Volume 11 Number 22 1983 Nucleic Acids Research

Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*

Vassie C.Ware, Brian W.Tague*, C.Graham Clark**, Richard L.Gourse+, Reindert C.Brand++ and Susan A.Gerbi

Division of Biology and Medicine, Brown University, Providence, RI 02912, USA

Received 29 August 1983; Revised and Accepted 19 October 1983

ABSTRACT

We have determined the complete nucleotide sequence of Xenopus laevis 28S rDNA (4110 bp). In order to locate evolutionarily conserved regions within rDNA, we compared the Xenopus 28S sequence to homologous rDNA sequences from yeast, Physarum, and E. coli. Numerous regions of sequence homology are dispersed throughout the entire length of rDNA from all four organisms. These conserved regions have a higher A+T base composition than the remainder of the rDNA. The Xenopus 28S rDNA has nine major areas of sequence inserted when compared to E. coli 23S rDNA. The total base composition of these inserts in Xenopus is 83% G+C, and is generally responsible for the high (66%) G+C content of Xenopus 28S rDNA as a whole. Although the length of the inserted sequences varies, the inserts are found in the same relative positions in yeast 26S, Physarum 26S, and Xenopus 28S rDNAs. In one insert there are 25 bases completely conserved between the various eukaryotes, suggesting that this area is important for eukaryotic ribosomes. The other inserts differ in sequence between species and may or may not play a functional role.

INTRODUCTION

RNA comprises a major portion of the mass of prokaryotic and eukaryotic ribosomes; yet the function of most ribosomal RNA (rRNA) is unknown. Certain regions within rRNA are likely to be involved in interactions with ribosomal proteins, other RNAs, or cofactors necessary for protein synthesis. The idea that RNAs were the only components of the original protein synthetic machinery and that proteins were a subsequent addition may have merit (1-3). If this is the case, then it seems plausible that RNA-RNA associations may be at the heart of the mechanism of protein synthesis, with ribosomal proteins assuming structural and/or functional roles later in evolution.

To identify areas within rRNA that may be involved in specific interactions necessary for ribosome structure or function, we have undertaken a comparative sequence analysis to locate evolutionarily conserved regions within rDNA; one would expect natural selection to favor the retention of rRNA gene sequences that maintain effective ribosomal function.

The existence of evolutionarily conserved regions within rDNA was first

© IRL Press Limited, Oxford, England.

detected by heterologous hybridization experiments (4) using rRNA from Xenopus laevis and DNA from a variety of other eukaryotes. Later experiments showed that this heterologous hybridization was due to a conserved subset of sequences within rDNA (5, 6), and these conserved regions were subsequently mapped by Southern blot hybridization (7, 8). Remarkably, three regions were found to be conserved between Xenopus and E. coli rDNA (7). The conserved nucleotide sequences between such diverse organisms should identify the most likely candidates for functionally important regions.

Only with direct sequence analysis and comparisons of sequence data from various species could the conserved regions be mapped unambiguously and rRNA structure be explored in detail. Xenopus rDNA was the first eukaryotic gene to be cloned (9) and much sequence data are already available: the external transcribed spacer (10, 11), 18S rDNA (12), internal transcribed spacers with the 5.8S rDNA (13, 14) and the nontranscribed spacer (15-17). We have completed the entire nucleotide sequence of Xenopus laevis 28S rDNA. In addition, we present evolutionary comparisons made by aligning nucleotide sequences so as to maximize homology between the Xenopus laevis 28S sequence and other complete non-organellar rDNA sequences which have been published: E. coli 23S (18), yeast 26S (19, 20) and Physarum polycephalum 26S (21).

MATERIALS AND METHODS

DNA

Xenopus laevis 28S rDNA was prepared as described by Brand and Gerbi (22) from plasmids containing portions of the rDNA repeat. 90% of the 28S rDNA is in clone pXlrl1, and the remainder is in clone pXlrl2 (Fig. 1). The construction of these Col E1 clones has been described previously (23). Subclones in pBR322 were provided by Drs. B.E.H. Maden and R. Reeder; subclones M3 and R20 contain rl1 fragments C+E and A+D, respectively (Fig. 1). NIH Guidelines were followed for recombinant DNA work.

Enzymes And Nucleotides

Restriction enzymes were purchased from New England Biolabs, BRL, or NEN; some Eco RI was kindly provided by Dr. H. Bäumlein. E. coli DNA polymerase I (large fragment) and terminal transferase were obtained from NEN. T4 polynucleotide kinase was from NEN and P-L Biochemicals, Inc. DNase I was from Worthington. The following nucleotides were used: $^{32}P\gamma$ -ATP (NEN: average specific activity, 3000 Ci/mmol or the less expensive and more effective ICN crude preparation: average specific activity, 9000 Ci/mmol) for 5' end kinase labeling; 3'-dATP [α-^{32}P] from the 3' end labeling kit from

NEN; deoxyNTPs and dideoxyNTPs (P-L Biochemicals, Inc.).

DNA Sequencing

5' or 3' labeled ends were usually separated by secondary restriction digestion. Occasionally strand separation (following alkaline denaturation of the DNA at room temperature) on neutral polyacrylamide gels was used to obtain singly labeled fragments.

Primarily the chemical modification/cleavage method of Maxam and Gilbert (24), with five reactions and standard gel conditions, was used (25). Occasionally a rapid enzymatic technique utilizing DNA polymerase, DNase I, deoxyNTPs, and dideoxyNTPs was employed as outlined by Seif et al. (26) with the following modifications: 0.5 µl of DNase I (10^{-2} µg/ml) was added to the 5 µl reaction mixture (previous experiments had shown that our DNA polymerase lacked sufficient nicking activity); 0.5 µl of DNA polymerase I (5 U/µl) was sufficient to complete the reactions; reactions were carried out in 1.5 ml Eppendorf tubes instead of capillary tubes. "Forward" and "Backward" reactions for each particular nucleotide were usually combined and run in one slot on the sequencing gel. The enzymatic method could only be used for double stranded fragments having one labeled end as strand separated fragments lack a template for polymerase to fill in on the opposite strand. Generally sequence information started about 15 bases from the 5' labeled end. With the Maxam and Gilbert method up to 15 bases were often illegible from the 3' labeled end, in contrast to the legibility from 5' labeled ends.

At times, gel conditions were varied to cope with band compression on one strand of the DNA. To maximize DNA fragment denaturation, high gel temperatures were generated using 1X or 1.5X Peacock's buffer (27) rather than 0.5X Peacock's Buffer (J. Leong, personal communication). This method was usually adequate to resolve compressed areas on gels. In addition, xylene cyanol/formamide (0.1% w/v) dye mixtures instead of xylene cyanol/urea (0.1% in 7M w/v) helped to relieve the compression problem (28).

Computer Analysis

The interactive version (29) of the Queen and Korn program (30) was used for restriction site analysis and sequence alignments.

RESULTS

Strategy For Sequence Determination

A detailed restriction map of Xenopus laevis rDNA was published previously (13, 15); additional sites were mapped in the current study by the method of Smith and Birnstiel (31) or by double digestion. Portions of the

