Volume 14 Number 23 1986 Nucleic Acids Research

A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes

Shona Murphy*, M.Tripodi[1] and Marialuisa Melli*

Sclavo Research Centre, Via Fiorentina 1, Siena and [1]Istituto di Biologia Cellulare, Dipartimento di Biopatologia Umana, Policlinico Umberto I, Rome, Italy

Received 21 October 1986; Accepted 6 November 1986

ABSTRACT

We have isolated and characterized two recombinant λ phages containing sequences homologous to 7SK RNA which code for a RNA 330 nucleotides long in an "in vitro" transcription system. S1 mapping of the transcript shows that this RNA corresponds to the 7SK RNA obtained from human cells, indicating that the two recombinant phages contain genes coding for 7SK RNA. The transcription of these genes is polymerase III dependent. Sequences upstream from the start of transcription are essential for "in vitro" synthesis of 7SK RNA, suggesting that internal promoter elements, if present, are not sufficient to support the synthesis of 7SK RNA. A region of homology with the upstream sequences of the genes for U6 RNA (1), 7SL RNA (2) and Bombyx mori alanine tRNA (3) is found within 50 bp from the transcription start point. Within the homologous region a motif common to the four genes is a "TATA"-like box, placed at position -30 to -25 of the 7SK RNA gene, which is typical of the polymerase II promoter region (4).

INTRODUCTION

Small RNAs, usually in the form of ribonucleoprotein particles (RNPs), perform a wide range of critical functions in every cell. For instance, members of the U family of RNPs are involved in post-transcriptional processing of RNA (5,6) and the 7SL RNP or signal recognition particle is involved in transport of secreted proteins across the endoplasmic reticulum (7). The 7SK RNA is an abundant RNA, 330 nucleotides long, which is also found complexed with proteins, in the form of a 12S RNP (E. Gundelfinger, D. Zopf and M. Melli, unpublished results). Like the 7SL RNA, the 7SK RNA is transcribed by RNA polymerase III (8). Although the precise cellular location and function of this RNP remains unknown, the ubiquity and high evolutionary conservation of the RNA suggest that also this RNP has a fundamental role in the cell (9). The RNA is homologous to a large family of repeated sequences which is composed

mainly of 3' truncated pseudogenes which appear to have been generated through a RNA intermediate. We have shown that there are relatively few full length copies of the RNA in the human genome and these must include the genes (10). The two recombinant DNAs described here contain a full length copy of 7SK RNA and direct the synthesis of 7SK RNA by polymerase III in an "in vitro" transcription system, strongly suggesting that they are genes coding for the 7SK RNA. Unlike most other eukaryotic genes transcribed by polymerase III, they require sequences located upstream from the coding region for transcription. These sequences contain a region of homology to the 5' flanking region of the genes for 7SL RNA, which is necessary for efficient transcription of the gene (2) and to the upstream region of a human gene for U6 RNA (1), suggesting that these three genes share a common transcriptional control region. Homology is also found with upstream sequences required for transcription of the alanine tRNA gene of Bombyx mori (3). All four genes have in common a "TATA"-like box placed in the same position as the polymerase II promoter sequence (4).

MATERIALS AND METHODS

Radionucleotides and enzymes

Restriction endonucleases, Escherichia coli DNA polymerase I (Klenow enzyme), polynucleotide kinase, exonuclease Bal 31, exonuclease S1 and alpha-amanitin were obtained from Boehringer Mannheim GmbH. RNA ligase was obtained from PL Biochemicals and radionucleotides from New England Nuclear.

Cloning and DNA sequencing

Two human genomic DNA λ phage libraries were screened with a DNA probe corresponding to the 3' end of 7SK cDNA (probe C, described by Murphy et al. (10)). The positive DNA recombinants were hybridized to a DNA probe corresponding to the 5' end of full length 7SK cDNA (probe A described by Murphy et al. (10)) and clones hybridizing to both probes were selected for further analysis. Clone 7SK45 was isolated from a library of human genomic DNA in the λ phage Charon 4a kindly provided by Dr. R. Lawn (11) and clone 7SK33 was isolated from a library of human genomic DNA in the λ phage EMBL4 kindly provided by Dr. R. Cortese (12). For sequence determination, restriction

fragments were subcloned into M13mp18 and M13mp19 phage vectors (13) or into pEMBL18 and pEMBL19 plasmid vectors (14), and sequencing was performed by the method of Sanger et al. (15).

