The EMBO Journal vol.3 no.13 pp.3303 – 3310, 1984

Human genes and pseudogenes for the 7SL RNA component of signal recognition particle

Elisabetta Ullu and Alan M. Weiner

Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 3333, New Haven, CT 06510, USA

Communicated by G. Bernardi

Of the several hundred 7SL RNA-like sequences that are dispersed in human DNA, no more than four are likely to represent genes for 7SL RNA; the majority are 7SL pseudogenes which appear to result from the reverse flow of genetic information from 7SL RNA back into genomic DNA. We present the sequence of five 7SL pseudogenes displaying an unprecedented diversity of structures. All are truncated copies of 7SL RNA, but the site of truncation can occur at either the 5′ end, the 3′ end or at both ends of the RNA sequence. We suggest that such diverse 7SL pseudogenes are generated by different but related pathways. In particular, we argue that two of the loci are secondary 7SL pseudogenes which derive from RNA polymerase III transcripts of primary (pre-existing) 7SL pseudogenes. We also report the isolation and characterisation of a human genomic clone carrying two linked 7SL RNA coding regions, 7L30.1 and 7L30.2. The 7L30.2 locus differs by several single base changes from the known human 7SL RNA sequences and does not appear to be expressed at a detectable level in HeLa cells. The 7L30.1 locus is an authentic 7SL RNA gene encoding one of the three sequence variants of human 7SL RNA.

Key words: Alu sequences/7SL RNA genes/pseudogenes/transcription

Introduction

In higher eukaryotes the co-translational insertion of secretory proteins into the lumen of the endoplasmic reticulum is mediated by a cytoplasmic ribonucleoprotein complex, called signal recognition particle or SRP (Walter and Blobel, 1982). SRP consists of six polypeptides and a single molecule of 7SL RNA, 300 nucleotides in length. Micrococcal nuclease digestion of dog SRP has shown that 7SL RNA is essential for SRP function (Walter and Blobel, 1982; Gundelfinger *et al.*, 1983). The synthesis of 7SL RNA is sensitive to high concentrations of α-amanitin, suggesting that 7SL RNA genes are transcribed by RNA polymerase III (Zieve *et al.*, 1977).

The structure of 7SL RNA attracted considerable interest after mammalian 7SL RNA was found to be partially homologous to Alu DNA (Weiner, 1980; Ullu *et al.*, 1982). Alu sequences constitute the dominant family of middle-repetitive DNA sequences in the mammalian genome (for a review, see Jelinek and Schmid, 1982). Rodent Alu-equivalent elements have a monomeric structure while human Alu elements are dimeric. The dimer consists of two similar sequences, ~130 bp long, arranged head to tail; the right hand monomer contains a 30-bp insertion which is usually not present in the left half (Deininger *et al.*, 1981). Previously, we found that human 7SL RNA consists of an Alu right monomer interrupted by 140 nucleotides of non-Alu DNA, termed the S sequence (Ullu *et al.*, 1982). The S sequence is diagnostic of 7SL RNA-related sequences and is present in multiple copies in the human genome.

The interrupted homology between 7SL RNA and Alu DNA suggested that their evolution might be related. Since the *Xenopus* and *Drosophila* 7SL RNA sequences are homologous to mammalian 7SL RNA, the Alu sequence in 7SL RNA must have appeared in evolution long before the mammalian radiation (Ullu and Tschudi, 1984). The genes coding for 7SL RNA are therefore likely to be the ancestors of the Alu sequence, and we favor the notion that Alu DNA arose from 7SL RNA (or DNA) through a deletion of the central 7SL-specific sequence.

We have used the central S sequence of the human 7SL cDNA clone p7L1 (Ullu *et al.*, 1982) as a probe to investigate the structure of the human 7SL RNA sequence family. We find that the S DNA family consists of several hundred cross-hybridising sequences, most of which show only limited homology to 7SL RNA. We report the nucleotide sequences of five such 7SL pseudogenes. In comparison with the sequence of human 7SL cDNA, two pseudogenes are truncated at the 3′ end, two at the 5′ end and one at both ends. We interpret the diversity of 7SL pseudogene structures as an indication that 7SL RNA pseudogenes might be generated by a variety of related mechanisms.

We also report the isolation and characterisation of a human recombinant DNA fragment which contains two neighbouring 7SL RNA coding sequences, 7L30.1 and 7L30.2. By several criteria, we show that 7L30.1 DNA represents one of the few functional 7SL RNA genes in the human genome. The 7L30.2 locus, instead, has several mismatches compared with the known human 7SL RNA sequences and does not appear to be expressed in HeLa cells.

Results

Human DNA contains several hundred imperfect copies of the S sequence

To isolate human sequences homologous to 7SL RNA, ~5 x 10^5 recombinant phages were screened with an S DNA probe. The probe spans the 125-bp *Sau*3A DNA restriction fragment of cDNA clone p7L1 and lacks homology with the Alu sequence (Ullu *et al.*, 1982). We had previously shown that S DNA defines a family of middle repetitive sequences in the human genome. The screen of genomic phage libraries confirmed this result, since 0.2% of the recombinant phages gave a positive signal under standard hybridisation conditions (50% formamide, 5 x SSC, 40°C). From this we estimated that human DNA contains ~500 sequences related to S DNA. We anticipated that 7SL RNA genes would be found among those S sequences bearing perfect homology to the S DNA of cDNA clone p7L1. To distinguish between conserv-

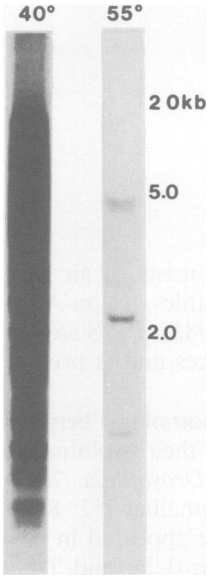

Fig. 1. Hybridisation of the S probe to human DNA. Human placental DNA (10 μg) was digested with *Eco*RI and *Hind*III restriction enzymes, electrophoresed through a 0.7% agarose gel, transferred to a nitrocellulose filter and hybridised with the S DNA probe as described in Materials and methods. The temperature of hybridisation was either 40°C or 55°C as indicated above each lane. Following hybridisation, the nitrocellulose filter was exposed at −70°C to a Kodak XAR-5 film using an intensifying screen. The exposure time was 16 h for both lanes.

ed and divergent S sequences, we tested the thermal stability of the hybrids formed between the S DNA probe and a large number of recombinant phages. The stability of the hybrids was assessed either by increasing the hybridisation temperature or by melting the hybrids in 0.1 x SSC at increasing temperatures. Only 1% of the hybrids displayed a thermal stability comparable with that of the S DNA fragment hybridised to itself (data not shown), indicating that the majority of genomic sequences hybridising to S DNA are imperfect copies of the S sequence. This observation was supported by the experiment shown in Figure 1. A Southern blot of human placental DNA was hybridised to the S DNA probe at low (40°C) or high (55°C) temperature. At 40°C a complex pattern of hybridisation was observed, but at 55°C the pattern is much simpler. Hybridisation temperatures above 55°C decreased the overall hybridisation signal, without any further decrease in the number of hybridising bands (data not shown). Taken together, these results first suggested that human DNA codes for relatively few 7SL RNA genes.