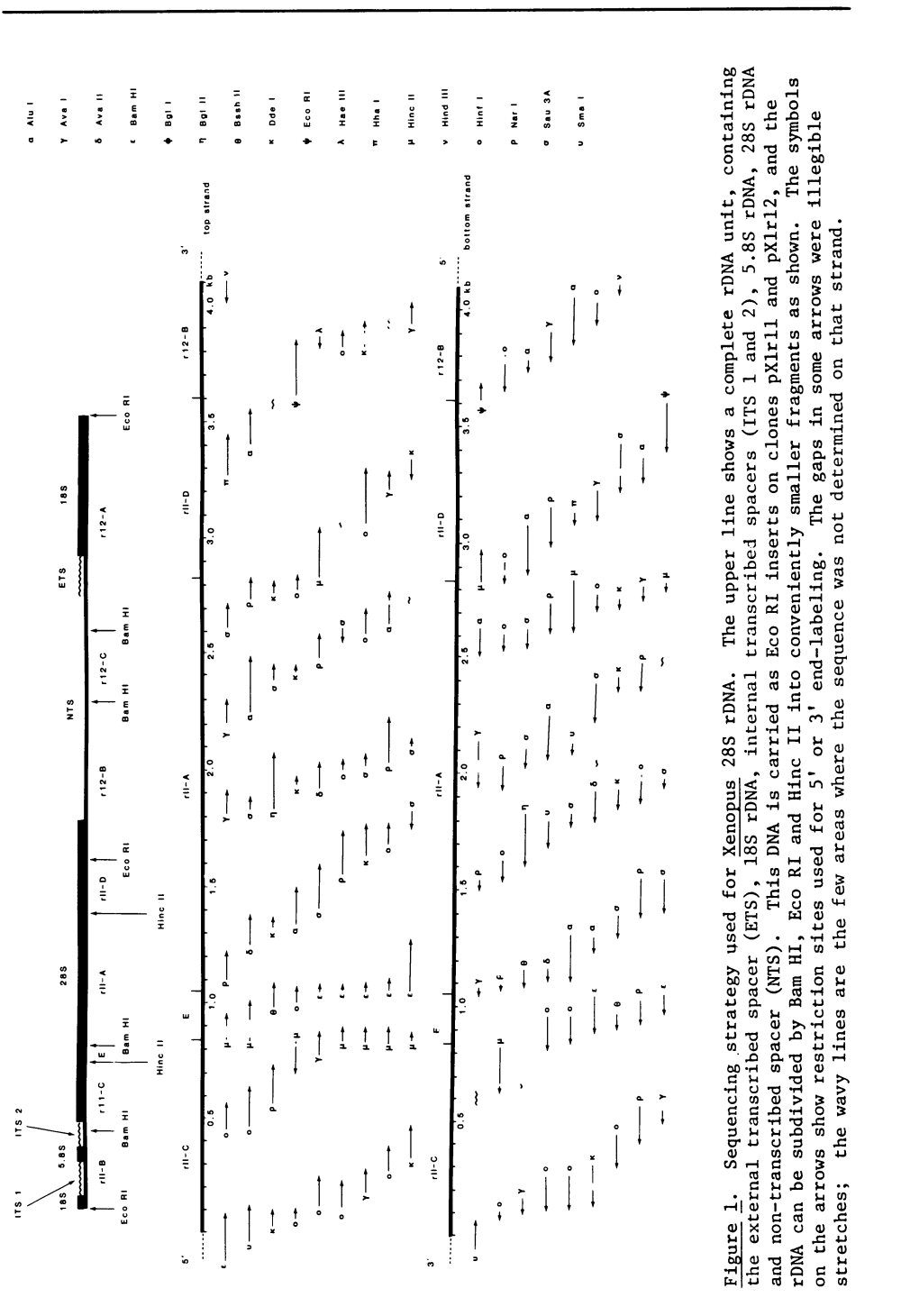


Figure 1. Sequencing strategy used for Xenopus 28S rDNA. The upper line shows a complete rDNA unit, containing the external transcribed spacer (ETS), 18S rDNA, internal transcribed spacers (ITS 1 and 2), 5.8S rDNA, 28S rDNA and non-transcribed spacer (NTS). This DNA is carried as Eco RI inserts on clones pXlrl1 and pXlrl2, and the rDNA can be subdivided by Bam HI, Eco RI and Hinc II into conveniently smaller fragments as shown. The symbols on the arrows show restriction sites used for 5' or 3' end-labeling. The gaps in some arrows were illegible stretches; the wavy lines are the few areas where the sequence was not determined on that strand.

Xenopus laevis 28S rDNA had been determined previously: the 5' terminus (14), parts of fragment r11-D (see Fig. 1; 25, 32), and the 3' terminus (16).

Our sequencing strategy is shown in Fig. 1. Using this approach, we were able to generate sequence data on both strands for 95% of the length and to overlap restriction enzyme sites. In the few areas where the sequence was obtained on only one strand, experiments were repeated several times to confirm the single stranded sequence. As a measure of the quality of our data, we compared a computer-generated restriction map resulting from our sequence data with the experimentally-derived restriction map. Overall, only about 1% ambiguity remains in the sequence.

Features Of The 28S rDNA Sequence

The sequence of the gene for Xenopus laevis 28S rRNA is 4110 base pairs long. Our calculated G + C content for Xenopus 28S rRNA is 66%, which is slightly greater than the previously determined values of 63% (33) or 65% (34). This value correlates well with the observation that there is an increase in the G+C content especially of the large rRNA of the large subunit in plants and animals of the higher taxa (35, 36). Due to the high G+C nature of Xenopus 28S rDNA and secondary structure effects resulting in extreme band compression on sequencing gels, certain areas of the sequence were difficult to determine. In one instance, ten G's appeared as one band. However, with the use of special conditions (see Methods), one could reduce or eliminate the compression problem. A representative sequencing gel is depicted in Fig. 2; note the G+C richness of the sequence.

There are a few minor differences in nucleotides in our DNA sequence (Fig. 3) compared to the partial sequence of another rDNA clone previously published (16). Nucleotide polymorphism between rDNA repeat units within a single individual most likely accounts for these differences.

Comparison Of Xenopus 28S rDNA With Other Homologous rDNAs

The nucleotide sequences of Xenopus laevis 28S, yeast 26S, Physarum 26S,and E. coli 23S rDNAs were aligned to maximize homologies in primary structure. Alignment of the nucleotides was accomplished by matching sequences at least 12 bases long with greater than 75% homology (Fig. 3). In addition, previously published secondary structure models for yeast (19) and E. coli (40-42) were taken into account for the primary sequence alignments. Table 1 shows the overall degree of homology for each species relative to Xenopus rDNA resulting from the alignments. As expected the extent of homology with Xenopus rDNA is greater for the eukaryotes than for E. coli. It is noteworthy, however, that E. coli 23S rDNA is 47% homologous to Xenopus 28S

Figure 2. Sequencing gel showing the separation of Xenopus 28S rDNA bases 2696 - 2737 (bottom strand). This stretch of sequence emphasizes the predominance of G's and C's in the overall sequence: here there are 28 C's, 11 G's, 2 T's, and 1 A. Standard gel conditions, except with 1X Peacock's buffer, were used.

rDNA; this degree of homology is far greater than was previously detected using fairly stringent hybridization conditions (7).

Specifically, fourteen areas of high homology between all four species were identified (Fig. 4). These conserved segments are distributed throughout the entire length of the rDNAs, although the majority are clustered within the first and last thirds of the rRNA genes. Compared to the bulk of the rDNA, the conserved regions have a higher A + T base composition (Table 1). There appears to be a particular bias in sequence conservation in these areas. Of the 933 bases conserved in all four species, 64.4% are purines; in particular the dinucleotide AA and the trinucleotides GAA and AAA are over-represented in conserved regions (as are GAAA, AAAA, and GAAAA). Not all poly-purine tracts are conserved since, for example, AGA is under-represented. Localization of

