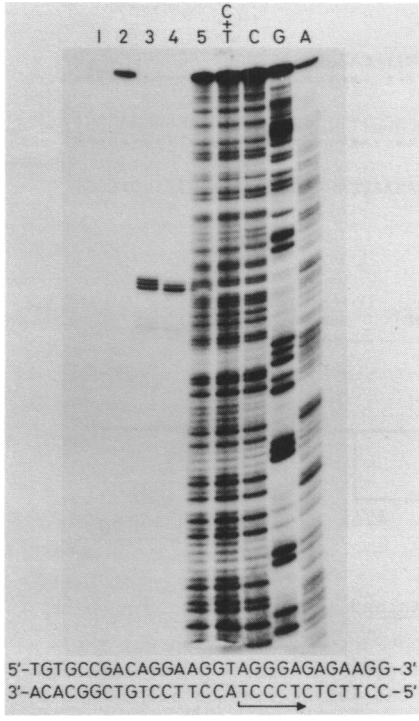

Figure 5 - S1 mapping of the 40S rRNA transcription initiation site

The transcribed (-) strand of the AvaI fragment labelled at the 5' end was hybridized to the 40S pre rRNA. The hybrids were digested with nuclease S1 and run in parallel with a sequence ladder of the same fragment.
1) (-) strand digested with 10^5u of S1
2) (-) strand untreated
3) (-) strand hybridized with RNA and treated with 10^5u of S1
4) (-) strand hybridized with RNA and treated with 5.10^5u of S1
5) same as in 4 but run together with the C+T sequence

C+T, C, G and A : sequence ladder of the (-) strand. The reading of the sequence around the 40S transcription initiation site is shown at the bottom of the Figure. The arrow indicates the 40S rRNA origin.

The 40S origin is repeated

To test whether, like in X.laevis, the sequence around the 40S starting site is repeated in the NTS, we made the following experiment. pXCl DNA digested with BglI and/or HinfI (Figure 2) was electrophoresed in a 1% agarose gel and transferred to Millipore filter (30). The transferred DNA was hybridized with the labelled BglI-HinfI fragment which overlaps the 5' end of the 40S rRNA (Figure 4). The result showed that the probe hybridizes to the large BglI fragment, but that none of the two larger HinfI fragments flanking the HinfI Island gives a positive signal (data not shown). This result strongly suggests that a "promoter duplication"(28) exists and that it occurs within the HinfI Island. To analyze in more detail this duplication, we sequenced the entire HinfI Island. Figure 6 shows that the 135 bp HinfI fragment is a nearly identical copy of the sequence just upstream from the 40S starting point.

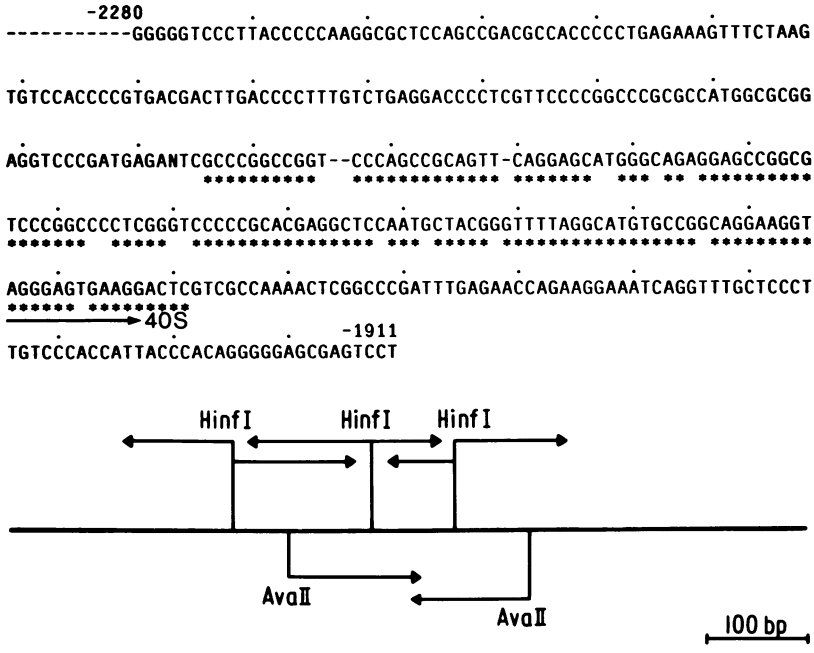

Figure 6 - Nucleotide sequence of the HinfI Island

The strategy employed to determine the nucleotide sequence is shown. The arrows indicate the origine, direction and length of each sequence obtained. Nucleotides identical to those found around the 40S initiation are marked by*. Gaps, indicated by (-), have been introduced to maximize homology.