Radioactive DNA probe preparation and hybridization

Radioactive DNA probes were prepared by end labelling electrophoretically purified DNA fragments with either (γ-^{32}P) ATP and the polynucleotide kinase enzyme, or (α-^{32}P) dNTPs and the Klenow enzyme. The specific activity of the DNA probes was 1 x 10^8 to 2 x 10^8 cpm/ug of DNA. Hybridization was carried out in 10x Denhart's solution (2 g polyvinylpyrrolidone/l, 2 g bovine serum albumin/l, 2 g Ficoll/l), 0.1% (w/v) sodium dodecylsulphate (SDS) and 2x SET (0.3 M NaCl, 40 mM Tris HCl (pH 7.5), 0.2 mM EDTA) at 70°C. After hybridization, the filters were washed in 2 x SET, 0.1% SDS at the hybridization temperature.

In vitro transcription

HeLa cell cytoplasmic (S100) and nuclear extracts were made by the method of Dignam et al. (16). The standard transcription mix contained: 25% nuclear extract, 25% cytoplasmic extract, 75 mM KCl, 2 mM $MgCl_2$, 20 mM Tris-HCl pH 7.9, 0.5 mM dithiothreitol, 10% glycerol, 0.5 mM ATP, UTP, CTP, 25 uM GTP, 10 uCi of (α-^{32}P)GTP (760 Ci/mM), 1 ug/ml alpha-amanitin and 5-20 ug/ml template DNA in a 25 ul reaction volume. Reactions were initiated by the addition of extract and incubated at 30°C for 1 hour. At the end of the incubation, reaction mixtures were digested with DNAse I (100 ug/ml final concentration) for 5' and 125 ul of 0.2 M sodium acetate solution, 0.5% SDS, 40 ug/ml E. coli tRNA were added. The mixtures were extracted twice with an equal volume of phenol/H_2O and twice with two volumes of $CHCl_3$ and the nucleic acids were precipitated with two volumes of ethanol. The resulting precipitates were collected by centrifugation, resuspended in 99% formamide, 0.1% bromophenol blue, 0.2% Xylene cyanol, heat denatured and electrophoresed on 6% polyacrylamide 7M urea gels. The gels were usually exposed overnight to Kodak X-OMAT XAR-5 film. The plasmid XBS1 which was used as a transcription control contains a Xenopus borealis somatic 5S RNA gene (17). A Caenorhabditis elegans leucine transfer RNA gene (18) which was also used as a template for transcription, was kindly provided by G. Ciliberto. All plasmid DNAs used as

transcription templates were purified on caesium chloride gradients and the DNA was used in the supercoiled form.

S1 mapping

S1 mapping of the "in vitro" transcripts was carried out essentially as described by Maniatis et al. (19). The hybridization reactions contained 10 µg of carrier E. coli tRNA and either 500 ng of DNA, for the single stranded M13 DNA probes or 200 ng of DNA, for the double stranded DNA probes. The hybridization mixtures were denatured at 80°C for 10' and annealed at 58°C overnight. S1 digestion was carried out at 30°C with 500 units of S1 exonuclease/ml. After digestion the reaction mixtures were extracted with an equal volume of phenol/H_2O, the nucleic acids were precipitated with 2 volumes of ethanol and processed as for the "in vitro" transcription products.

Construction of the subclones used for transcription

The EcoRI-SalI restriction fragment of the λ clone 7SK33 which contains the 7SK RNA coding region was subcloned into the plasmid vector pEMBL 8 using the restriction enzyme sites for EcoRI and SalI present in the polylinker (20). All other subclones were derived from this construction by using appropriate restriction enzyme sites and insertion into the polylinker of pEMBL 8 (33b, 33C, 33D) or pEMBL19 (33E and 33F). The subclones of the λ clone 7SK45 were obtained in the same way. All constructions were checked by restriction enzyme analysis and limited sequencing.

Labelling of "in vivo" 7SK RNA

"In vivo" 7SK RNA was obtained from a cytoplasmic fraction of HeLa cells enriched for the 7SK ribonucleoprotein and was a kind gift from E. Gundelfinger. The RNA was labelled with (^{32}P)pCp as described by England and Uhlenbeck (21).