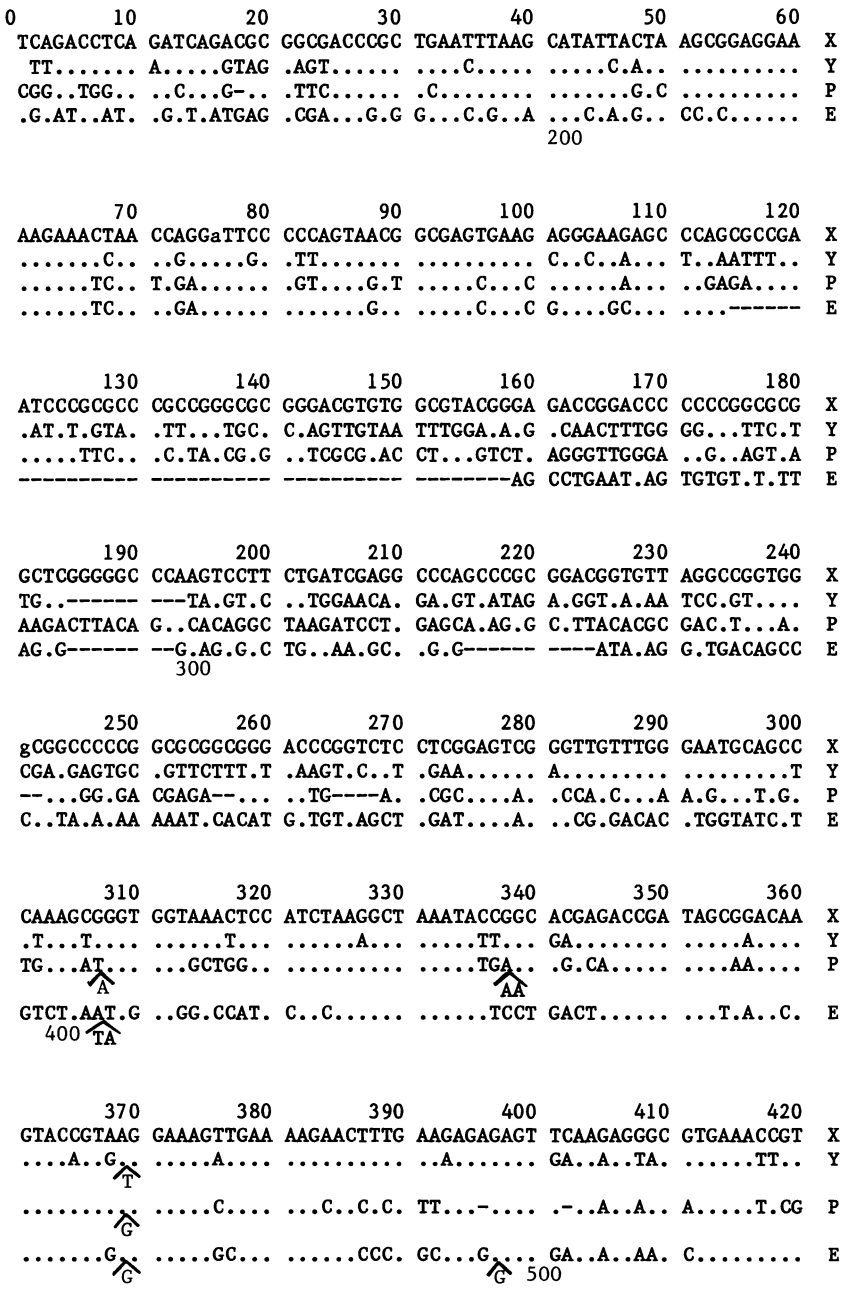

0 10 20 30 40 50 60
TCAGACCTCA GATCAGACGC GGCGACCCGC TGAATTTAAG CATATTACTA AGCGGAGGAA X
TT..... A.....GTAG .AGT..... ....C..... .....C.A.. ..... Y
CGG..TGG.. ..C...G-.. .TTC..... .C..... .....G.C ..... P
.G.AT..AT. .G.T.ATGAG .CGA...G.G G...C.G..A ...C.A.G.. CC.C..... E
200
70 80 90 100 110 120
AAGAAACTAA CCAGGaTTCC CCCAGTAACG GCGAGTGAAG AGGGAAGAGC CCAGCGCCGA X
.....C.. ..G.....G. .TT..... ..... C..C..A... T..AATTT.. Y
.....TC.. T.GA..... .GT....G.T .....C...C .....A... ..GAGA.... P
.....TC.. ..GA..... .....G.. .....C...C G....GC... ....------ E
130 140 150 160 170 180
ATCCCGCGCC CGCCGGGCGC GGGACGTGTG GCGTACGGGA GACCGGACCC CCCCGGCGCG X
.AT.T.GTA. .TT...TGC. C.AGTTGTAA TTTGGA.A.G .CAACTTTGG GG...TTC.T Y
.....TTC.. .C.TA.CG.G ..TCGCG.AC CT...GTCT. AGGGTTGGGA ..G..AGT.A P
---------- ---------- ---------- --------AG CCTGAAT.AG TGTGT.T.TT E
190 200 210 220 230 240
GCTCGGGGGC CCAAGTCCTT CTGATCGAGG CCCAGCCCGC GGACGGTGTT AGGCCGGTGG X
TG..------ ---TA.GT.C ..TGGAACA. GA.GT.ATAG A.GGT.A.AA TCC.GT.... Y
AAGACTTACA G..CACAGGC TAAGATCCT. GAGCA.AG.G C.TTACACGC GAC.T...A. P
AG.G------ --G.AG.G.C TG..AA.GC. .G.G------ ----ATA.AG G.TGACAGCC E
300
250 260 270 280 290 300
gCGGCCCCCG GCGCGGCGGG ACCCGGTCTC CTCGGAGTCG GGTTGTTTGG GAATGCAGCC X
CGA.GAGTGC .GTTCTTT.T .AAGT.C..T .GAA..... A..... .....T Y
--...GG.GA CGAGA--... ..TG----A. .CGC....A. .CCA.C...A A.G...T.G. P
C..TA.A.AA AAAT.CACAT G.TGT.AGCT .GAT....A. ..CG.GACAC .TGGTATC.T E
310 320 330 340 350 360
CAAAGCGGGT GGTAAACTCC ATCTAAGGCT AAATACCGGC ACGAGACCGA TAGCGGACAA X
.T...T.... .....T... .....A... .....TT... GA..... .....A.... Y
TG...AT... ...GCTGG.. ....TGA.. .G.CA..... ....AA.... P
A AA
GTCT.AAT.G ..GG.CCAT. C..C..... .....TCCT GACT..... ...T.A..C. E
400 TA
370 380 390 400 410 420
GTACCGTAAG GAAAGTTGAA AAGAACTTTG AAGAGAGAGT TCAAGAGGGC GTGAAACCGT X
....A..G.. .....A.... ....A..... GA..A..TA. .....TT.. Y
T
.....G.. .....C.... ...C..C.C. TT...-.... .-..A..A.. A.....T.CG P
G
.....G.. .....GC... .....CCC. GC...G.... GA..A..AA. C..... E
G G 500