The 86 bp Hinf fragment shows only little homology with the ETS. This homology progressively disappears as the boundary of the Hinf Island close to repetitive region 3C (Figure 2) is reached.

B) X.BOREALIS

General structure of the NTS

We have sequenced the NTS boundaries of the two different large EcoRl fragments from X.borealis rDNA, pXB3 and pXB8. The sequence data obtained for pXB8 are presented in Figure 7. Analysis of these data shows that DNA segments of various sizes and base sequences are repeated and interspersed in a rather complex pattern as diagrammed in Figure 8. In the sequenced region, pXB3 differs from pXB8 only by a few base substitutions. However, pXB3 is 209 bp shorter than pXB8 and misses one HincII site and three repeated canons of 118, 47 and 44 bp, respectively.

It is interesting to note that the 44 bp repeat in pXB8 which shows a high degree of homology to the 40S starting region in X.laevis (15) and in X.clivii (see preceeding section) is found 5 times in the part of the DNA that was sequenced (4 times in pXB3). The 44 bp repeats can be grouped into two categories as shown at the bottom of Figure 8. One group consists of region III* and V* and contains a stretch of 13 nucleotides which are identical to those found at the 40S RNA precursor origin in X.laevis and X.clivii. Another group is made of three repeats (two in pBX3) which are all identical and which contain a C instead of a T in front of the 40S initiation sequence AGGG. The sequences on the right of repeat V* have no repetitive character and most likely represent the ETS region.

Location of the 40S precursor RNA starting site

The sequence we have obtained in X.borealis extends over two BglI fragments (see bottom of Figure 7). Since the separation of the fragments was more convenient in pXB3, this clone has been used to map the transcription initiation point.

In pXB3 the left BglI fragment is 499 bp and the right fragment is 440 bp long. Like for X.clivii we designed an S1 protection experiment to map the 40S RNA origin using both BglI fragments. The results (data not shown) indicate that the 440 bp BglI fragment is completely protected whereas with the 499 bp fragment only minimal protection ($\sim$ 10% of the input counts) was

```
                1
 74 GGGCCGGCCCCGGGGCGGGCGAGCCAGCCCGGCCCCGGAGTTCTCGGAGCGC--GGGCACGGCGCCTCC--GCTCCCC
 44 GGGCCCGCTCCGGCAGGAAGGCAGGGGCGAGGCCCTCCTCTCCC
 47 GGGACCGGAGGCGCAGCCGGGGGCGCGATGTCAACACCACTGCTCGG
118 AAGTCCCGATGAGGACGGATTCACCCCGGCCCGGCCAGGCCGGAGTACAAGGCGCCCGG
    GCGAGCGGAGCGGCGGCTTGGCACTCGCGGGCACGCACGCACGGGGCCTCCCCTCCCCC
 44 GGGCCCGCTCCGGCAGGAAGGCAGGGGCGAGGCCCTCCTCTCCC
 47 GGGAACGGAGGCGCAGCCGGGGGCGCGATGTCAACACCACTGCTCGG
118 AAGTCCCGATGAGGACGGATTCACCCCGGCCCGGCCAGGCCGGAGTACAAGGCGCCCGG
    GCGAGCGGAGCGGCGGCTTGGCACTCGCGGGCACGCACGCACGGGGCCTCCCCTCCCCC
 46 GGGTACTGCTCCGGCAGGAAGGTAGGGACTGAGGTACTAATCACCC
  6 TGCACC
 61 GGCCGCCGCCTCCGCCCGGCCCGGGGTTCCAAGAGCTCGGGCAGGGGAAGCCGGTGC-CCCC---
 18 TTTCCCTCGGAGGTCCGA
 65 GGCAGCCGCTTCCGCCCGACCCGGGGTTCCAAGAGCTCGGGCAGGGGAAGCCGGTCCTCCCCCGG
 16 AAGTCCCGACGAGGAC
 78 GGGCCGGCCCCGGGGCGGCGGAGCCAGCCCGGCCCCGGAGCTCTCGGGGCGCGCGGGCACGGCGCCTCCGCGCTCCCC
 44 GGGCCCGCTCCGGCAGGAAGGCAGGGGCGAGGCCCTCCTCTCCC
 46 GGGAACGGAGGCGCAGCCGGGGCGCGACGTTGA-CACCACCGCTCGT
 44 GGGCACGCTCCGGCAGGAAGGTAGGGACGAGGTCCTCCTCACCC
    CGCCAGGAGCGCAGCCCAACCCCAACCCCGGCCCCGGCCCTGGCCCCGACCCTAGCCCGGGTTAGCGCCCGGCTCCTC
    CGACCCCACCCCCGGCCCGGTCGCCGAGGCCAGGGGCCGGGCCCTCCTCTCCCGCGGGCGGAAATTTCCGCAGACTGT
    CGCGCCGACGCCGACTTGCACGGAGGGGCCCCGTTCTCCTCCCCACCCCCCGGGCCCCCGAGCCCCCGCGGAGGTCCC
                  1176
    GATGAGGACGGATTTGCCNNNNNGGC
```