RESULTS

Structural organization of 7SK33 and 7SK45 recombinant clones

Figures 1 and 2 show the organization and nucleotide sequence of two putative genes for human 7SK RNA: 7SK33 and 7SK45. The recombinant clones differ by 4 single base changes outside the region coding for 7SK RNA. In addition 7SK45 has a 181 bp tandem duplication of the sequence from base 171 to 351, which is not present in 7SK33. In the 7SK33 clone the 7SK RNA coding region is closely

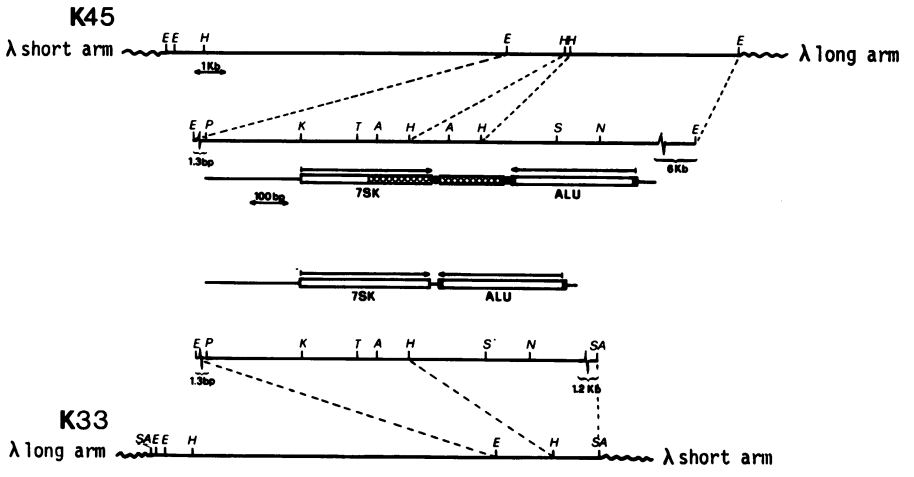


Figure 1
Restriction maps of the two recombinant DNA clones 7SK45 and 7SK33. The sequence homologous to 7SK RNA and the adjacent Alu DNA are boxed. The arrows above the 7SK RNA homology and the Alu sequence indicate the presumptive direction of transcription. In the map of 7SK45, hatching indicates the duplication of the presumptive 7SK RNA coding region and filled circles the duplication of the 3' non coding sequences. The direct repeats flanking the Alu sequence are shown as closed boxes. Restriction enzyme symbols are: E, EcoRI; H, HindIII; P, PstI; K, KpnI; T, TaqI; A, AvaI; S, SmaI,; SA, SalI.

followed (19 bp) by a full length Alu element which is flanked by direct repeats, and is in the opposite transcriptional orientation. The Alu sequence immediately follows the tandem duplication in the 7SK45 recombinant phage. The Alu DNA contains sequences homologous to the bipartite polymerase III promoter, which may be sufficient for transcription (22). The 7SK RNA coding region of the genomic clones differ by the addition of a G at position 236 from the 7SK cDNA previously described (10). However, HeLa cell 7SK RNA sequenced by dideoxy chain termination sequencing, is identical to the genomic clones in this region (data not shown). At the 3' end of the RNA coding sequence there is a stretch of T residues which is known to be an efficient transcription terminator for polymerase III (23,24).

Identification of the 7SK33 and 7SK45 sequences in human genomic DNA

Restriction maps of the two clones differ due to the 181 bp duplication of the 3' end of the coding region of 7SK45. A 695 bp PstI-SmaI fragment in 7SK33 is