Figure 3

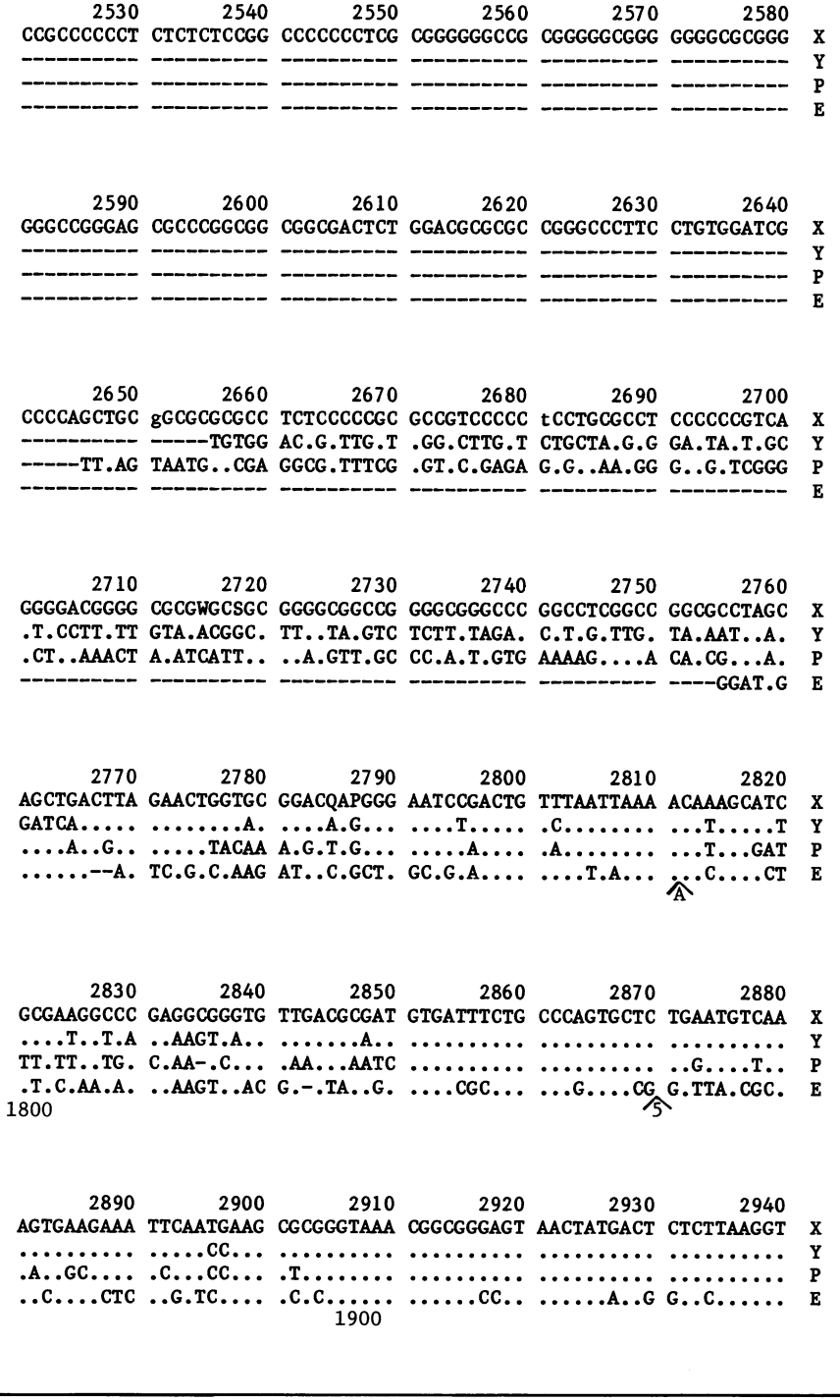
```
       430         440         450         460         470         480
TAAGAGGTAA ACGGGTGGGG CCGTGCGGTC CGCCCGGAGG ATTCAACCcG GCGGGTCAGC  X
.G.A...G.. GG.CA.TT.A T.AGA.ATGG T.TTTT.TTC TCG..TTT.A CT...C....  Y
                                      1
CCCA.T.AG. ....TAAATC TATACG.C.. GT.GAAAGAC GGCGCTGAGT A...CA....  P
                                       2
GT.CGTAC.. G.A.TG..A. .ACGCTTAGG ..TGT.ACAC C..TTGTGAT AAT.......  E
                                      TGCGT

       490         500         510         520         530         540
GCCGCCGGGA CCGGGCCACT CGGCGGACCC CCCCGCCGGC CCCCTCCCCT CGCCGGGAGG  X
-ATCAGTTTT GGT...AGGA TAAATCCATA GGAATGTA.. TTG.CT.GG. AAGTATT.TA  Y
TG..G.T.AG G..ACG.TGG GAAA..TGA. .TAACGGCCA AT.TCG.TTA ..-G.C....  P
.A--------- ---------- ---------- ---------- ---------- ----------  E

       550         560         570         580         590         600
GCGGTCGCGC GGGGGGGACG CGGCCCGGGC GSCCCGGCGC CSPSPGCGST TTCCTCCGCG  X
..CTGT.G.A ATACT.CCA. .T.GGACT.A .GA.T.CGA. GTAAGT.AAG GATGCTG..A  Y
A.....AA.G ACACC...AC ..CGGGACCT CT..GTCTAG GCTTACACTC ACG.GTGAAT  P
---------- ---------- ---------- ---------- ---------- ----------  E

       610         620         630         640         650         660
GCGGTGCGCG CGCCGGCTCC GGGCCGCGTG GGAAGGCCCA GGGGTTCAGG GGAAGGTGGC  X
TAATG.TTAT ATG------- ---------- ---------- ---------- ----------  Y
.G.T.CGC.C .ATAAC.GTG A.A.GTACA. C.TGTCAGAG CTTAG.TT-- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E

       670         680         690         700         710         720
CGGCCGCCCC CGCGCGCGGC TACAGCCCCC CCCCAPGCAG CAGCACTCGC CGTCGCCCGG  X
---------- ---------- ---------- ---------- ---------- ----------  Y
---------- ---------- ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E

       730         740         750         760         770         780
GGCCGAGGGA GACGcCgGCC TCCGCGGtCC TCCTCCCCGG AGCGCGTCCC gCCGCTCCCC  X
---------- ---------- ---------- ---------- ---------- ----------  Y
---------- ---------- ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E

       790         800         810         820         830         840
CCCGGGGGGG CGGCGCGCGG GGCGGGGAAG GGGGAAGGGG CCCCCCGCTC CCGGCGCGGC  X
---------- ---------- ---------- ---------- ---------- ----------  Y
---------- ---------- ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E
```

```
      850         860         870         880         890         900
TGTCAACCGG GGCGGACTGC CCCCAGTGCG CCCCGTCCGC GCCGCGCCGC CGAGGCGGGA  X
---------- ---------- ---------- ---------- ---------- ----------  Y
---------- ---------- ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E

      910         920         930         940         950         960
GGGCCGCCGG GAGCCGCCRA GGGTCCGCGG CGATGTCGGT GTCCCACCCG ACCCGTCTTG  X
---------- ---------- ---------- ---------- -------... -.........  Y
---------- ---------- ---------- ---------- -------A.C ..........  P
---------- ---------- ---------- ---------- ---------- ----------  E

      970         980         990        1000        1010        1020
AAACACGGAC CAAGGAGTCT AACGCGCGCG CGAGTCGGAG GGACTCTGCG CGAAACCCTG  X
.......... .......... ....TCTAT. .....GTTT. ..T------. TA......AT  Y
.......... .......... ....TA.AT. .A.....A.C A.CT.TCA.. .TG.GA.GGC  P
---------- ------C.TA T.TT.TGTA. .A..GTTA.C C--------- -...TAGGG.  E
                            600

     1030        1040        1050        1060        1070        1080
TGGCGCAATG AAGGTGAGGG CCGGGGCGCC CCGGCTGAGG TGGGATCCCG CCGCCCCTCC  X
AC...T.... ..A....AC. TA..TT---- GG...CTC.C AA.AGGTG.A .AAT.GAC.G  Y
GAAGCT.... C.AAAAGCAC ..A...---. GA.C..AT.. C-...C.... .AATG.TCT.  P
A.C..A.GG. ..ACC...TC TTAACTG.G. ---------- ---------- ----------  E

     1090        1100        1110        1120        1130        1140
CTCCGCCCCC CCGGGGGCGG GGGGGggGCG CCGGCGGGCS CACCACCGGC CCGTCTCGCC  X
A...TGATGT .TTC------ ---------- ---------- ---------- ----------  Y
..AAC.GGG. T.TC------ ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E

     1150        1160        1170        1180        1190        1200
CGCCCCGTCG GGGRGGTGGN GCGTGAGCGC GCGCGATTAG GACCCGAAAG ATGGTGAACT  X
--------G. AT.GAT.T.A .TAA....-A TA..TG..G. .......... ..........  Y
-------.GGC TT.GAT...A .TT.....AT .TCATG.... .T.......A .C.AC..G..  P
---------- ---------- -GT.A..TTG CA.GG-.ATA .........C CC.....T..  E

     1210        1220        1230        1240        1250        1260
ATGcCTGGGC AGWGCGAAGC CAGAGGAAAC TCTGGTGGAG GTCCGTAGCG GTCCTGACGT  X
.......AAT ..G.T..... .......... .......... .CT....... ..T.......  Y
....T..A.T ..GC...... ...GA....T CT.......A .GT.....A ..A.......  P
                                                  C
.GC.A..... ..GTT....G TT.G.T..CA CTAAC..... .A...A.C.. ACTAAT-GT.  E
          700
```

```
      1270        1280        1290        1300        1310        1320
GCAAATCGGT CGTCCGACCT PGGTATAGGG GCGAAAGACT AATCGAACCA TCTAGTAGCT  X
........A. ....GA.TT. G......... .......... .......... ..........  Y
........T. T...AA..T. GA........ .........C ......GT.G .T........  P
.A...ATTAG ..GAT...T. GT.GC.G... .T.....G.C ....A....G GGAGA.....  E
                                                        800

      1330        1340        1350        1360        1370        1380
GGTTCCCTCC GAAGTTTCCC TCAGGATAGC TGGCGCTSGT CCGTCGCAGT TTTATCCGGT  X
......TG.. .......... .......... A.AA...C.. AT----.... .....GA...  Y
....T..A.. .......... .......... AAAG.AAAAA GT....-... A.GGGG....  P
.....T.C.. ...AGC.ATT .-...-.... GCCTCG.GAA TT-----..TC .CCGGG-...  E

      1390        1400        1410        1420        1430        1440
AAAGCGAATG ATTAGAGGTC TTGGGGCCGA AATCGATCTC AACCTATTCT CAAACTTTAA  X
.......... .........T CC....T... ...-..C..T G......... ..........  Y
....AC.... .....GA.-. ACC....G.T TTGACCGT.. GG.TC..... ........T.  P
.G...AC-.. T..C.GCA-A GG...T.ATC CCGACT.AC. ....CGA.G- .....-----  E
                                          900

      1450        1460        1470        1480        1490        1500
ATGGGTAAGA AGCCCGGCTC GCTGGCTTGG AGCCgGGGCG TGGAATGCGN VGCACGCCAT  X
..AT...... ..T..TTG.T A..TAA...A .CGT..ACAT .T.....AAG A..TTT---.  Y
...CTCT.-. ....ATTGA. CAC.CG.C.T ...A.CTCT. G.C.G.ATCG G.TCTC.TT.  P
                                    GCTGG
-..C.-..T. ---------- ---------- -------C.. GA.....TTA T-....GG.G  E

      1510        1520        1530        1540        1550        1560
AGTGGGCCAC TTTTGGTAAG CAGAACTGGC GCTGCGGGAT GAACCGAACG CCGGGTTAAG  X