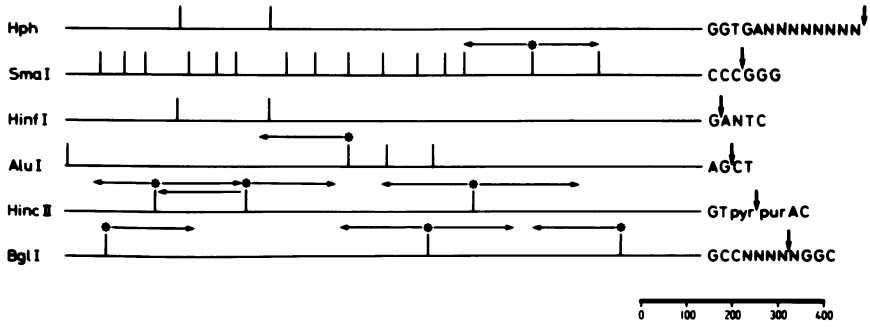

Figure 7

obtained.

The low amount of protection could be due to a short unstable hybrid; however, the fact that the protected counts disappeared under denaturing conditions makes the interpretation more difficult. The same experiment was done using other restriction fragments containing the 44-46 bp type sequence, as well as the corresponding BglI fragment from pXB8. In no case were we able to observe a protected fragment of a discrete size on denaturing gels. No conclusion is therefore possible as to the location of the 40S origin in X.borealis. It is even conceivable that - despite the sequence homology with the other species - none of the 44-46 bp canons shown in Figure 8 is actually used to initiate transcription.

C. INTERSPECIES COMPARISON

General structure of the DNA region preceeding the 40S initiation site

Figure 9 shows the general sequence arrangement for pXL108 and pXC1 of the region extending from the Bam and Hinf Islands respectively, to the NTS-40S boundary as well as the sequence arrangement of the corresponding region of pXB8. This figure shows to the same scale the various repeating blocks which constitute most of this part of the spacer in the three species. The starting point of the 40S RNA has been precisely mapped for X.laevis and X.clivii whereas it is still uncertain for X.borealis. It can be seen that the general structure of pXL108 and pXC1 is quite similar whereas pXB8 shows a much more complex arrangement. In X.laevis and X.clivii the sequence corresponding to the 40S starting point is duplicated in the NTS. In pXC1 one such duplication has been detected while in pXL108 two have been described(15). Assuming that the 40S initiation site in X.borealis lies whithin a 44 bp repeat, five duplications would be present, all clustered in a DNA segment of approximately 1000 nucleotides.

Figure 7 - Nucleotide sequence of the NTS region surrounding the 40S RNA initiation site in X.borealis

The sequence of the NTS-ETS boundary in pXB8 is represented. The strategy employed to determine the nucleotide sequence is shown. The arrows indicate the origin, direction and length of each sequence obtained. The breakdown of the sequence into different lines has been done by computer in order to emphasize repetitive pattern. The numbers on the left side of the sequence indicate the number of nucleotides present in each line.