Using the criteria described above, we isolated four independent clones (7L30, 7LEM1, 7L3 and 7L28) with a perfect or nearly perfect S sequence. In addition, three clones (7L23, 7L63 and 7L7) with divergent S sequences were chosen for further analysis. Restriction enzyme mapping and Southern hybridisation revealed that six of the seven recombinant clones were derived from different loci. Each recombinant except 7L30 contained only a single region homologous to the S DNA probe. Clone 7L30 had two regions of homology to S DNA, denoted 7L30.1 and 7L30.2 (Figure 2). The seventh recombinant clone 7L3 was found to overlap the 7L30.2 locus.

The variety of human 7SL RNA pseudogenes

By DNA sequence analysis we confirmed that the cloned genomic S DNAs were indeed structurally homologous to

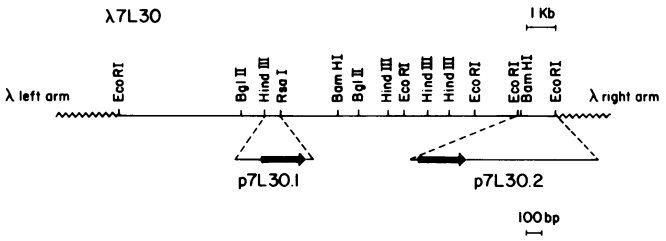

Fig. 2. Restriction enzyme map of recombinant phage 7L30. Recombinant phage 7L30 contains a 15-kb human DNA fragment inserted in the lambda vector Charon 4A (Lawn *et al.*, 1978). Solid arrows indicate the position and orientation of 7L30.1 and 7L30.2 loci. p7L30.1 and p7L30.2 represent plasmid subclones derived from the two loci.

7SL RNA. The nucleotide sequences of the seven 7SL RNA homologies are compared in Figure 3 with the three sequence variants of human 7SL RNA corresponding to 7SL cDNA clones 7L1a, 7L1b and 7L1c. We originally identified the 7L1a and 7L1b sequence variants which differ at positions 52, 53 and 69, as shown in Figure 3 (Ullu *et al.*, 1982). By restriction enzyme mapping and DNA sequence analysis of additional human 7SL cDNA clones, we have now characterised a new sequence variant of human 7SL RNA called 7L1c (data not shown). 7L1c is identical to 7L1b at positions 52, 53 and 69 but is distinguished by a G to A transition at position 275 of the 7SL cDNA sequence (Figure 3).

Five loci (7L28, 7LEM1, 7L7, 7L23 and 7L63) bear incomplete and divergent copies of human 7SL RNA and represent 7SL RNA pseudogenes. Both 7L28 and 7LEM1, which show almost perfect conservation of the central S sequence, are 3′-truncated 7SL pseudogenes. They are homologous to the 7L1a and 7L1b sequences, respectively. The homology with the 7SL cDNA sequences starts precisely at the 5′ end, but in both cases it is lost downstream from position 235. 7L7 and 7L23, which were originally scored as divergent S sequences, represent 5′-truncated pseudogenes since they lack 27 and 19 nucleotides at the 5′ end, respectively. The 3′ end of both 7L7 and 7L23 DNAs coincides with that of human 7SL cDNA. On the other hand, 7L63 is truncated both at the 5′ and at the 3′ end relative to the 7SL cDNA sequence: 12 nucleotides at the 5′ end and 69 nucleotides at the 3′ end are missing. The overall divergence of the 7SL pseudogenes from the 7SL cDNA sequences varies from a minimum of 0.4% (7LEM1) to a maximum of 15.0% (7L7).

Short direct repeats flank the 7SL RNA homologies (Figure 4). In the 3′-truncated pseudogenes (7L28 and 7LEM1) the direct repeats (13 and 15 nucleotides long, respectively) abut the 7SL DNA sequence (Figure 4B). We note that the direct repeats at the 3′ end overlap the sequence of the 7SL pseudogenes in both cases. In Figure 4C we have underlined the possible direct repeats flanking 7L63 DNA. However, due to the presence of several mismatches, these direct repeats are less convincing than the ones flanking the other pseudogenes.

The direct repeats flanking the 5′-truncated pseudogenes 7L23 and 7L7 do not abut the 7SL DNA sequence (Figure 4D). At the 3′ end 7L23 and 7L7 pseudogenes are followed by an A-rich sequence. The downstream direct repeat is adjacent to the A-rich sequence both in 7L23 and 7L7 DNA. The upstream direct repeat is located 15 (7L23) or 16 nucleotides (7L7) away from the 5′ end of both 7SL sequences. Interestingly, 12 of these 15 or 16 nucleotides in 7L23 and 7L7 are very similar (Figure 4D).