.........T .......... .......... .A........ .......... TA.A......  Y
.......... G-........ ..AGGA.... AAAAA...T. C..A.TTGT. ..C.....-.  P
.CACA-.GG. GGG..C...C GTCCGTC.TG AAGAG..A.A C....C.GAC .GCA.C....  E
                                                        C 1000

      1570        1580        1590        1600        1610        1620
GCGCCCGATG CCGACGCTCA TCAGACCCCA GAAAAGGTGT TGGTTGATAT AGACAGCAGG  X
.T...G..AT A.-....... ......A... C.......... .A...C..C. .......C..  Y
.TT.T...A. A.-.TTGG.. -AT....A.T C.......A. ..CC.C..GA T......T..  P
.T-...A.A. T.-.T.G.T. --..TGG--- ....C.A... G..AA.GCCC ..........  E
                                                        C

      1630        1640        1650        1660        1670        1680
ACGGTGGCCA TGGAAGTCGG AATCCGCTAA GGAGTGTGTA ACAACTCACC TGCCGAATCA  X
.......... .......... .......... .......... .......... G.......G.  Y
..A....... ......C... C...T..... ......C... ...G...... A.......GG  P
.T.T....T- .A....CA.C C...ATT... A..AA.C... .T.G.....T G.T...G..G  E
                                          1100
```

```
      1690        1700        1710        1720        1730        1740
ACTAGCCCTG AAAATGGATG GCGCTGGAGC GTCGGGCCCA TACCCGGcCG TCGCCGGCGC  X
.......... .......... .....CA... ..GTTA..T. ...T.TA... ..--------  Y
GG...T.... ........C. AG.T.CA.C. AGT.CAA... ....-....A AG..TCTTAT  P
G.CT..G.G. ..G...T.AC .G.GCTA.A. CAT.CA..G. AG.TGC.G.A G.--------  E

      1750        1760        1770        1780        1790        1800
TGGGTCAGTC CGCGGGGGCT AGGCCGCGAC TGAGTAGGAG GGCCCCGGCG GTGGGCGCGG  X
A....TGA.A T--------G .T...CT... -.......CA ...GTG.AG. TCA.TGA..A  Y
ATTCGGCTCG .AA.CTATTC CATGAGTA.A C.......GC ..ATA.CC.A .GTC.G.TT.  P
3
---.A.GC.T ATGC.----- ------TTGT ..G.....G. A..GTTCTGT AA.CCT..-.  E
                                              1200

      1810        1820        1830        1840        1850        1860
AAGCGCGCGC GAGGGCCCGG GTGGAGCCGC CGCGGGTGCA GATCTTGGTG GTAGTAGCAA  X
.GC.TA.AC. .TAA.GT... ..C..A.G.. .T.TA..... .......... ..........  Y
.....T..TG TT.ACAG.AC .....C.G.. ..G..CA... .........A ..........  P
...GTGTGCT .T.A.G.AT. C.....GTAT .AGAA....G A..GC..ACA TA....A.G.  E

      1870        1880        1890        1900        1910        1920
ATATTCAAAC GAGAACTTTG AAGGCCGAAG TGGAGAAGGG TTCCATGTGA ACAGCAGTTG  X
.........T .......... ...A.T.... ...G...A.. .....C..C. ..........  Y
G......T.T .CA....C.. .......... .....G.... ....TCAAC. ..........  P
           C
TA.AG.GGGT ..A..GCCC. CTC....G.A GACC-..... ....TGTCC. ..GTT.A.C.  E
                      1300

      1930        1940        1950        1960        1970        1980
AACATGGGTC AGTCGGTCCT AAGAGATGGG CGAGCGCCGT TCGGAAGGGA CGGGCGATGG  X
G..G.....T .....A.... .......... GA...T.... .TCA..---- G.CCT...TT  Y
.TTGG....T ......AT.. ...C.TCAA. G..AAC.TAG GTTA.CCTCG G.CT.TCGTC  P
              A
GGGCA....G .....AC..C T.AG.CGA.. ...AA.G... AG-------- ----------  E
                        C

      1990        2000        2010        2020        2030        2040
CCTCCGTCGC CCTCGGCCGA TCGAAAGGGA GTCGGGTTCA GATCCCCGAA CCCGGAGTGG  X
TA.G.AGGC. A.C------. .......... A..C....A. ...T..G... ..T...TAT.  Y
..CGT.G.TT GGG.TCGGAC G...C..A.. AG......A. T..T..T.C. .T.AACTC.A  P
4
---------- ---------- ....-T.... AA.A....A. T..T..T.T. .TT..T..TA  E
                                              1400

      2050        2060        2070        2080        2090        2100
CGGAGACGGG CGCCCGCGGC CCCCCCCGAC GCCTCGCGGC GGCGGGGGGG CGGGGGCGTC  X
GATT------ ---------- ---------- ---------- ---------- ----------  Y
T.TG------ ---------- ---------- ---------- ---------- ----ACA..G  P
.T.C------ ---------- ---------- ---------- ---------- ----------  E
```

```
      2110        2120        2130        2140        2150        2160
CAGTGCGGCG ACGCGACCGA TCCCGGAGAA GCCGSGSGGG AGSCCCGGGa GAAGAGTTCT  X
.TTCA...TA ...TA..T.. ATGT.....C .T..G-C.C. ..C..T.... .G..T--.A.  Y
GGCGA.CTTA .A....TATC CGAAACCAGT .A..TAACT. G.AA...T.. ..G.GA-.T.  P
---------- ---------- ------GA.G .GG.GAC..A GAAGG.TATG TTG.CCGGGC  E

      2170        2180        2190        2200        2210        2220
CTTTTCTTTG TGAAGGGCAG GGCCGCCCTP GAATGGGTTC GCCCCGAGAG AGGGGCCCGC  X
........CT .-..CA..T- TAT.A...CG ....T....T AT..G....T G...T.TTAT  Y
TCCC..C... .T.C...AGT TAA.TG.TCT .GTAA.C..G CATTGTGC.. .T.CC.AGCT  P
                                                          GCAAC
GACGGT.G.C CCGGTTTA.. C.TGTAGGCT .GT.TTCCAG ..AAATCCG. .AAAT.AA.G  E
                                                 1500

      2230        2240        2250        2260        2270        2280
GCCgTTGGAA AGCGTCGCGG TTCCGGCGGC GTCCGGTGAG CTtCTCGCTG GCCCTTGAAA  X
.--.C..... GAG.C.AGCA CCTTT..T.G C.......C. ...G.-.AC. ....G.....  Y
A.TCC.C..G ...A.GT..T .CTTC.AT.T ..T..CG.TC TCCT..CG.T .....A....  P
CTGAGGC.TG .TGACGAG.C ACTAC.GT.. TGAA.CAACA AA.GC.CTGC TT..AG....  E

      2290        2300        2310        2320        2330        2340
ATCCGGGGGA GAGGGTGTAA ATCTCtGCGC CGGGCCGTAC CCATATCCGC AGCAGGTCTC  X
....ACA... AG.AA--..- G.T.TCAT.. AA..T..... TG...A.... ..........  Y
.G.T..CA.. -T.....A.. G..CT.CG.T T.AA...... .T.-...... ..........  P
                  CGTGTT
.G..-TCTA. .CATCA---- --GGTAA.AT .AAAT..... ..CA.A...A CA.....GGT  E
                                    1600

      2350        2360        2370        2380        2390        2400
CAAGGTGAAC AGCCTCTGGC ATGTTAGAAC AATGTAGGTA AGGGAAGTCG GCAAGTCAGA  X
.......... .......A.T -..A.....T .......A.. .......... ....AAT...  Y
....A...G. ..T..... -GCA...... ..A..... ......T... .....C.G..  P
                                      C
..G.TA..GA .TA.CAA... -GC..GAG.G ..CT.G...G .A....C.A. ....AATG.T  E

      2410        2420        2430        2440        2450        2460
TCCGTAACTT CGGGATAAGG ATTGGCTCTA AGGGCTGGGT CGGTCGGGCT GGGGCGCGAA  X
.......... .......... .......... ....TC.... A.TGA....C TT..TCA..C  Y
.T........ .......... .......... TA........ GTCG...... ...TAAG.CT  P
                                                 T
G......... .....G.... CAC-...--- ---.A.AT.. A...GA..TC CCT-...---  E
              1700

      2470        2480        2490        2500        2510        2520
GCGGGGCTGG GCCGCGCCGC GGCTGGACGa aGGCGCCcCC GtGGCGCtCt TTCCCTCCTC  X
..A.C.GGC. TG.T------ ---------- ---------- ---------- ----------  Y
CGC...A..T ..-------- ---------- ---------- ---------- ----------  P
---------- ---------- ---------- ---------- ---------- ----------  E
```