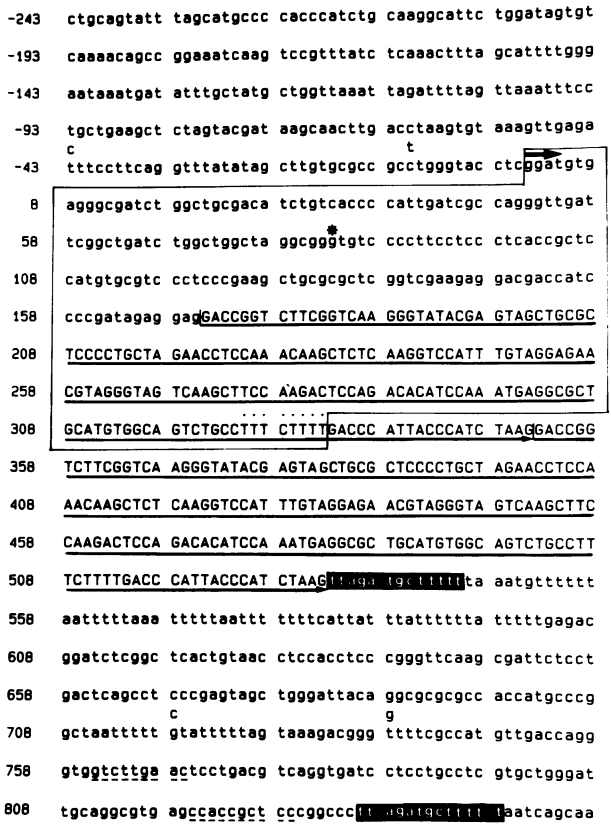


Figure 2
Sequence of the 7SK45 DNA. The bases above the sequence indicate where 7SK33 differs from 7SK45. The tandem duplication of 7SK45 is shown in upper case letters and underlined by a continuous line. The arrowheads delimit the sequence duplication, which spans from base 414 to 594 and from base 595 to 775. The presumptive 7SK RNA coding sequence is boxed and the large arrow indicates the start and direction of transcription. The negative numbers indicate the sequence before the start of the 7SK RNA homology. A possible polymerase III terminator is marked by dots above the sequence. An asterisk marks the G which is absent from the cDNA clone previously sequenced (10). The direct repeats flanking the Alu sequence are shown as white letters on a black background. The possible bi-partite polymerase III promoter of Alu DNA is underlined by a broken line.

replaced by a 876 bp fragment in 7SK45. DNA probes complementary to these fragments were hybridized to human DNA to find out whether both sequences are present within the human genome. Figure 3 shows a Southern blot of human DNA

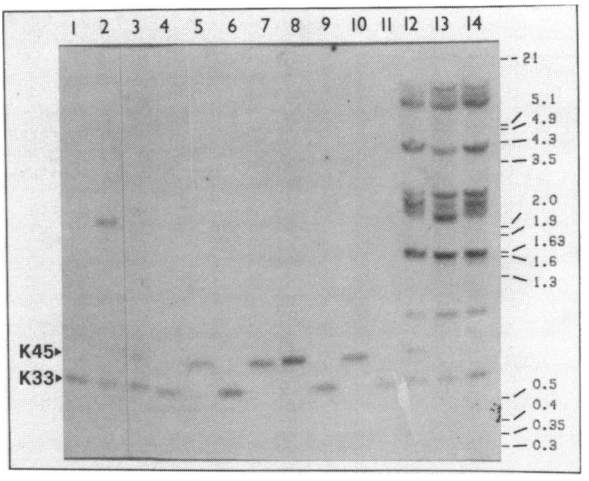
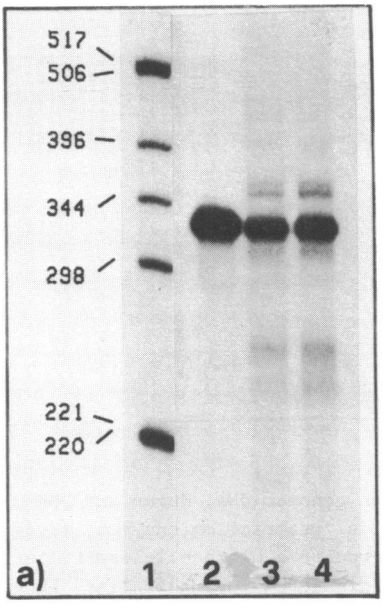
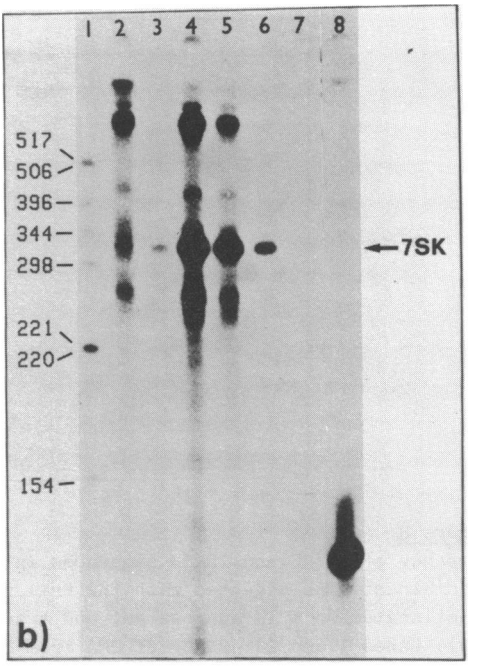
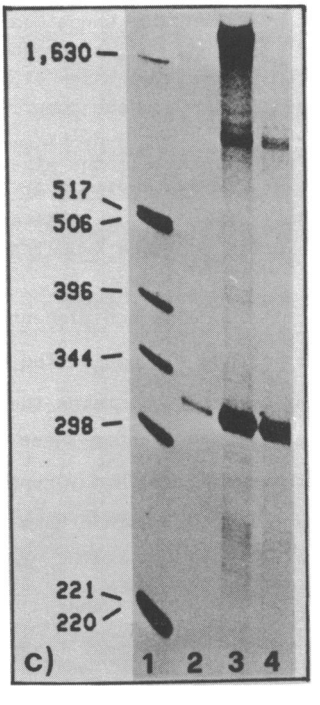