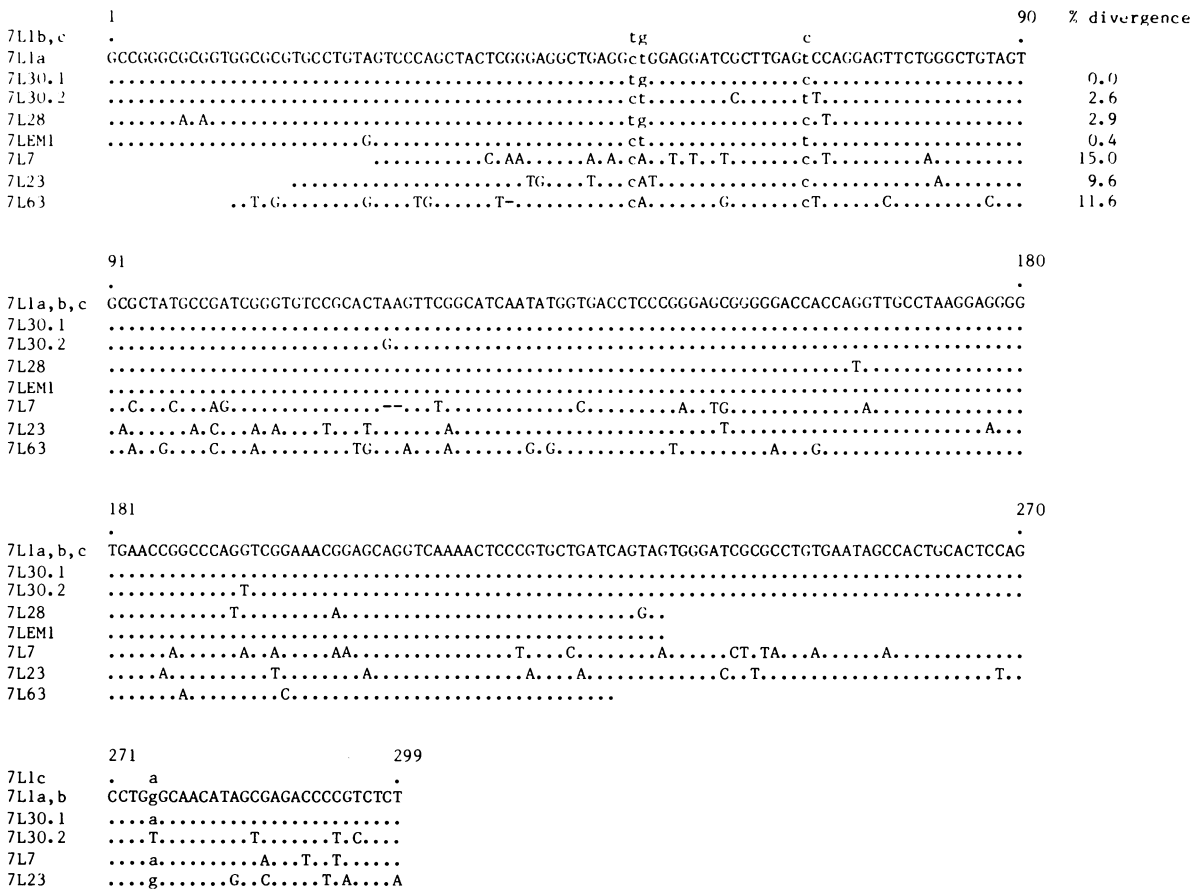

Fig. 3. Comparison of the sequences of human 7SL cDNAs with the sequences of seven human 7SL RNA homologies. The anti-coding strand of human 7SL cDNA clone 7L1a is shown at the top line. Lower case letters mark the base changes found in human 7SL cDNA clones 7L1b and 7L1c at positions 52, 53, 69 and 271 of the sequence (see text for details). The nucleotides present at the corresponding positions in the genomic 7SL DNAs are also indicated by lower case letters. Dot, homology with the corresponding base in 7L1a cDNA sequence; dash, deletion of a nucleotide.

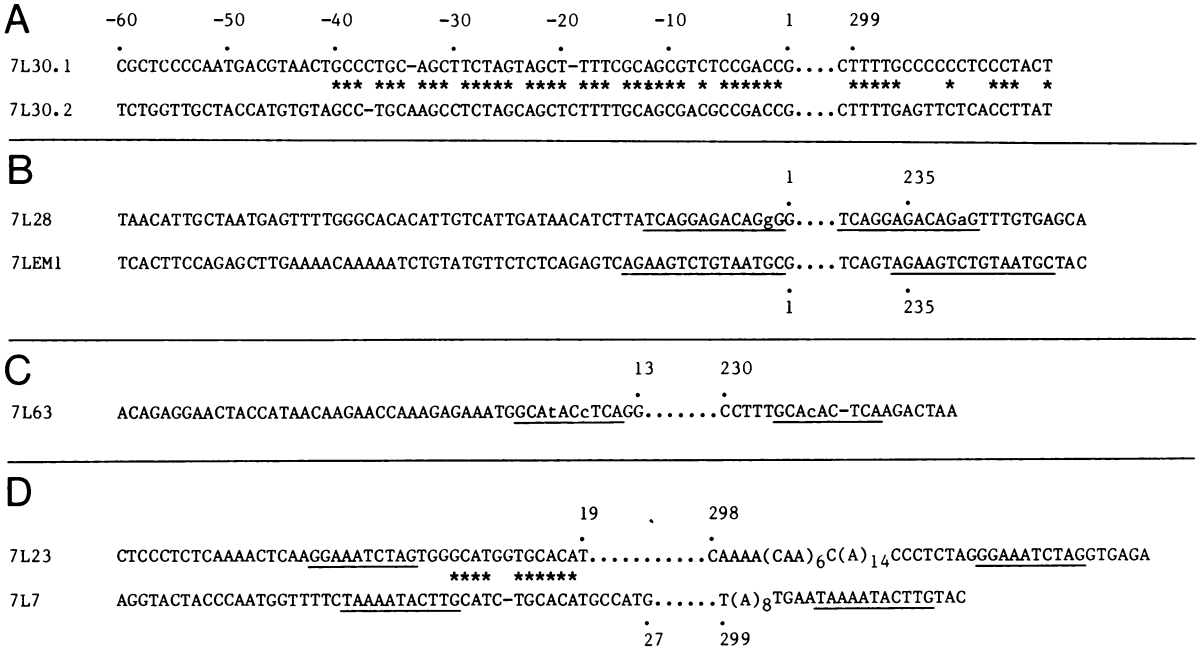

Fig. 4. The DNA sequences flanking seven human 7SL RNA homologies. Comparison of the DNA sequences flanking 7L30.1 and 7L30.2 coding sequences **(A)**, 3′-truncated pseudogenes 7L28 and 7LEM1 **(B)**, 5′- and 3′-truncated pseudogene 7L63 **(C)**, and 5′-truncated pseudogenes 7L23 and 7L7 **(D)**. A dotted line represents the 7SL DNA sequence of each recombinant clone as shown in Figure 3. Positive numbers mark the first and the last nucleotide of the 7SL RNA homologies. Sequence homologies are indicated by asterisks. Dashes have been occasionally introduced to maximise homology. Flanking direct repeats are underlined. Mismatches within each set of direct repeats are indicated by lower case letters.

The structure of human loci 7L30.1 and 7L30.2

Recombinant phage 7L30 contained two regions of homology to the S DNA probe, termed 7L30.1 and 7L30.2 (Figure 2). These S DNA homologies were located ~8-kb apart on the DNA and were oriented in the same direction. There was no obvious conservation of restriction sites surrounding the 7L30.1 and 7L30.2 loci, indicating that the genomic context of the two regions was different.

The nucleotide sequences of 7L30.1 and 7L30.2 are compared with human 7SL cDNA sequences in Figure 3. Both sequences are co-linear with the 7SL RNA molecule. The 7L30.1 sequence is identical to the 7L1c sequence variant. 7L30.2, although most closely related to the 7L1a sequence, has eight mismatches relative to it.