2530 2540 2550 2560 2570 2580
CCGCCCCCCT CTCTCTCCGG CCCCCCCTCG CGGGGGGCCG CGGGGGCGGG GGGGCGCGGG X
---------- ---------- ---------- ---------- ---------- ---------- Y
---------- ---------- ---------- ---------- ---------- ---------- P
---------- ---------- ---------- ---------- ---------- ---------- E
2590 2600 2610 2620 2630 2640
GGGCCGGGAG CGCCCGGCGG CGGCGACTCT GGACGCGCGC CGGGCCCTTC CTGTGGATCG X
---------- ---------- ---------- ---------- ---------- ---------- Y
---------- ---------- ---------- ---------- ---------- ---------- P
---------- ---------- ---------- ---------- ---------- ---------- E
2650 2660 2670 2680 2690 2700
CCCCAGCTGC gGCGCGCGCC TCTCCCCCGC GCCGTCCCCC tCCTGCGCCT CCCCCCGTCA X
---------- -----TGTGG AC.G.TTG.T .GG.CTTG.T CTGCTA.G.G GA.TA.T.GC Y
-----TT.AG TAATG..CGA GGCG.TTTCG .GT.C.GAGA G.G..AA.GG G..G.TCGGG P
---------- ---------- ---------- ---------- ---------- ---------- E
2710 2720 2730 2740 2750 2760
GGGGACGGGG CGCGWGCSGC GGGGCGGCCG GGGCGGGCCC GGCCTCGGCC GGCGCCTAGC X
.T.CCTT.TT GTA.ACGGC. TT..TA.GTC TCTT.TAGA. C.T.G.TTG. TA.AAT..A. Y
.CT..AAACT A.ATCATT.. ..A.GTT.GC CC.A.T.GTG AAAAG....A CA.CG...A. P
---------- ---------- ---------- ---------- ---------- ----GGAT.G E
2770 2780 2790 2800 2810 2820
AGCTGACTTA GAACTGGTGC GGACQAPGGG AATCCGACTG TTTAATTAAA ACAAAGCATC X
GATCA..... ....A. ....A.G... ....T..... .C..... ...T.....T Y
....A..G.. .....TACAA A.G.T.G... .....A.... .A..... ...T...GAT P
......--A. TC.G.C.AAG AT..C.GCT. GC.G.A.... ....T.A... ...C....CT E
A
2830 2840 2850 2860 2870 2880
GCGAAGGCCC GAGGCGGGTG TTGACGCGAT GTGATTTCTG CCCAGTGCTC TGAATGTCAA X
....T..T.A ..AAGT.A.. ....A.. .... Y
TT.TT..TG. C.AA-.C... .AA...AATC ....G....T.. P
.T.C.AA.A. ..AAGT..AC G.-.TA..G. ....CGC... ...G....CG G.TTA.CGC. E
1800
5
2890 2900 2910 2920 2930 2940
AGTGAAGAAA TTCAATGAAG CGCGGGTAAA CGGCGGGAGT AACTATGACT CTCTTAAGGT X
....CC... .... Y
.A..GC.... .C...CC... .T........ .... P
..C....CTC ..G.TC.... .C.C...... ......CC.. ......A..G G..C...... E
1900

```
      2950        2960        2970        2980        2990        3000
AGCCAAATGC CTCGTCATCT AATTAGTGAC GCGCATGAAT GGATGAACGA GATTCCCACT  X
.......... .......... .......... .......... ....T..... ..........  Y
.......... ........T. ....T..... .......... ....T..T.. ..........  P
...G....T. ..T...GGG. ..G.TCC... CT...C.... ..CGT..T.. TGGC.AGG..  E

      3010        3020        3030        3040        3050        3060
GTCCCTACCT ACTATCTAGC GAAACCACAG CCAAGGGAAC GGGCTTGGCG GAATCAGCGG  X
.......T.. .......... .......... .......... .........A ..........  Y
.......... .......... .......... .......... .........A C...T.....  P
...T.C...C GAG.CTC..T ....TTGA.C .GCT.T...G AT..AGT.TA CCCG.G..AA  E
     2000                        ^T

      3070        3080        3090        3100        3110        3120
GGAAAGAAGA CCCTGTTGAG CTTGACTCTA GTCTGCAACT GTGAAGAGAC ATGAGAGGTG  X
.......... .......... .......... ..T..AC.T. .......... ..AGAG....  Y
.......... .......... .......... .G.ATAG..- .C..G.T..T TCT.A.....  P
.ACGGA.... ...C..GA.C ...T...A.. .CT..AC... .AAC.TT..G CCTT..T...  E
                                                    2100

      3130        3140        3150        3160        3170        3180
TAGGATAAGT GGGAGGCCCC CGCGCTCGTC GCAAAGGGGC GCCGCCGGTG AAATACCACT  X
...A...... .....CTT.- ---------- ---------- .G....A... ..........  Y
...C...G.. ......G... ATG------- ---------- C...T.AA.. .........C  P
.......G.. .......TTT GAA.TGT.AG T.TGCAT..A ....A.CT.. .........C  E
                       GACGCC

      3190        3200        3210        3220        3230        3240
ACTCTTATCG TTTTTTCACT TACCCGGTGA GGCGPPGGGG CGAGCCCCGA GGgGCTCTCG  X
..CT....A. ...C..T... ..TT.AA... ------.C.. A.CTGGAATT CATTT..CAC  Y
...T.CGA.A .CGC..TG.. A.TG.T..A. C.AACGAAC. GA.C.G.GTC CCTCTCAC..  P
CT.TAATGTT .GA.G.TCTA ACGTT.ACCC .TA------- ---------- ----------  E
                      2200

      3250        3260        3270        3280        3290        3300
CTTCTGGACC CAAGCGCSCG GCCCCCGCGC CGGGCGCGAC CCGCTCCGAG GACAGTGGCA  X
G....A..TT ....GT.C.- ------ATT. G....-T..T ...GGTT..A ....T..T..  Y
         ^C
TAAAATCGG. .TCTGTGGGC TT.ATGC.CG GTACGTAATT G..TGTA.CA ...TA.CTAT  P
   ^6
---------- ---------- ---------- --------.T ...GGTT.C. .......T.T  E

      3310        3320        3330        3340        3350        3360
GGTGGGGAGT TTGACTGGGG CGGTACACCT GTCAAACCGT AACGCAGGTG TCCTAAGGCG  X
.......... ...G...... ...C...T.. ..T....GA. .......A.. .........G.  Y
.T........ ...G...... ...A.A.-.. .CT.C..G.C ....GCA..C .........TC  P
......T... .......... ....-.T... CCT...GA.. ....G...A. CA.G....TT  E
```

```
      3370        3380        3390        3400        3410        3420
AGCTCAGGcG AGcGACAGAA ACCTCCCGTG GAGCAGAAGG GCAAAAGCTC GCTTGATCTT  X
G.....T.-. ..-A...... .T....A..A ..A..A.... .T......C. C......T..  Y
CA.....A-. .C-....... ..G..T...A ...A-T..A. .......TGG ....A.CTCG  P
G...A.TC-C T.-.T.G..C .T.AGGA.GT T..TGC..T. ...T....CA ......CTGC  E
     2300

      3430        3440        3450        3460        3470        3480
GATTTTCAGT ATGAATACAG ACCGTGAAAc GCGGGWGCCT CACGATCCTT CTGACTTTTT  X
.......... G........A ...A.....- .T.T.-.... AT........ TA.T.CC.CG  Y
C......... .GT...GTGA .G.AA....- TT.C.-..T. A......... AGCGGCGGG.  P
..GCG.G.CG GC.CGAG... -GT.C....- ..A..-T.A. AGT.....GG TG.TTC.GAA  E
                                                      2400

      3490        3500        3510        3520        3530        3540
GGGTTTTAAG CAGGAGGTGT CAGAAAAGTT ACCACAGGGA TAACTGGCTT GTGGCcGGCC  X
.AA...G.G. .TA......C .......... .......... .......... .....-A.T.  Y
.CCAGCC.C. .TT......A G......... .......... .......... .....-C...  P
          C
T..AAGGCCA TC.CTCAACG G.T.....G. ..TC.G.... ....A....G A.AC.-.C..  E
        G

      3550        3560        3570        3580        3590        3600
AAGCGTTCAT AGCGACGTCG CTTTTTGATC CTTCGATGTC GGCTCTTCCT ATCATTGTGA  X
.......... ......A.T. .........T .......... .......... .....ACC..  Y
.......... ........G. .......... .......... .......... .....AC.A.  P
...A...... .T.....G.. G.G....GCA .C........ .....A..AC ...C.G.G.C  E
                                  2500

      3610        3620        3630        3640        3650        3660
AGCAGAATTC ACCAAGCGTT GGATTGTTCA CCCACTAATA GGGAACGTGA GCTGGGTTTA  X
.......... GGT....... .......... .......... .......... ..........  Y
...-...... .GT...T... .......... ...T..T.-. .......... ..........  P
T.A..T.GGT C.....G..A T.GC.....G ..ATT...AG T..T...C.. ..........  E

      3670        3680        3690        3700        3710        3720
GACCGTCGTG AGACAGGTTA GTTTTACCCT ACTGATGATG TGTTGTCGCA ATAGTAATCC  X
.......... .......... .......... ........A- ....AC.... ........TG  Y
...-...... .......... .......... C..AC.TGC. GTG..C.CTG ......GAT.  P
                                            A
..A....... ......T.CG ..CCCTAT.. G.C.TG.GGC ..GA.AACTG .GG.GGGCTG  E
          2600                              C

      3730        3740        3750        3760        3770        3780
TGCTCAGTAC GAGAGGAACC GCAGGTTCAG ACATTTGGTG TATGTGCTTG GCTGAGGAGC  X
AA..T..... .........A .TTCA...G. .T.A.....T .T..C.GC.. T....TC..G  Y
ACT....... .......... AGT.A..... ..CAA..... ..A.C...GC T.GA.C.G.-  P
CT.CT..... .......C.G .AGT.GA.GC .TCAC..... .TC.G.T.GT CA..CCA.TG  E
                                                      2700
```

```
      3790        3800        3810        3820        3830        3840
CAATGGGGCG AAGCTACCAT CTGTGGGATT ATGACTGAAC GCCTCTAAGT CAGAATCCCC  X
..T..CC... .......... .C.CT..... ...G...... .......... ........AT  Y
..GA.C.CTA G......-G. ....T....G .A.G...G.A ..A......C .C...G..AT  P
GC.CT.CC.. GTAGCTAA.. GC.G.A...A .GTG.....A ..A......C AC...A.TTG  E
                          AA

      3850        3860        3870        3880        3890        3900
CCTAAACGTG ACGATACCGC AGCGCCGCGG AGCCTCGGTC GGCCTCGGAT TAGCCGGCgC  X
G.....C.C. GT...TT.TT T..T..A.AC .ATA.A.A.- ..ATA..A.. A..G..T.CT  Y
       G

G...G.TAG. GGACGT.GCA .T.AAA.ATA C..T.GCT.T TAAG..T.GC A...G.T.TT  P
                                                      C

..CCG.GA.. .GTTCT..CT GA.C.TT--- ---------- ---------- ----------  E

      3910        3920        3930        3940        3950        3960
CCCCCCCGGG GGGCGCCGGC GGGCAGAGCC GCTCGCCTCG GGACCGGAGC GCGGACGAAA  X
---------- -T.T.G..T. .CTG.ACCAT AGCA.G..A. CA..GT.CA. TT..CG....  Y
                                             G

TTA..ATCAA CA..CAGTCA ACTG..CAGG A.G..ATG.A AAGAG.TC.A C.C.TAA...  P
                                                  AATTAG

---------- ---------- ---------- ---------- ---------- ----------  E

      3970        3980        3990        4000        4010        4020
GGGGGCCGCC TCTCTCCCGG AGCGCACCGC ACGTTCGTGG GGAACCTGGT GCTAAATCAT  X
..CCTTG.GT G..TG.TG.- --..A.TT.. .ATG..A.TT T.--.G...G .A.........  Y
..TCAGT... ..GT.AAGCC GC.C.CA..G GG.CATCAAC ..CG.TCC.C CT....C..C  P
                                                         8

---------- ---------- ---------- ---------- -T..GGGTCC TGA.GGAACG  E
                                          2800

      4030        4040        4050        4060        4070        4080
TCGTAGACGA CCTGATTCTG GGTCAGGGTT TCGTGCGTAG CAGAGCAGCT ACCTCGCTGC  X
.T...T.... .T.AGA.G.A CAA.G....A .T..AA.C.. T....T...C TTG.T.T.A.  Y
A..C.T.... .TGTGCGTA. C.GA.AT.G. GTTA.T.ATC A..CTGGCTG G..GTCTA..  P
A

.T.A...... .GACG..GAT A.G.C....G .GTAAGCGCA GC..TGC.T. GAGCTAACCG  E

      4090        4100        4110
GATCTATTGA AAGTCATCCC TTGGCCAAGC                                    X
.....GC... G.T.A.G..T ...TTGTCTG ATTTGT                             Y
AT.GCGAGAC GCTGAGC.AG .GTTTGCGTT TTTTTGTCGG C                       P
.TA...A... .CCGTGAGG. ..AA..TT                                      E
                      2900
```