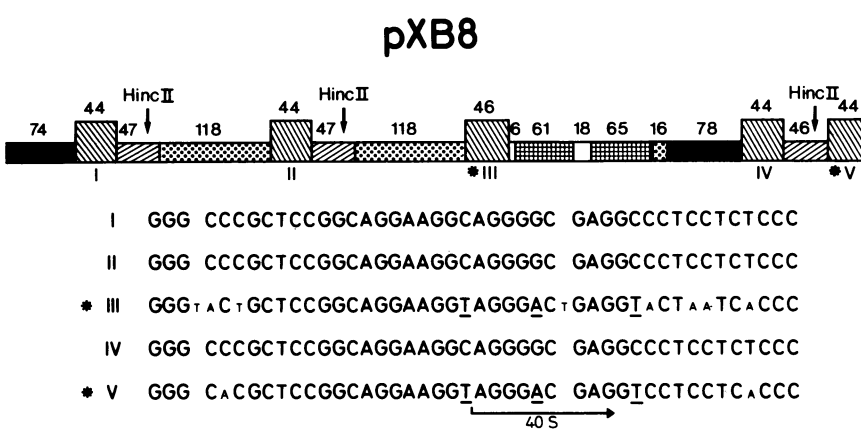

Figure 8 - Structural organization of the nucleotide sequence of the NTS region surrounding the 40S RNA initiation site in X.borealis

Regions of homologous sequences in pXB8 are drawn with the same symbols. White areas indicate unique sequences. The number on top of each region indicates the number of nucleotides present. The nucleotide sequences of the five 44-46 bp repeats are compared in the lower part of the Figure. The arrow indicates the position of the corresponding transcription initiation site of the 40S RNA in X.laevis and X.clivii.

Sequence comparison

Figure 10 shows the results of a computer search for homology in X.laevis, X.clivii and X.borealis. The X.laevis sequence established by Boseley at al (15) has been used as a basis for comparison. The sequences have been aligned using the 40S RNA initiation point as reference. For X.borealis a presumptive 40S rRNA initiation point located in the III repeat of Figure 8 has been used, chosen on the basis of maximum homology. In the same perspective a (-) sign has been introduced in the sequence in order to maximize homology.

Using this alignment, a stretch of 13 identical nucleotides around the 40S initiation point is detected between pXL108 and pXC1 (Figure 11).

The same is true for pXB8 if one assumes that the 40S rRNA is initiated at the same sequence in X.borealis as in the other two species. Another region of homology starts at around position -70 and extends to position -150. Minor mismatches are however present as it is indicated in Figure 10.

When the sequences present in the repetitive regions are analyzed for

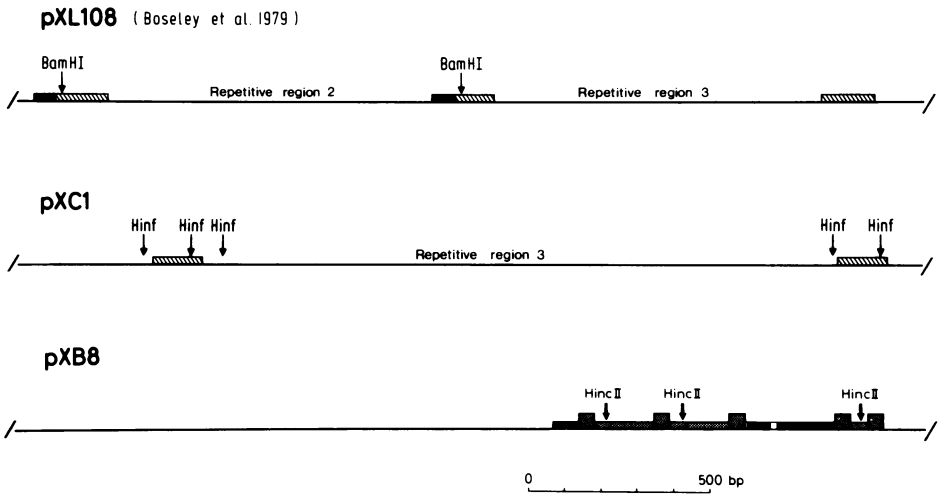

Figure 9 - Comparison of the sequence organization of the NTS in X.laevis, X.clivii and X.borealis

The general structure of the part of the NTS that we studied in X.clivii and X.borealis is compared to the data available for X.laevis (15). For each species the arrangement of the repeating blocks is represented. The arrow (→) indicates the origin of the 40S. For X.borealis, since the origin of the 40S is uncertain, the alignment has been arbitrarily chosen.

homology, the following results are observed :
A stretch of eight nucleotides GGCTTAAG is shared between X.laevis and X.clivii in the canons of repetitive region 0 and 1, respectively. Another common stretch of fourteen nucleotides GGAGTTCCAGGAGC can be observed in the canons of repetitive regions 3C (in pXC1) and of repetitive region 2 (in pXL108). The same sequence, with minor mismatches, is detected in the repetitive blocks 74 and 78 of the X.borealis (pXB8).