Figure 4A shows the comparison of the DNA sequences immediately flanking the 7SL DNA of 7L30.1 and 7L30.2. Upstream from the 7SL RNA coding sequences a stretch of 40 nucleotides appears to be almost perfectly conserved (see Discussion). At the 3′ end, four T residues, surrounded by G+C-rich sequences, follow the 7SL RNA coding sequences. Such a structure is known to be an efficient transcription terminator for RNA polymerase III (Bogenhagen and Brown, 1981), the RNA polymerase that transcribes 7SL RNA genes (Zieve *et al.*, 1977). Downstream from the T residues, the two sequences display only limited homology. Based on these observations, we concluded that both 7L30.1 and 7L30.2 DNAs had the structural features of 7SL RNA genes.

In vitro transcription of 7L30.1 and 7L30.2 DNAs

To ascertain whether 7L30.1 and 7L30.2 DNAs exhibit the functional features of 7SL RNA genes, we tested their ability to direct the synthesis of 7SL RNA in an *in vitro* transcription system prepared from HeLa cell nuclei (Dignam *et al.*, 1983). 7L30.1 and 7L30.2 DNAs were separately subcloned into the plasmid vector pUC13 (see Figure 2). When these DNAs were used as templates for *in vitro* transcription, products resembling 7SL RNA in mobility were obtained (Figure 5A); moreover, when the transcripts were eluted and rerun on a high-resolution sequencing gel, they co-migrated with ^{32}P-labelled 7SL RNA from HeLa cells (Figure 5B). The efficiency of expression of the 7SL DNA templates is comparable with that of other RNA polymerase III genes tested in the same *in vitro* system, such as the genes coding for EBER RNAs (Rosa *et al.*, 1981) (Figure 5A). Transcription of 7SL DNA was inhibited by concentrations of α-amanitin >200 μg/ml (data not shown). This confirms that RNA polymerase III transcribes 7SL DNA. We found that efficient expression of 7SL DNA requires a low concentration of Mg^{2+} ions. Depending on the extract preparation and on the extract dilution, the optimum Mg^{2+} concentration for 7SL DNA transcription ranged between 0 (no addition) and 2 mM. These values are significantly lower than those reported for other class III genes. In the same extract and under the same conditions, the optimum transcription rate of a human methionine tRNA gene (Santos and Zasloff, 1981) occurred at 5 mM Mg^{2+} (J. Stefano, personal communication). This indicated that the low Mg^{2+} requirement for *in vitro* transcription of 7SL DNA is template-dependent and it is not an anomalous property of the nuclear extract used.

The structure of the 7L30.1 and 7L30.2 transcripts

The 7SL RNA synthesised *in vitro* co-migrates with HeLa cell 7SL RNA. To characterise further the transcripts we subjected them to T1 ribonuclease fingerprint analysis. The T1 fingerprint of *in vivo* ^{32}P-labelled 7SL RNA from HeLa cells

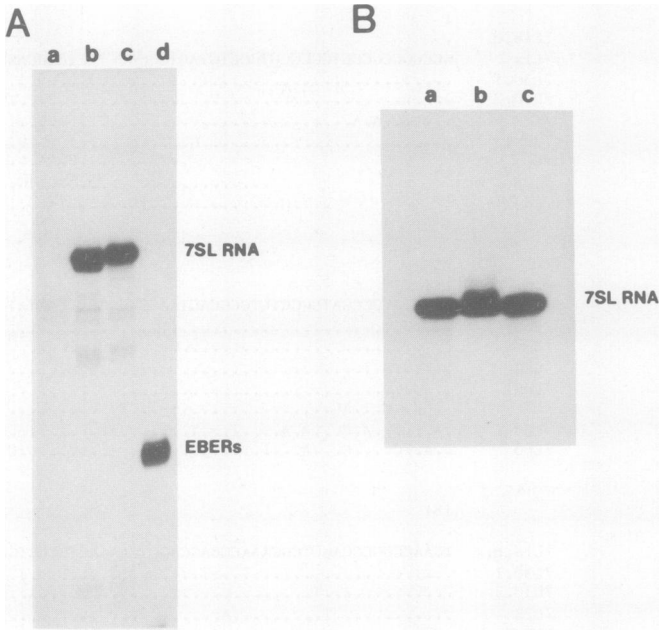

Fig. 5. *In vitro* transcription of 7L30.1 and 7L30.2 DNAs. **(A)** Plasmid subclones p7L30.1 **(lane a)** and p7L30.2 **(lane c)**, were transcribed in a HeLa cell nuclear extract, as described in Materials and methods. The transcription products were fractionated on a 6% polyacrylamide gel containing 7 M urea. **Lane a** shows the transcripts produced by incubation of the transcription extract with pUC13 plasmid vector DNA as template. The products of transcription of a plasmid containing the Epstein Barr virus EBERI and EBERII genes are shown in **lane d. (B)** Approximately equal amounts of gel purified 7L30.1 **(lane a)** and 7L30.2 **(lane c)** transcripts were electrophoresed on a 6% polyacrylamide sequencing gel (40 cm x 20 cm x 0.3 cm) until the xylene cyanol blue dye migrated 1.5 times the length of the gel. Gel purified 7SL RNA from HeLa cells was labelled at the 5′ end with [γ-^{32}P]ATP and polynucleotide kinase and used as a size marker.

is shown in Figure 6B and is compared with the T1 fingerprints of the 30.1 (Figure 6A) and 30.2 (Figure 6C) RNAs, synthesised *in vitro* using [α-^{32}P]GTP and [α-^{32}P]UTP as radioactive precursors. The fingerprint of the 30.2 transcript differs significantly from the fingerprint of HeLa cell 7SL RNA. In particular, the T1 oligonucleotide indicated by an arrow in Figure 6C is not found in any of the 7SL RNA variants labelled *in vivo*. From this result we concluded that the 7L30.2 locus is not expressed at a detectable level in HeLa cells.

On the other hand, the 30.1 transcript appears to be a genuine human 7SL RNA molecule, although the analysis is somewhat complicated by the existence of three human 7SL RNA sequence variants. 7L30.1 DNA represents the 7L1c sequence variant in which a T replaces a C residue at position 69 of the 7L1a sequence (Figure 3). This single base substitution changes the sequence of a T1 oligonucleotide from UCCAGp in 7L1a RNA to CCCAGp in 7L1c RNA. Consequently, T1 digestion of the 30.1 transcript generates the CCCAGp oligonucleotide (indicated by an open arrow in Figure 6A) but not the UCCAGp oligonucleotide (indicated by an open arrow head in Figure 6B). The 5′ end of HeLa cell 7SL RNA and of the 30.1 RNA synthesised *in vitro* was determined to be pppG by 5′ end group analysis (data not shown).