$G(A)_n$ and $A(A)_n$ oligonucleotides on a secondary structure model for Xenopus 28S rRNA (43) reveals that 85% of these bases are in single stranded regions. The stretches of homologous sequences in the three eukaryotes are separated by segments of sequence which differ in length and overall base composition between species but occur in the same positions (Fig. 4, Table 1). Within the coding region for the mature rRNA of Xenopus, the 28S rDNA has nine major tracts of DNA sequence not found in the E. coli 23S rDNA. There are a few small inserts only found in yeast and/or Physarum (Fig. 4). The base composition of the inserts varies significantly between species. The base composition of the Xenopus inserts is 83% G+C, resulting in the high (66%) G+C content of Xenopus 28S rDNA as a whole.

Figure 3. DNA sequence alignment of large subunit rDNA sequences. "X" represents Xenopus laevis 28S (this paper); "Y" represents yeast 26S (19); "P" represents Physarum 26S (21); "E" represents E. coli 23S (18). Sequence is from the RNA-like strand in all cases. Numbering at the top refers to the Xenopus sequence. Numbering at the bottom refers to the E. coli sequence; last digit of a number appears under the specified base. The E. coli sequence begins at base 158 due to the presence of 5.8S-like sequences at the 5' end of E. coli 23S (37-39). The following symbols are used; ambiguities account for only about 1% of the sequence:

P = purine	S = G or C	"." = same base as Xenopus
Q = pyrimidine	V = A or C	"-" = base missing relative to Xenopus
R = A or T	W = G or T	lower case = unclear if base is present or absent

A single digit beneath a sequence indicates long DNA sequences not present in Xenopus 28S rDNA:

Number	Organism and Coordinates		# bases
1	Yeast	450- 476	27
2	Physarum	452- 479	28
3	Physarum	1405-1421	17
4	Physarum	1564-1694	31
5	E. coli	1847-1861	15
6	Physarum	2726-2872	147
7	Physarum	3315-3339	25
8	Physarum	3659-3678	20

Short DNA sequences not present in Xenopus are shown directly in the figure. There are no differences between our sequence and the 5' 118 bases determined by Hall and Maden (14). The following six differences were found with the 3' 263 bases determined by Sollner-Webb and Reeder (16): position 3850 is a G in this paper, an A in their paper. Bases 3891, 3892, 3897, and 3899 are not present in their sequence. An extra base (A) is inserted between bases 4103 and 4104 in their determination.

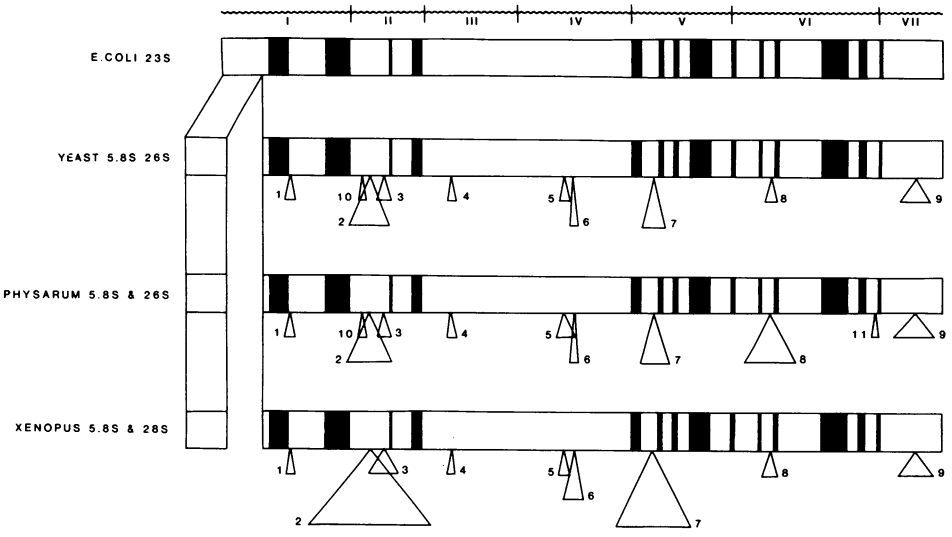


Figure 4. Comparison of known eukaryotic large sub-unit rDNA structure with E. coli showing conserved, non-conserved and inserted regions relative to E. coli. Black areas indicate regions of ≥ 12nt with ≥ 75% conservation between all four organisms according to the alignment of Figure 3. The second conserved region from the 5' end represents 3 areas of conservation separated by 10 and 6 bases, respectively, that could not be resolved in this figure. White areas represent non-conserved regions. Triangles represent inserted sequences ≥ 20 bases relative to E. coli. Inserts are numbered with respect to Xenopus, for ease of reference. Roman numerals at the top refer to secondary structural domains I-VII of E. coli (40, 41). The eukaryotic 5.8S molecule is shown as corresponding to the 5' end of E. coli 23S based on primary and secondary structure conservation (37-39).

Table 1: Percentage homologies and base compositions were calculated from the alignments in Figure 3.

	Xenopus	Yeast	Physarum	E. coli
% Homology to Xenopus	100	68.6	52.9	46.9
G+C Base Composition:				
Total	65.6	47.9	53.5	53.4
Conserved Regions	49.4	46.2	47.2	53.5
Non-conserved Regions	58.7	46.9	52.9	53.4
Inserted Regions	82.8	51.3	60.3	-

DISCUSSION

Conserved Primary Sequence

We have determined the entire primary structure of Xenopus laevis 28S rDNA; this constitutes the first completed nuclear 28S rDNA sequence from a multicellular organism. The Xenopus 28S rDNA sequence, coupled with the previously determined Xenopus sequences of the 18S rDNA (12), some tRNAs (44-48), 5S RNA (49-51), and 5.8S rDNA (13, 14), yields a complete array of RNA sequences for this model system of the eukaryotic translational machinery. Our direct sequence analysis of Xenopus 28S rDNA has identified evolutionarily conserved areas not mapped by our previous heterologous hybridization studies (7). Fourteen areas of 28S primary sequence are highly conserved even in the distant taxonomic comparison of the prokaryote E. coli and the vertebrate Xenopus. This figure of fourteen is not absolute, as it represents only the number of areas fulfilling our definition of homology: shorter stretches of homology also may occur and may be of functional significance.

Areas of sequence conservation between such diverse organisms most likely play vital functional roles; mutations within such areas would be deleterious or lethal and not perpetuated during evolution. One of the earliest notions was that the role of rRNA was only structural, to act as a scaffold for ribosomal proteins during ribosome biogenesis or reconstitution. A dozen E. coli ribosomal proteins can interact directly with E. coli 23S rRNA, and some of these have been mapped to discrete locations (reviewed by 52). E. coli ribosomal protein L1 is the only one which also binds to a eukaryotic rRNA, and the L1 binding site in Dictyostelium 26S rRNA also shares both primary and secondary structure homology with the same region in Xenopus 28S rRNA (Xenopus 28S coordinates 3096-3180, including a conserved region in Fig. 4) (32). Other areas of rDNA sequence conservation may also function to bind ribosomal proteins.

However, in addition to binding ribosomal proteins, the large subunit large rRNA has been implicated in other functions. Using the sequence alignment shown in Fig. 3, and by comparison to the models proposed for E. coli 23S (40-42) and yeast 26S rRNA (19), we have been able to derive a secondary structure model for Xenopus 28S rRNA (43). Conserved areas are described in our model which are candidates for regions of functional significance: the GTPase center, A and P sites for tRNA binding, 5S RNA binding site, and the peptidyl transferase center. Many of the conserved regions (fig. 4) fall outside the putative areas for the functions listed above, and may play roles for other functions not yet localized.

Insertions in rDNA

As shown in Fig. 4, the increased size of eukaryotic 28S rRNA relative to E. coli 23S rRNA can be accounted for largely by blocks of inserted sequence in the former. Generally these inserts occur at the same position in all known eukaryotic rDNAs, but their size and sequence vary. However, in one case, at the 3' end of insert 2, there is conservation of sequence in all three eukaryotes (Fig. 3). This region corresponds to Xenopus 28S rDNA coordinates 952-976, and suggests that these 25 completely conserved bases, present in all known rDNAs of eukaryotes but absent in E. coli and organelles, may play a function common to eukaryotic ribosomes.