Sequence similarities at the boundaries of the NTS region

A striking feature emerges upon examination of the sequences very near the boundaries of the NTS region. The NTS-40S boundary was determined by mapping the 40S rRNA initiation point. For X.borealis the 40S initiation point was arbitrarily chosen on the basis of maximum homology. Since the 3' end of the 28S RNA in X.clivii and X.borealis was not directly mapped, the 28S-NTS boundary was established by analogy with the situation observed in X.laevis (29).

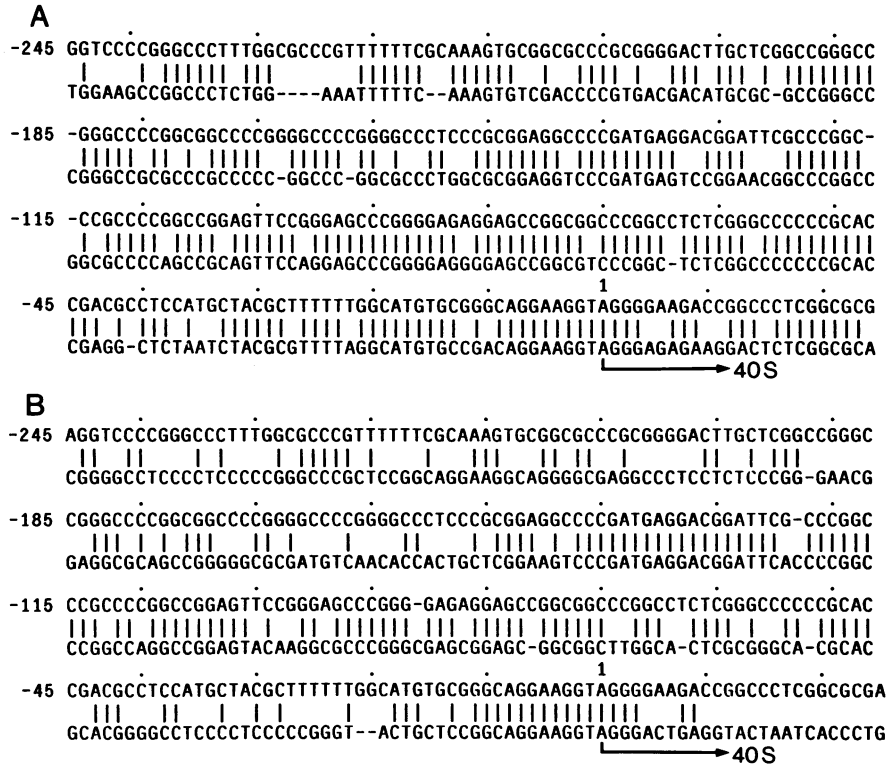

Figure 10 - Interspecies sequence comparison in the region surrounding the 40S transcription initiation site

A - comparison between X.laevis and X.clivii
B - comparison between X.laevis and X.borealis
The X.laevis sequence (upper line) is from Boseley et al (15). Gaps, indicated by (-), have been introduced to maximize homology. The coordinates refer to the first nucleotide of the 40S rRNA precursor.

Figure 11 - Sequence homology around the 40S transcription initiation site

The 13 identical bases around the 40S transcription initiation site are indicated. The heavy line corresponds to the beginning of the 40S sequence.

Figure 12 shows the results of a computer search for homology between the 28S-NTS boundary and the NTS-40S boundary. The results clearly show a striking sequence homology at the two ends of the NTS region.

For each individual species the probability that the observed homologies were occuring by chance is rather low (from 5% to 9%). It is obvious that this probability must become even lower since this homology occurs in three species which widely differ in the overall nucleotide sequence of their NTS and ETS regions.