T1 oligonucleotides derived from the 3′ ends of HeLa cell and 30.1 7SL RNA are indicated by solid arrow heads in Figure 6. The relative position of the spots indicates that both RNAs have heterogeneous termini. In 30.1 RNA, 3′ end heterogeneity simply reflects the presence of one or two ter-

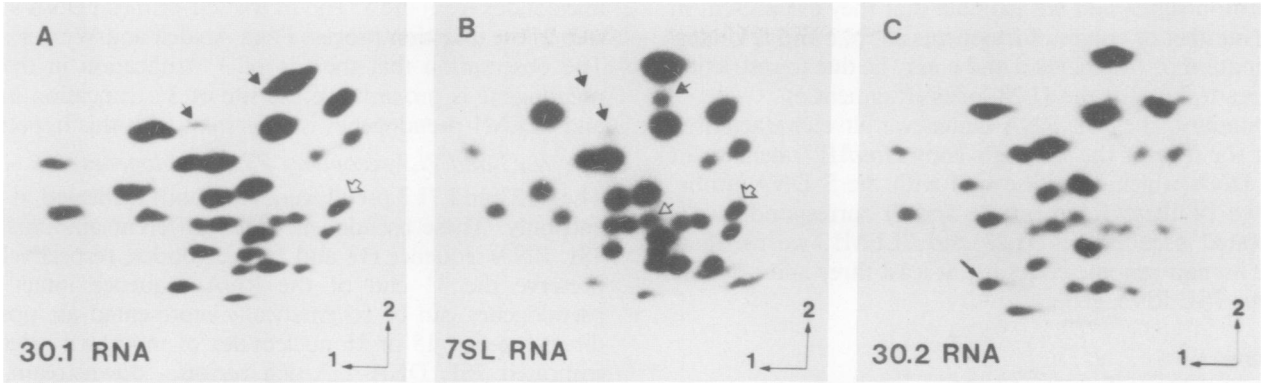

Fig. 6. T1 RNase fingerprint analysis of human 7SL RNA and of 7L30.1 and 7L30.2 transcripts. 7L30.1 and 7L30.2 RNAs were synthesised *in vitro* as described in Materials and methods and fractionated on a 6% polyacrylamide gel. *In vivo* ^{32}P-labelled 7SL RNA was isolated from HeLa cells and purified by hybridisation to p7L1 plasmid DNA bound to a nitrocellulose filter. The hybridised RNA was eluted from the filter and fractionated by electrophoresis on a 6% polyacrylamide gel. Each gel-purified RNA species was digested with ribonuclease T1 and the T1 oligonucleotides were separated by electrophoresis on Cellogel strips (horizontal dimension) followed by homochromatography on PEI thin-layer plates (vertical dimension). An open arrow head indicates the oligonucleotide UCCAGp; an open arrow indicates oligonucleotide CCCAGp. In **panel C** an arrow points to a T1 oligonucleotide which is unique to 30.2 RNA. T1 oligonucleotides derived from the 3′ ends of 30.1 RNA and of HeLa cell 7SL RNA are indicates by solid arrow heads.

minal U residues, as often observed in RNA polymerase III primary transcripts. In contrast, secondary analysis (Ullu and Weiner, in preparation) of the T1 oligonucleotides derived from the 3′ ends of 7SL RNA synthesised *in vivo* suggested a more complex set of structures: $UCUC(U)_{1-2}OH$ and $UCUC(U)_{1-2}A_{OH}$. Thus, 3′ end heterogeneity of human 7SL RNA reflects both a variable number of terminal U residues as well as the presence or absence of one A residue at the extreme 3′ end. A similar finding was reported for rat 7SL RNA (Li *et al.*, 1982). In addition, dog 7SL RNA, isolated from SRP, was shown to terminate with one A residue (Walter and Blobel, 1982). Because HeLa cell 7SL RNA is a mixture of at least three sequence variants, we cannot establish whether each variant has a different, but homogeneous, terminus or whether 3′ end heterogeneity is common to all three RNA species. However, the presence of a 3′-terminal A residue in 7SL RNA from different organisms strongly suggests that 7SL RNA is post-transcriptionally modified by adenylation. An alternative but much less likely possibility is that the canonical RNA polymerase III terminator, a cluster of Ts, can be interrupted by a single A residue.

The human genome contains only a few 7SL RNA genes

The abundance of 7SL RNA genes in human DNA was determined by gene-counting experiments using recombinant clones 7L30, 7LEM1, 7L28 and 7L3 (an independent clone of the 7L30.2 locus) as hybridisation standards. Figure 7 shows an example of such an experiment. A one copy equivalent of each 7L recombinant phage DNA was digested with *Hind*III (7LEM1, 7L28 and 7L3) or with *Hind*III plus *Eco*RI (7L30) and co-electrophoresed on an agarose gel with a *Hind*III digest of human placental DNA (lanes P). The various DNAs were then transferred to a nitrocellulose filter and hybridised with S DNA at 55°C in 50% formamide, 5 x SSC. Under these hybridisation conditions only perfect or nearly perfect S sequences can be detected. Six prominent hybridising bands (a–f) can be observed in placental DNA. The placental *Hind*III fragments b, c, e and f co-migrate with the corresponding restriction fragments of 7L30.1, 7L30.2, 7L28 and 7LEM1 recombinants and hybridise with single-copy intensity. DNA fragments d and e consistently gave a reduced