Figure 3
Presence of 7SK45 and 7SK33 sequences in human genomic DNA. Human and 7SK recombinant DNAs digested with the PstI and SmaI restriction enzymes, were fractionated on a 1% agarose gel and transferred to a nitrocellulose filter (29). Lanes 1 and 14: 20 ug of DNA from HL60 tissue culture cell line. Lanes 2 and 13: 20 ug of DNA from HeLa tissue culture cell line. Lanes 3 and 12: 20 ug human placenta DNA. Lanes 4 and 11: one gene equivalent of 7SK33 DNA. Lanes 5 and 10, one gene equivalent of 7SK45 DNA. Lanes 6 and 9: two gene equivalents of 7SK33 DNA. Lanes 7 and 8: two gene equivalents of 7SK45 DNA. Lanes 1 to 7 were hybridized to a DNA probe comprising bases -160 to -3 of Figure 2 (probe A). lanes 8 to 14 were hybridized to a probe comprising bases 270 to 329 of Figure 2, i.e. the 3' end of the 7SK RNA homology (probe B). The numbers at the right hand side of the figure indicate the size markers in Kb. The arrows on the left of the figure indicate the position of the hybridization bands corresponding to the PstI-SmaI fragments of 7SK45 and 7SK33 respectively. The nitrocellulose filter was exposed to Kodak X-OMAT XAR-5 film for 14 days.

from HL60 and HeLa tissue culture cells and placenta, digested with the PstI and SmaI restriction enzymes. Probe A contains the 5' flanking region of the 7SK RNA coding sequence, while probe B contains the 3' end of the RNA coding sequence. As we have previously shown, under these conditions probe B hybridizes to several DNA fragments, which may represent 7SK genes or pseudogenes (lanes 12, 13 and 14)(10). Both probes hybridize to a DNA fragment migrating in the same position of the gel as the PstI-SmaI fragment of the 7SK33 recombinant phage (lanes 1,2,3,12,13 and 14). A hybridization band corresponding to the 7SK45 specific PstI-SmaI fragment is detected only in


517
506
396
344
298
221
220
a)
1 2 3 4


1 2 3 4 5 6 7 8
517
506
396
344
298
221
220
154
7SK
b)


1,630
517
506
396
344
298
221
220
c)
1 2 3 4

placental DNA (lanes 3 and 12). A hybridization band of higher molecular weight (2.2 kb) than those found in either recombinant is present in the DNA of HeLa tissue culture cells (lanes 2 and 13).

In the same experiment we also investigated the copy number of these putative 7SK RNA genes. A comparison of the intensity of hybridization of the two probes to human placenta (lanes 3 and 12) and measured amounts of recombinant DNAs (lanes 4-11) shows that there are one to two copies of these sequences per haploid genome. In placenta DNA, hybridization to the 7SK45 specific PstI-SmaI fragment is less intense than the hybridization to the 7SK33 specific fragment. This may be explained by the fact that the DNA was extracted from several placentas and only the genome of a few individuals contains this type of fragment. Hybridization of the probe containing the 3' end of the 7SK RNA coding region is twice as strong to the recombinant with the duplication, than to the one without, while the probe containing the 5' flanking sequence hybridizes equally well to both. The larger fragment

Figure 4

Polyacrylamide-urea gel electrophoresis of RNA synthesized using 7SK33 and 7SK45 DNAs as templates for transcription.

a. Lane 1: ^{32}P-labelled pBR322 DNA digested with the HinfI restriction enzyme was used in this and subsequent experiments as a molecular weight marker. The numbers on the left of the figure indicate the sizes in base pairs of the DNA fragments. Lane 2: Human ^{32}P-labelled 7SK RNA purified from HeLa tissue culture cells. Lane 3: ^{32}P-labelled RNA synthesized using as a template a recombinant plasmid DNA containing the 7SK33 sequence indicated as 33C in Figure 6a. Lane 4: ^{32}P-labelled RNA synthesized using as template a recombinant plasmid DNA containing the 7SK45 sequence indicated as 45B in Figure 6a. Template DNA concentration was 5 ug/ml.