DISCUSSION

As already pointed out for X.laevis by Boseley et al (15), the sequence arrangement found in the NTS of the two additional Xenopus species presented here implies that a series of different mechanisms like duplication, unequal crossing over and mutation must have come into action during evolution. However, the nucleotide arrangement is much too complex to enable us to deduce, like it was done for the satellite DNA (30), in which order the different events might have occured. Yet it seems clear from our data that the appearance of

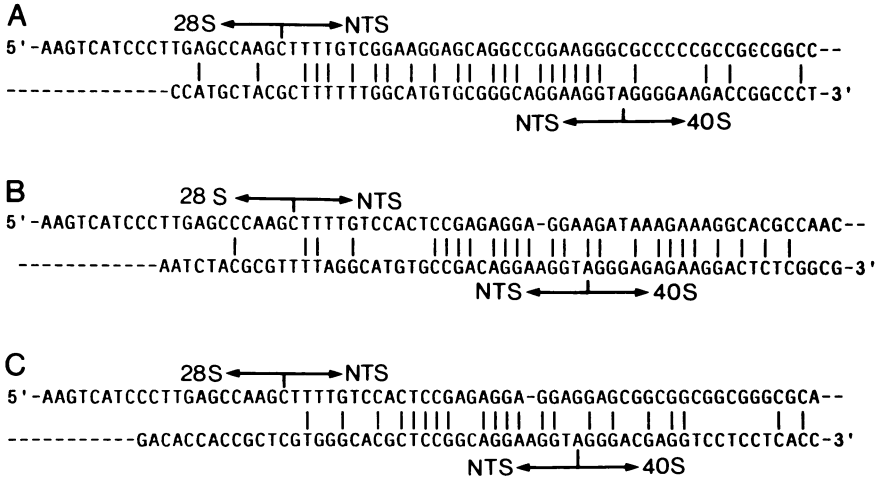

Figure 12 - Sequence comparison at the 5' and 3' NTS boundaries

The nucleotide sequences across the 2 ends of the NTS are aligned in order to detect homologous bases. A - X.laevis; B - X.clivii; C - X.borealis. For X.laevis, the data of Boseley et al (15) and Sollner-Webb et al (29) have been used.

sequence differences between species (speciation)has preceeded the horizontal expansion of the spacer sequences. Comparative analysis of the NTS from the two Xenopus species described here as well as from several others (P.A.Briand, C.Reymond, F.Stutz and M.Crippa, in preparation) has never revealed a sequence composition indicative of a mixed "interspecies" arrangement.

If sequence conservation throughout evolution reflects functional significance, one would conclude that the common nucleotides found at the 40S initiation point as well as the conserved sequences observed at position -70/-150 in the NTS should play some role in the regulation of transcription of the Xenopus ribosomal genes. The elucidation of this point will have to await the availability of a suitable transcription system with which "surrogate genetics" (31) experiments could be performed. However, one should also stress that though sequence data on the ribosomal genes of different systems are still limited, no sequence homology has been found at the transcription initiation site between evolutionary very distant genera such as Saccharomyces, Drosophila, Xenopus and Mus (32,33).

The presence of "promoter duplications" (28) which was first discovered in the NTS of the X.laevis ribosomal genes and has now been shown to exist in X.clivii and perhaps X.borealis seems to be a general feature of the Xenopus ribosomal unit. The strong homology between these duplications and the actual transcription initiation site, argues in favor of a possible functional role for these DNA regions. Whether or not they actually represent an advantage for the system can only be a subject for speculation at the present time.

Interestingly, it has been reported that some of these duplicated regions are indeed used to initiate transcription (34; T.Moss, personal communication). Bakken at al (35) have furthermore suggested that in X.laevis a strong transcription termination signal should exist upstream from the normal 40S rRNA initiation point. Transcripts which might have initiated in the spacer would then be prevented from going through the 40S coding region. The sequences found in front of the 40S initiation point, homologous to those present at the 28S-NTS boundary,could indeed represent such a signal.

ACKNOWLEDGEMENTS

We wish to thank Dr. I. Sures for very helpful suggestions during the preparation of the text and Prof. M. Fischberg for making his collection of Xenopus species available to us.

We are indebted to Ms H. Achermann and Mr. P.A. Briand for skillful technical assistance; to Mr. Freddy Bourquin and Mr. O. Jenny for the graphic work and to Ms M. Weiss for the careful typing of the manuscript.

This work has been supported by the grants of the Fonds National Suisse pour la recherche scientifique to MC and BA.