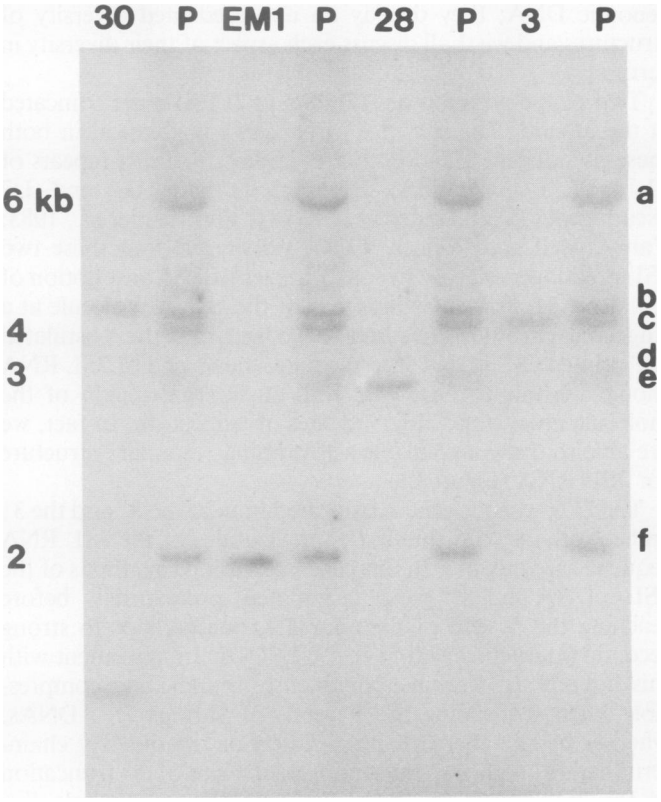

Fig. 7. The human genome codes only for a few 7SL RNA genes. Human placental DNA (10 μg) was digested with *Hind*III, fractionated by electrophoresis on a 0.7% agarose gel, blotted onto a nitrocellulose filter and hybridised to ^{32}P-labelled S DNA (**lanes P**), as a described in Materials and methods. To estimate the 7SL RNA gene copy number, the hybridisation intensity of the S DNA probe to a one copy-equivalent of various 7L recombinant phage DNAs was used as a reference. 7LEM1, 7L28 and 7L3 DNAs (**lanes EM1, 28** and **3**) were digested with *Hind*III, while 7L30 DNA (**lane 30**) was digested with *Hind*III plus *Eco*RI. Two bands of hybridisation are detected in **lane 30**. The slow migrating band corresponds to the 4.3-kb *Hind*III DNA fragment which spans the 7L30.1 locus, while the fast migrating band represents the 1.3-kb *Eco*RI DNA fragment derived from the 7L30.2 locus (see Figure 2). Lower case letters indicate the six most prominent human DNA fragments which consistently hybridised to the S DNA probe in different experiments.

hybridisation signal and we estimate that they are present in half the number of copies of fragments aa, b, c and f. Underrepresentation of fragments d and e may be due to restriction site polymorphism at the 7L28 locus (fragment e).

In summary, the 7SL RNA clones we have characterised account for four of the six single-copy *Hind*III fragments of human DNA which hybridise well with the S DNA probe. Since two of these fragments (e and f) correspond to the 3′-truncated pseudogenes 7L28 and 7LEM1, we conclude that the human genome codes for at least three and no more than four 7SL RNA genes.

Discussion

Human 7SL pseudogenes might be generated by a variety of mechanisms

We have shown that the central S sequence of 7SL RNA is repeated several hundred times in the human genome. The majority of these S sequences are 7SL RNA pseudogenes having only limited homology with 7SL RNA. All 7SL pseudogenes characterised appear to be generated by the reverse flow of genetic information from 7SL RNA back into genomic DNA; they display an unprecedented diversity of structures and we shall discuss each aspect of their diversity in turn.

Two of the pseudogenes (7L28 and 7LEM1) are truncated at the 3′ end. The site of 3′ truncation is identical in both these pseudogenes, and both are flanked by direct repeats of 13 and 15 nucleotides. By analogy with U3 and U2 pseudogenes (Van Arsdell *et al.*, 1981; Bernstein *et al.*, 1983; Van Arsdell and Weiner, 1984), we suggest that these two 7SL pseudogenes arose by self-primed reverse transcription of 7SL RNA, followed by insertion of the cDNA molecule at a staggered chromosomal break. To generate the postulated self-primed 7SL cDNA intermediate, the 3′ end of 7SL RNA should be able to base pair with an internal region of the molecule near to the observed sites of truncation. In fact, we are able to draw a plausible self-priming secondary structure for 7SL RNA (Figure 8).

The 7L63 pseudogene is truncated both at the 5′ and the 3′ ends. At the 5′ end the first 12 nucleotides of the 7SL RNA sequence are missing. In this case, self-primed synthesis of the 7SL cDNA might have terminated prematurely before reaching the 5′ end of the template, due perhaps to strong secondary structure within the 7SL RNA. In agreement with this hypothesis, we have consistently found band compression when sequencing the 5′ ends of various 7SL DNAs, whether by the chemical degradation or the dideoxy chain-termination methods. Interestingly, the site of 3′ truncation of the 7SL homology in the 7L63 pseudogene is only five nucleotides away from the site of truncation in the 7L28 and 7LEM1 pseudogenes discussed above. We previously suggested that heterogeneous sites of 3′ truncation, as observed in U2 RNA pseudogenes, could reflect a variable loss of nucleotides from the 5′ end of the self-primed cDNA at some step in the insertion process (Van Arsdell and Weiner, 1984). The observation that the site of 3′ truncation in the 7L63 pseudogene is proximal to the site of 3′ truncation in 7L28 and 7LEM1 pseudogenes is consistent with this hypothesis.

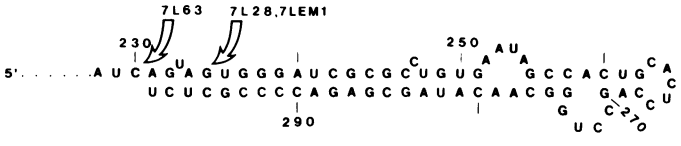

Fig. 8. A hypothetical secondary structure for the 3′ end of human 7SL RNA. A potential folding for the 3′ end of 7SL RNA is shown. The structure has a calculated free energy of formation of −26.5 kcal/mol (Tinoco *et al.*, 1973; Salser, 1977). Arrows point to the sites of truncation found in 3′-truncated 7SL pseudogenes.

Are 7L23 and 7L7 secondary 7SL pseudogenes?

The 7L23 and 7L7 pseudogenes are both truncated at the 5′ end only. These pseudogenes lack different amounts of 5′ 7SL RNA sequence (18 and 26 nucleotides, respectively) but preserve the 3′ end of the RNA sequence intact. Both pseudogenes can be schematically represented as: upstream direct repeat - 15 or 16 nucleotides of unknown origin - 5′ truncated 7SL DNA - A-rich region - downstream direct repeat. The insertion of a reverse transcript of 7SL RNA cannot account for the existence in these two pseudogenes of 15 or 16 extra nucleotides between the upstream direct repeat and the beginning of the homology with 7SL RNA. Rather, the structure of these two pseudogenes suggests that the extra 15 or 16 nucleotides were already part of the cDNA at the time of insertion.

We suggest that 7L7 and 7L23 pseudogenes represent secondary pseudogenes derived by reverse transcription of an aberrant 5′-substituted 7SL RNA which is itself transcribed from a primary 5′-truncated pseudogene. Because 7SL RNA genes contain an internal promoter for RNA polymerase III (our unpublished observations), 7SL RNA pseudogenes which preserve the internal control region might still function as RNA polymerase III transcription units. Internal control regions in both tRNA (Hofstetter *et al.*, 1981; Sharp *et al.*, 1981) and 5S RNA (Sakonju *et al.*, 1980) genes instruct the polymerase to initiate at a fixed distance upstream, so RNA polymerase III transcription of a 5′-truncated 7SL pseudogene would produce a 5′-substituted 7SL RNA. In this case, the extra 15 or 16 nucleotides would be derived from whatever upstream sequences happened to flank the primary 5′-truncated 7SL pseudogenes.