Similarly, the lack of sequence conservation for all the other inserts in eukaryotic 28S rDNA raises the possibility that they play no role at all, but are not deleterious for ribosome function, and hence are tolerated and allowed to remain in the mature rRNA. Although Xenopus 28S rRNA has not been sequenced directly, the lack of S1 nuclease cuts in rRNA-rDNA hybrids at the insert positions (22, 53) suggests that transcripts of the inserts form part of the mature 28S rRNA. In some groups of organisms, however, there may have been sufficient negative pressure to cause removal of the insert from the final rRNA product. Specifically, the bases have been localized in the fungus fly which are cut out during processing, thereby dividing 28S rRNA into the α and β halves (R. Renkawitz, personal communication), and the bases which are removed include insert 6. In a second example, eukaryotic insert 9 appears to be treated like a "spacer" in chloroplast rDNA of higher plants and is not present in mature rRNA. The nucleotides 3' to this "spacer" are found as a 4.5S RNA; they are homologous to the 3' end of E. coli 23S rRNA but are no longer covalently bound to the bulk of the large ribosomal RNA (54-56; 39).

These observations on eukaryotic rDNA inserts can be extended to suggest that they form a subclass of the DNA sequences inserted into rDNA. The subclasses are defined by the RNA processing fates of the insertions. The first subclass of insertions are what we have called "inserts" above. Except for the insect and chloroplast cases above, transcripts of the eukaryotic 28S rDNA inserts shown in Fig. 4 appear not to be removed by processing cuts, and presumably are present in the mature 28S rRNA. The transcribed intervening sequences found in the rDNA of some non-dipteran species represent a second class. They represent insertion sequences whose transcripts may interfere with ribosome function, and therefore are removed by splicing events from the rRNA. Indeed, the areas in which intervening sequences are found are highly conserved in primary sequence (reviewed in 57), implying that a functional

site has been interrupted by such an insertion.

The internal transcribed spacer 2 (ITS 2) which separates 5.8S from 28S rDNA falls within a third class of rDNA insertions. 5.8S RNA is homologous to the 5' end of E. coli 23S rRNA (37-39); the ITS 2 insertion, which separates 5.8S from the remainder of the large ribosomal RNA, is cleaved out during processing. However, unlike the intervening sequence class of insertions, the processing cuts are not followed by a splicing ligation. Perhaps it is not necessary for 5.8S RNA to be covalently joined to 28S rRNA, because both of the 5.8S RNA termini are already hydrogen bonded to 28S rRNA (58, 59).

The three classes of rDNA insertions generally do not have the sequence properties of IS elements, but this could simply be a reflection of mutational change over evolutionary time. Nonetheless, it is satisfying to think of these classes of rDNA insertions as foreign DNA sequences which have integrated into rDNA. Depending on their site of integration within rDNA, their transcripts may be handled differently during rRNA maturation, and thereby define the three subclasses of rDNA insertions as discussed above.

ACKNOWLEDGEMENTS

We thank Drs. Albert Dahlberg and Christian Zwieb for helpful criticism, and Mrs. Carol King for typing this manuscript. This research was supported by grant PHS-GM 20261 to S.A.G., and fellowship # 2175 from the American Cancer Society to V.C.W.

CURRENT ADDRESSES

* Department of Biology, University of California-San Diego, La Jolla, California

** The Rockefeller University, New York, New York 10021

\+ Institute for Enzyme Research, University of Wisconsin, Madison, Wisconsin 53706

++ Genetisch Laboratorium, Katholieke Universiteit, Nijmegen, The Netherlands

REFERENCES

1. Crick, F.H.C. (1968) J. Mol. Biol. 38, 367-379.
2. Orgel, L.E. (1968) J. Mol. Biol. 38, 381-393.
3. Woese, C.R. (1980) in Ribosomes: Structure, Function, and Genetics, Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 357-373. University Park Press, Baltimore, Maryland.
4. Sinclair, J. and Brown, D.D. (1971) Biochemistry 10, 2761-2769.
5. Birnstiel, M.L. and Grunstein, M. (1972) FEBS Symp. 23, 349-366.
6. Gerbi, S.A. (1976) J. Mol. Biol. 106, 791-816.
7. Gourse, R.L. and Gerbi, S.A. (1980) J. Mol. Biol. 140, 321-339.
8. Cox, R.A. and Thompson, R.D. (1980) Biochem. J. 187, 75-90.
9. Morrow, J.F., Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Goodman, H.M. and Helling, R.B. (1974) Proc. Natl. Acad. Sci. USA 71, 1743-1747.
10. Salim, M. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 2871-2884.

11. Maden, B.E.H., Moss, M. and Salim, M. (1982) Nucleic Acids Res. 10, 2387-2398.
12. Salim, M. and Maden, B.E.H. (1981) Nature (London) 291, 205-208.
13. Boseley, P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nucleic Acids Res. 5, 1121-1137.
14. Hall, L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 5993-6005.
15. Boseley, P.G., Moss, T., Mächler, M., Portmann, R. and Birnstiel, M.L. (1979) Cell 17, 19-31.
16. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499.
17. Moss, T., Boseley, P.G. and Birnstiel, M.L. (1980) Nucleic Acids Res. 8, 467-485.
18. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. USA 77, 201-204.
19. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 6935-6952.
20. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.R., Zakharyev, V.M. and Bayev, A.A. (1981) Nucleic Acids Res. 9, 6953-6958.
21. Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, S. and Sakaki, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3163-3167.
22. Brand, R.C. and Gerbi, S.A. (1979) Nucleic Acids Res. 7, 1497-1511
23. Dawid, I.B. and Wellauer, P.K. (1976) Cell 8, 443-448.
24. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
25. Course, R.L. and Gerbi, S.A. (1980) Nucleic Acids Res. 8, 3623-3637.
26. Seif, I., Khoury, G. and Dhar, R. (1980) Nucleic Acids Res. 8, 2225-2240.
27. Peacock, A.C. and Dingman, C.W. (1968) Biochemistry 7, 668-674.
28. Frank, R., Müller, D. and Wolff, C. (1981) Nucleic Acids Res. 9, 4967-4979.
29. Sege, R., Söll, D., Ruddle, F.H. and Queen, C. (1981) Nucleic Acids Res. 9, 437-444.
30. Queen, C.L. and Korn, L.J. (1979) in Methods in Enzymology, Grossman, L. and Moldave, K., Eds., Vol. 65, pp. 595-609.
31. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2398.
32. Course, R.L., Thurlow, D.L., Gerbi, S.A. and Zimmermann, R.A. (1981) Proc. Natl. Acad. Sci. USA 78, 2722-2726.
33. Birnstiel, M., Speirs, J., Purdom, I., Jones, K. and Loening, U.E. (1968) Nature (London) 219, 454-463.
34. Lava-Sanchez, P.A., Amaldi, F. and Posta, A. (1972) J. Mol. Evol. 2, 44-55.
35. Amaldi, F. (1969) Nature (London) 221, 95-96.
36. Loening, U.E., Jones, K. and Birnstiel, M.L. (1969) J. Mol. Biol. 45, 353-366.
37. Nazar, R.N. (1980) FEBS Letts. 119, 212-214.
38. Jacq, B. (1981) Nucleic Acids Res. 9, 2913-2932.
39. Clark, C.G. and Gerbi, S.A. (1982) J. Mol. Evol. 18, 329-336.
40. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.-P., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 4303-4324.
41. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 3287-3306.
42. Noller, H.F., Kop, J., Wheaton, V., Brosius, H., Gutell, R.D., Kopylov, A., Dohme, F. and Herr, W. (1981) Nucleic Acids Res. 9, 6167-6189.
43. Clark, C.G., Tague, B.W., Ware, V.C. and Gerbi, S.A. (1983) in preparation.
44. Kressman, A., Hofstetter, H., DiCapua, E., Grosschedl, R. and Birnstiel, M.L. (1979) Nucleic Acids Res. 7, 1749-1763.
45. Telford, J.L., Kressman, A., Koski, R.A., Grosschedl, R., Mueller, F., Clarkson, S.G. and Birnstiel, M.L. (1979) Proc. Natl. Acad. Sci. USA 76,

2590-2594.
46. Mueller, F. and Clarkson, S.G. (1980) Cell 19, 345-353.
47. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Nature (London) 294, 626-631.
48. Hofstetter, H., Kressman, A. and Birnstiel, M.L. (1981) Cell 24, 573-585.
49. Brownlee, G.G., Cartwright, E., McShane, T. and Williamson, R. (1972) FEBS Letts. 25, 8-12.
50. Wegnez, M., Monier, R. and Denis, H. (1972) FEBS Letts. 25, 13-20.
51. Ford, P.J. and Southern, E.M. (1973) Nature New Biol. 241, 7-12.
52. Zimmermann, R.A. (1980) in Ribosomes: Structure, Function, and Genetics, Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 135-169, University Park Press, Baltimore, Maryland.
53. Botchan, P. (1978) Carnegie Inst. Yearbook 77, 123.
54. Mackay, R.M. (1981) FEBS Letts. 123, 17-18.
55. Machatt, M.A., Ebel, J.-P. and Branlant, C. (1981) Nucleic Acids Res. 9, 1533-1550.
56. Edwards, K., Bedbrook, J., Dyer, T. and Kössel, H. (1981) Biochem. Int. 2, 533-538.
57. Gerbi, S.A., Gourse, R.L. and Clark, C.G. (1982) in The Cell Nucleus, Busch, H. and Rothblum, L. Eds., Vol. X, pp. 351-386, Academic Press, New York.
58. Pace, N.R., Walker, T.A. and Schroeder, E. (1977) Biochemistry 16, 5321-5328.
59. Nazar, R.N. and Sitz, T.O. (1980) FEBS Letts. 115, 71-76.