b. ^{32}P-labelled RNA synthesized using as template a recombinant plasmid DNA containing the 7SK33 sequence indicated as 33B in Figure 6a. The transcription mixture contains, lane 2: 25% cytoplasmic extract; lane 3: 25% nuclear extract; lanes 4, 5 and 6: 25% cytoplasmic plus 25% nuclear extracts (standard transcription mixture). The template DNA concentration of the transcription mixtures was, lanes 2, 3 and 6: 5 ug/ml; lane 4: 15 ug/ml; lane 5: 10 ug/ml. Lane 7: 5 ug/ml pEMBL18 DNA in a standard transcription mixture. Lane 8: 5 ug/ml XBS1 recombinant plasmid DNA in a transcription mixture containing 25% cytoplasmic extract. Lane 1: molecular weight marker.

c. ^{32}P-labelled RNA synthesized using the same 7SK33 recombinant DNA plasmid as in Figure 4b, in a standard transcription mixture containing variable concentrations of alpha-amanitin, lane 2: 100 ug/ml; lane 4: 1 ug/ml; lane 3: no alpha-amanitin. Lane 1: molecular weight marker. Template DNA concentration was 10 ug/ml.

detected in HeLa cell DNA, hybridizes twice as strongly to the 3' end DNA probe as to the 5' end DNA probe, suggesting that it may have a structure similar to that of the 7SK45 specific fragment.

Transcriptional analysis of the 7SK33 and 7SK45 DNA recombinants

The transcriptional potential of the 7SK DNA recombinants was tested using an "in vitro" transcription system prepared from HeLa cells. Figure 4a shows the results of the transcription of 7SK33 (lane 3) and 7SK45 (lane 4) recombinant plasmid DNAs in a transcription mixture containing nuclear and cytoplasmic cell extracts prepared by the method of Dignam et al. (16). Under the conditions used, the major RNA species synthesized is identical in size to purified HeLa cell 7SK RNA (lane 2). The experiments shown in Figure 4b were carried out to investigate the transcriptional requirements of the system. The absence of nuclear extract in the transcription mixture results in a drastic reduction of the synthesis of 7SK-sized RNA and an increased synthesis of larger and smaller RNAs (lanes 2 and 6). Under the same conditions a Xenopus borealis somatic 5S RNA gene is transcribed efficiently (lane 8). With nuclear extract alone, only a small amount of 7SK-sized transcript is synthesized (lane 3). The relative proportion of nuclear and cytoplasmic extracts in the transcription mixture is important since reduction of the amount of nuclear extract results in a decrease of 7SK RNA transcription (data not shown). Increasing the concentration of template DNA allows synthesis of a larger amount of 7SK-like RNA, as well as of other RNA species similar to those observed in lane 2 (lanes 4 and 5). Although at the concentration of 5 ug/ml of template DNA, the background transcription from the plasmid DNA is not detectable (lane 7), some of the additional RNA transcripts observed at higher template DNA concentration may derive from transcription of the vector DNA (Figure 6b, lane 11).

Figure 4c shows the results of addition of alpha-amanitin to the transcription mixture. The synthesis of the 7SK-like RNA is insensitive to 1 ug/ml alpha-amanitin (lane 4) but is inhibited by 100 ug/ml alpha-amanitin (lane 2), confirming that the transcription is polymerase III dependent (8). The higher molecular weight transcripts visible in lane 4 are also synthesized by polymerase III.

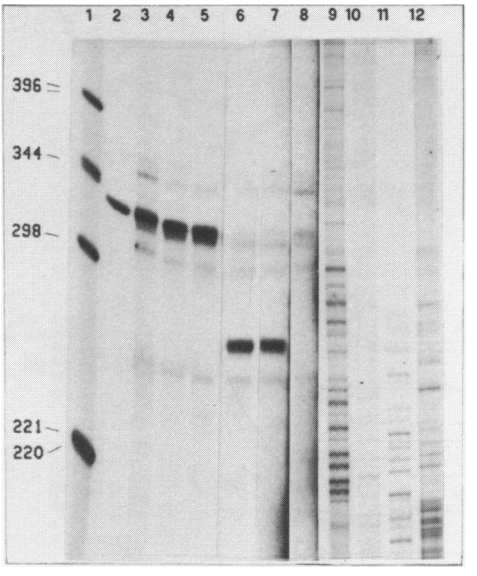


Figure 5

S1 mapping of RNA synthesized "in vitro". Molecular weight marker, lane 1; ^{32}P-labelled 7SK RNA of HeLa tissue culture cells, lane 2; ^{32}P-labelled RNA synthesized on the same 7SK33 template as in Figure 4a, lane 3; S1 nuclease digestion of the RNA after hybridization: to 7SK33 double stranded DNA, lanes 4 and 6; to 7SK33 sense strand DNA, lanes 5 and 7; to 7SK33 antisense strand DNA, lane 8. A set of sequencing reactions, lanes 9-12.