Transcription of tRNA genes (Hofstetter *et al.*, 1981; Sharp *et al.*, 1981) has been shown to be directed by two intragenic promoter regions, designated the A and B boxes. In both 7L23 and 7L7 DNAs the homology to the A box of the tRNA promoter consensus sequence (Traboni *et al.*, 1982), located between positions 5 and 13 of 7SL DNA, is lost, and we would expect these 5′-deleted 7SL pseudogenes to be transcriptionally inactive. The presence of a very similar 15 or 16 nucleotide long substitution at the 5′ end of 7L23 and 7L7 pseudogenes (Figure 4D) might suggest that this sequence has functionally replaced the A box of the intragenic promoter in the primary 5′-truncated 7SL pseudogene. Alternatively, it is possible that the transcription of the primary 7SL pseudogene might occur in the absence of the A box sequences and we are currently testing this hypothesis.

The 7SL RNA sequence is a composite corresponding to an Alu monomer interrupted by the S sequence which is unique to 7SL RNA (Ullu *et al.*, 1982). The production of secondary 7SL pseudogenes depends on the internal promoter for RNA polymerase III, presumably located within the 5′ Alu portion of the 7SL sequence; this same internal promoter is responsible for the ability of Alu elements to transpose through a cDNA intermediate as proposed by ourselves (Van Arsdell *et al.*, 1981) and others (Jagadeeswaran *et al.*, 1981). We have also proposed, on the basis of extensive sequence homology between the 7SL RNAs of humans, *Xenopus laevis* and *Drosophila melanogaster*, that the mammalian Alu sequence

family was derived from 7SL RNA by deletion of the central S sequence (Ullu and Tschudi, 1984). Thus, we could think of Alu sequences as an extended family of 7SL pseudogenes, in which one (or more) primary Alu sequence(s) has given rise to an extensive array of secondary, tertiary and higher order Alu elements.

7L30.1 is an authentic human 7SL RNA gene

We have found that the human genome contains no more than four 7SL RNA genes. We have also isolated and characterised a human genomic DNA fragment which carries two linked 7SL RNA coding sequences, the 7L30.1 and 7L30.2 loci. The 7L30.1 locus qualifies as an authentic 7SL RNA gene by several criteria: (i) the DNA sequence agrees perfectly with one of the three sequence variants of human 7SL cDNA; (ii) the cloned DNA can be efficiently transcribed *in vitro* to produce an RNA identical to HeLa cell 7SL RNA; (iii) the 7L30.1 locus represents one of the six most prominent S DNA-containing fragments in a Southern blot of human genomic DNA.

The 7L30.2 locus has all the features of a 7SL RNA gene but does not appear to be expressed at a detectable level in HeLa cells. 7L30.2 may represent a gene which has accumulated deleterious mutations, possibly affecting the stability and/or the packaging of the RNA in SRP, or a gene whose expression may be regulated in the cell.

Clustering and homology of 7SL RNA genes

The structure of recombinant clone 7L30 indicates that 7SL genes can be clustered in the chromosome, but we do not know whether the other human 7SL RNA genes are closely linked to the 7L30 locus. Genes coding for similar but not identical RNAs are often clustered. Examples include different tRNA genes, as observed in various organisms (Hovemann *et al.*, 1980; Sekiya *et al.*, 1982), and the human genes for the Ro RNAs (Wolin and Steitz, 1983). We presume that the 7L30.1 and 7L30.2 genes, as suggested for these other multicopy genes, have arisen by duplication and divergence of a single ancestral gene.

Unlike other polymerase III genes such as tRNA and Ro RNA genes, we find that the homology between the 7L30.1 and 7L30.2 loci extends beyond the RNA coding sequence as far upstream as position −40. This sequence conservation might indicate that the 7L30.1 and 7L30.2 DNAs were generated by a relatively recent gene duplication. Alternatively, a gene conversion could have homogenised the 5′-flanking sequences of the two 7SL loci after the duplication of the ancestral 7SL RNA gene. We favor a different explanation: the observed sequence conservation could reflect functional constraints acting on the genes. This would be consistent with our observation that the DNA sequences upstream from the site of transcription initiation are required for efficient and accurate expression of the 7SL RNA gene *in vitro* (Ullu and Weiner, in preparation). These conserved sequences might enable the transcriptional apparatus to distinguish a true 7SL RNA gene from those 7SL RNA pseudogenes which retain intact promoter sequences.

Materials and methods

The S DNA probe

S DNA, the central 125-bp *Sau*3A restriction DNA fragment of 7SL cDNA clone p7L1 (Ullu *et al.*, 1982), was ligated to *Eco*RI linkers and inserted at the *Eco*RI site of plasmid vector pBR322. For use as a probe S DNA was excised from the vector DNA by digestion with the *Eco*RI restriction enzyme, fractionated by electrophoresis through 1.5% agarose gel, electroeluted and purified by chromatography on DEAE-52 cellulose. S DNA fragments were ligated by using T4 DNA ligase and labelled with ^{32}P by nick-translation to a specific activity of $5-10 \times 10^7$ c.p.m./μg.

Cloning and DNA sequencing

Human genomic phage libraries were screened with the S DNA probe by the procedure of Benton and Davis (1977). Hybridisations were carried out in 50% formamide, 5 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate) at either 40°C or 55°C. Recombinant phages 7L30, 7L7, 7L23 and 7L63 were isolated from the human phage library of Lawn *et al.* (1978). To isolate recombinant phage 7LEM1 a library of 20-kb partial *Sau*3A DNA fragments of human placental DNA was constructed in the lambda vector EMBL4 (Frischauf *et al.*, 1983). Recombinant phages 7L3 and 7L28 were obtained by cloning human placental *Hind*III restriction DNA fragments, ~4 kb and 3 kb in length, respectively, in the lambda vector 590 (Murray *et al.*, 1977). For sequence analysis, restriction fragments from individual 7L recombinant phages were subcloned either in the M13 vectors, mp8 and mp9, or in the plasmid vectors pUC8 and pUC13 (Vieira and Messing, 1982). Both the chemical degradation (Maxam and Gilbert, 1980) and the dideoxy-chain termination (Sanger *et al.*, 1977) methods were used for sequence analysis. All experiments involving recombinant molecules were performed under P2 + EK2 conditions as specified by the revised National Institutes of Health Guidelines for Recombinant DNA Research.

Genomic blots

Genomic blots were prepared by the method of Southern (1975). Nitrocellulose filters were hybridised for 16 h either at 40°C or at 55°C in a solution containing 50% formamide, 5 x SSC, 2 x Denhardt's solution (1 x Denhardt's = 0.1% bovine serum albumin, 0.1% ficoll, 0.1% polyvinylpyrrolidone), 10% dextran sulphate and ~100 000 c.p.m./ml of ^{32}P-labelled S DNA. Carrier DNA was omitted from the hybridisation mixture. Following hybridisation, the filters were washed in 2 x SSC 0.1% SDS at the hybridisation temperature.