* present address : Department of Computer Sciences, Stanford University Stanford, California, U.S.A.

** present address : Biogène S.A., 3, route de Troinex, Carouge, Genève

*** to whom correspondence should be addressed

REFERENCES

1. Wallace, H. and Birnstiel, M.L. (1966) Biophys. Biochem. Acta 114, 296-310
2. Brown, D.D. and Weber, C.S. (1968) J. Mol. Biol. 34, 681-697
3. Dawid, I.B., Brown, D.D. and Reeder, R.H. (1970) J. Mol. Biol. 51, 341-360
4. Birnstiel, M.L., Speirs, J., Purdam, I., Jones, K. and Loehning, U.E. (1968) Nature 219, 454-463
5. Speirs, J. and Birnstiel, M. (1974) J. Mol. Biol. 87, 237-256
6. Wensink, P.C. and Brown, D.D. (1971) J. Mol. Biol. 60, 235-247
7. Boseley, P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nuc. Acids Res. 5, 1121-1137
8. Wellauer, P.K., Reeder, R.H., Carrol, D., Brown, D.D., Deuch, A., Higashinakagawa, T. and Dawid, I.B. (1974) Proc. Natl Acad. Sci. USA 71, 2823-2827
9. Brown, D.D., Wensink, P.C. and Jordan, E. (1971) Proc. Natl Acad. Sci. USA 68, 3175-3179
10. Wellauer, P.K. and Reeder, R.H. (1975) J. Mol. Biol. 94, 151-161
11. Botchan, P., Reeder, R. and Dawid, I.B.(1977) Cell 11, 599-607
12. Fedoroff, N.V. (1979) Cell 16, 697-710
13. Wellauer, P.K., Dawid, I.B., Brown, D.D. and Reeder, R.H. (1976a) J. Mol. Biol. 105, 461-486
14. Wellauer, P.K., Reeder, R.H., Dawid, I.B. and Brown, D.D. (1976b) J. Mol. Biol. 105, 461-486
15. Boseley, P.G., Moss, T., Mächler, M., Portmann, R. and Birnstiel, M.L. (1979) Cell 17, 19-31
16. Moss, T., Boseley, P.G. and Birnstiel, M.L. (1980) Nuc. Acids Res. 8, 467-485
17. Flavell, R.A. (1980) Nature 418, 358-375
18. Barth, L.G. and Barth, L.J. (1959) J. Embryol. Exptl. Morphol. 7, 210-217
19. Mandel, M. and Higa, A. (1970) J. Mol. Biol. 53, 159-162
20. Grunstein, M. and Hogness, D. (1975) Proc. Natl Acad. Sci. USA 72, 3961-3965
21. Rungger, D. and Crippa, M. (1977b) Exptl. Cell. Res. 107, 227-237
22. Katz, L., Kingsbury, D.T. and Helinski, D.R. (1973) J. Bacteriol. 114, 577-591
23. Smith, H.O. and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398
24. Allet, B. and Rochaix, J.-D. (1979) Cell 18, 55-60
25. Allet, B, (1979) Cell 16, 123-129

26. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl Acad. Sci. USA 74, 560-564
27. Sakano, H., Hüppi, K., Heinrich, G. and Tonegawa, S. (1979) Nature 280, 288-294
28. Moss, T. and Birnstiel, M.L. (1979) Nuc. Acids Res. 6, 3733-3743
29. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 19, 485-498
30. Southern, E.M. (1975) J. Mol. Biol. 94, 51-69
31. Kressmann, A. and Birnstiel, M.L. (1980) in: Transfer of Cell Constituents into Eukaryotic Cells, Eds. Celis, J.E., Grässmann, A. and Loyter, A. (Plenum Press) 31, 383-407
32. Long, E.O., Rebbert, M.L. and Dawid, I.B. (1981) Proc. Natl Acad. Sci. USA 78, 1513-1517
33. Bach, R., Grummt, I. and Allet, B. (1981) Nucl. Acids Res. 9, 1559-1569
34. Rungger, D., Achermann, H. and Crippa, M. (1979) Proc. Natl Acad. Sci. USA 76, 3957-3961
35. Bakken, A., Morgan, G., Sollner-Webb, B., Roan, J., Busby, S. and Reeder, R.H. (1981) Proc. Natl Acad. Sci. USA (in press)