S1 mapping of 7SK RNA synthesized "in vitro"

S1 mapping analysis of the 330 nucleotide RNA synthesized "in vitro" demonstrates its identity with 7SK RNA synthesized "in vivo" (Figure 5). When a double stranded DNA fragment including 7bp upstream and 126 bp downstream from the coding region is hybridized to the 330 nucleotide RNA, the complete transcript is protected (lane 4). The same result is obtained by hybridizing a single stranded DNA probe complementary to 7SK RNA and containing 243 bases upstream and 126 bases downstream from the 7SK RNA coding region (lane 5). On the contrary, single stranded DNA of the same region but not complementary to the 7SK RNA does not protect the RNA from S1 nuclease digestion (lane 8). These results show that the 7SK-like RNA synthesized "in vitro" is transcribed only from one DNA strand corresponding to the sense strand of the 7SK RNA coding region.

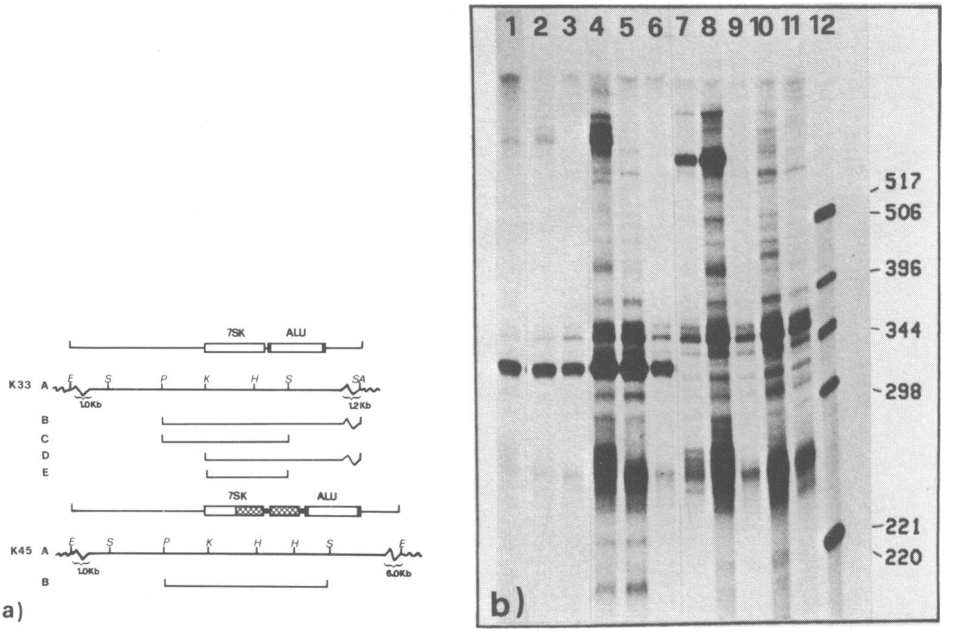


Figure 6

a. Restriction maps of recombinant DNAs obtained from plasmid clones of the 7SK33 and 7SK45 genes. 7SK33A and 7SK45A represent plasmids containing the 7SK33 and 7SK45 DNA fragments shown on an expanded scale in Figure 1. 7SK33B,C,D,E and F and 7SK45B indicate the same DNA fragment deleted according to the restriction map shown above. The relative positions of the 7SK RNA coding region and of the Alu DNA are indicated schematically above the restriction maps. Restriction enzyme symbols and the structural organization of the sequences are indicated as in Figure 1.

b. Transcription analysis of the recombinant plasmid DNAs shown in a. Templates for transcription and DNA concentrations were, lane 1: 7SK33A DNA, 10 ug/ml; lane 2: 7SK33B DNA, 10 ug/ml; lane 3: 7SK33C DNA, 10 ug/ml; lane 4: 7SK33B DNA, 20 ug/ml; lane 5, 7SK33C DNA, 20 ug/ml; lane 6: 7SK45B DNA, 10 ug/ml; lane 7: 7SK33D DNA, 10 ug/ml; lane 8: 7SK33D DNA, 20 ug/ml; lane 9: 7SK33E DNA, 10 ug/ml; lane 10: 7SK33E DNA, 20 ug/ml; lane 11, pEMBL18 DNA, 10 ug/ml. Lane 12: molecular weight marker.

Lane 6 shows the RNA fragment protected by a double stranded DNA probe containing 7 bp upstream from the 7SK coding region and 274 bp of the 7SK RNA coding region. If the transcript corresponds to correctly initiated 7SK RNA, the protected RNA should be 274 nucleotides long. Comparison of the migration of this protected RNA with a set of sequencing reactions (lanes 9-12) shows that the RNA fragment is 273-274 nucleotides long. In agreement with the

previous data, a RNA of the same size is protected by a single stranded DNA, complementary to 7SK RNA, which contains the same portion of the 7SK RNA coding region (lane 7).