In vitro transcription

A HeLa cell nuclear extract was prepared by the method of Dignam *et al.*, (1983). For transcription of 7SL DNA templates, a standard 20 μl reaction contained 50% extract, 50 μg/ml plasmid DNA, 80 mM KCl, 0.5 mM $MgCl_2$, 12 mM Hepes-KOH pH 7.9, 12% glycerol, 0.5 mM DTT, 0.5 mM ATP, UTP and CTP, 25 μM GTP and 10 μCi of [α-^{32}P]GTP (400 Ci/mmol). For transcription of plasmid pEBVRIJ, containing the genes for EBERI and EBERII RNAs (kindly provided by E. Gottlieb), the $MgCl_2$ concentration was raised to 5 mM. Transcription reactions were incubated at 30°C for 2 h. At the end of the incubation reaction mixtures were digested with proteinase K (200 μg/ml) in the presence of 0.5% SDS for 10 min at 60°C. Following addition of one volume of 4 M ammonium acetate pH 5.0, nucleic acids were precipitated with three volumes of ethanol and collected by centrifugation. The products of transcription were fractionated on a 6% polyacrylamide gel containing 7 M urea.

RNA analyses

To label 7SL RNA, HeLa cells were incubated with $^{32}PO_4$, as described by Wolin and Steitz (1983). For T1 fingerprinting ^{32}P-labelled 7SL RNA was purified by hybridisation to p7L1 DNA immobilised on nitrocellulose filters as previously described (Ullu and Melli, 1982). ^{32}P-Labelled 7SL RNA and RNAs synthesised *in vitro* were fractionated on a 6% polyacrylamide-7 M urea sequencing gel. Individual bands were excised from the gel and the RNA was extracted by soaking the gel slices in ~5 volumes of a solution containing 0.5 M KCl, 50 mM Tris-HCl pH 7.5. T1 fingerprints were performed as described by Barrell (1971). For 5′ end-group analysis, the RNAs were digested with P1 nuclease followed by separation on DEAE paper at pH 3.5. The 3′ ends of the RNAs were analysed by digesting purified T1 oligonucleotides with T2, P1 and pancreatic ribonucleases. The digestion products were resolved either by chromatography on PEI thin-layer plates using 1.0 M LiCl as solvent or by two-dimensional chromatography on Avicel plates as described by Wise and Weiner (1981).

Acknowledgements

We wish to thank Ellen Gottlieb, Carl Hashimoto, Steve Mount, Sandy Wolin and Chris Tschudi for much helpful advice; Ellen Gottlieb for sharing her transcription extracts and providing HeLa cells. E.U. was the recipient of a long-term EMBO fellowship and is on leave of absence from the Istituto di Biologia Generale, Universitá di Roma. This Research was supported by a grant from the National Institutes of Health.

References

Barrell,B.G. (1971) *Progr. Nucleic Acids Res. Mol. Biol*, **8**, 751-779.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.

Bernstein,L.B., Mount,S. and Weiner,A.M. (1983) *Cell*, **26**, 461-472.
Bogenhagen,D.F. and Brown,D.D. (1981) *Cell*, **24**, 261-270.
Deininger,P.L., Jolly,D.J., Rubin,C.M., Friedmann,T. and Schmid,C.W. (1981) *J. Mol. Biol.*, **151**, 17-33.
Dignam,J.D., Russell,M.L. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475-1489.
Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Gundelfinger,E.D., Krause,E., Melli,M. and Dobberstein,B. (1983) *Nucleic Acids Res.*, **11**, 7363-7374.
Hofstetter,H., Kressmann,A. and Birnstiel,M.L. (1981) *Cell*, **24**, 573-585.
Hovemann,B., Sharp,S.J., Yamada,H. and Söll,D. (1980) *Cell*, **19**, 889-895.
Jagadeeswaran,P., Forget,B.G. and Weissman,S.M. (1981) *Cell*, **26**, 141-142.
Jelinek,W.R. and Schmid,C.W. (1982) *Annu. Rev. Biochem.*, **51**, 813-844.
Lawn,R.M., Fritsch,E.F., Parker,R.C., Blake,G. and Maniatis,T. (1978) *Cell*, **15**, 1157-1174.
Li,W., Reddy,R., Henning,D., Epstein,P. and Busch,H. (1982) *J. Biol. Chem.*, **257**, 5136-5142.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Murray,N., Brammar,W.J. and Murray,K. (1977) *Mol. Gen. Genet.*, **150**, 53-61.
Rosa,M.D., Gottlieb,E., Lerner,M.R. and Steitz,J.A. (1981) *Mol. Cell. Biol.*, **1**, 785-796.
Sakonju,S., Bogenhagen,D.F. and Brown,D.D. (1980) *Cell*, **19**, 13-25.
Salser,W. (1977) *Cold Spring Harbor Symp. Quant. Biol.*, **42**, 985-1002.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Santos,T. and Zasloff,M. (1981) *Cell*, **23**, 699-709.
Sekiya,T., Nishizawa,R., Matsuda,K., Taya,Y. and Nishimura,S. (1982) *Nucleic Acids Res.*, **10**, 6411-6419.
Sharp,S., DeFranco,D., Dingermann,T., Farrell,P. and Söll,D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6657-6661.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40-41.
Traboni,C., Ciliberto,G. and Cortese,R. (1982) *EMBO J.*, **1**, 415-420.
Ullu,E. and Melli,M. (1982) *Nucleic Acids Res.*, **10**, 2209-2223.
Ullu,E. and Tschudi,C. (1984) *Nature*, in press.
Ullu,E., Murphy,S. and Melli,M. (1982) *Cell*, **29**, 195-202.
Van Arsdell,S.W. and Weiner,A.M. (1984) *Nucleic Acids Res.*, **12**, 1463-1471.
Van Arsdell,S.W., Denison,R.A., Bernstein,L.B., Weiner,A.M., Manser,T. and Gesteland,R.F. (1981) *Cell*, **26**, 11-17.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 269-276.
Walter,P. and Blobel,G. (1982) *Nature*, **299**, 691-698.
Weiner,A.M. (1980) *Cell*, **22**, 209-218.
Wise,J.A. and Weiner,A.M. (1981) *J. Biol. Chem.*, **256**, 956-963.
Wolin,S.L. and Steitz,J.A. (1983) *Cell*, **32**, 735-744.
Zieve,G., Benecke,B. and Penmann,S. (1977) *Biochemistry (Wash.)*, **16**, 4520-4525.

Received on 1 October 1984