Based on these results, we conclude that the 330 nculeotide RNA synthesized "in vitro" corresponds to 7SK RNA found "in vivo" and that 7SK33 and 7SK45 contain genes coding for 7SK RNA.

Role of the regions flanking the 7SK gene in transcription

The 7SK coding region does not contain internal promoters which obviously conform to the known models of polymerase III dependent RNA transcription (10). We have investigated, by deletion analysis, the possible role in transcription of the 3' and 5' sequences flanking the 7SK RNA coding region. At the 3' end of the gene a convenient SmaI restriction enzyme site was used to delete the majority of the Alu DNA, leaving approximately 120 bp after the termination sequence of the 7SK RNA gene. At the 5' end, resection using either the PstI or KpnI restriction enzymes leaves 243bp and 8bp of 5' flanking DNA respectively. The structure of these deleted DNAs is shown in Figure 6a and Figure 6b shows the polyacrylamide-urea gel electrophoresis of the products of transcription using the deleted DNAs. Resection of the DNA sequences at the 3' end of the gene does not affect the transcription of the 7SK RNA (lanes 3 and 5), while extensive deletion of the 5' flanking DNA, leaving only 8bp before the coding sequence abolishes the synthesis of 7SK RNA (lanes 7,8,9 and 10). The latter result is independent of the orientation of the deleted DNA within the plasmid, the particular batch of transcription extract used (data not shown), or the DNA concentration during the reaction (Fig. 6b).

A comparison of the products of transcription of the recombinant DNAs plus (lanes 2,4,7 and 8) and minus (lanes 3,5,9 and 10) the 3' end flanking sequences, shows that some of the transcripts, larger or smaller than 7SK RNA, are derived from these sequences. The experiment of Figure 6b also shows that the efficiency of transcription of the 7SK gene with and without the 3' end duplication, is essentially the same (lanes 3 and 6), indicating that neither the duplication nor the two base substitutions upstream from the coding region affect the rate of 7SK RNA transcription.

DISCUSSION

Structure of the 7SK RNA genes

We have previously described several 7SK RNA pseudogenes, characterized by the presence of frequent 3' end truncations, deletions, insertions and base substitutions. These pseudogenes are often flanked by direct repeats, suggesting transposition. Some of the pseudogenes are also in close association with Alu DNA and generally they are embedded in a repetitive environment (10). The two recombinant clones described here contain a sequence identical to the "in vivo" 7SK RNA, followed by a convincing polymerase III termination signal. The sequence immediately 5' to the 7SK RNA homology is not repetitive and is identical in both clones, and no direct repeats, signalling an insertion event, flank the putative 7SK RNA coding region. In summary, these are the only genomic sequences among those that we have isolated which have all the structural features of 7SK RNA genes. Since both putative genes direct the "in vitro" synthesis of a RNA 330 nucleotides long which corresponds to the 7SK RNA synthesized "in vivo", we conclude that we have isolated two genes coding for human 7SK RNA.

One similarity between the genes and the pseudogenes is the association of the 7SK coding region with Alu DNA. The presence of direct repeats flanking the Alu sequence in the 7SK33 and 7SK45 clones suggests that Alu DNA was inserted next to the 3' end of the 7SK DNA. The vicinity of the 7SK RNA coding sequence to a highly repeated mobile element may cause instability of this region and this may explain the presence of the duplication found in the 7SK45 DNA, which has generated a pseudogene between the end of the 7SK RNA coding region and the beginning of Alu DNA. The sequence of the 5' truncated pseudogene is identical to the corresponding region of the gene, suggesting that the duplication was a recent event. A similar arrangement is found in the Xenopus laevis oocyte 5S RNA gene cluster, where a 3' truncated pseudogene follows the gene in each repeating unit (25).

Transcription of the 7SK RNA genes

To obtain significant transcription of the 7SK RNA genes it is necessary to use a mixture of cytoplasmic and nuclear extracts.

The low level of transcription observed with cytoplasmic or nuclear extracts

alone suggests that neither of the two extracts contain a sufficient amount of the factors necessary for the efficient transcription of 7SK RNA genes. However, the cytoplasmic extract contains all the factors necessary for efficient transcription of a 5S RNA gene of Xenopus borealis and of a leucine transfer RNA gene of Coenorhabditis elegans (data not shown). Since several specific factors are known to interact with the different polymerase III promoters (26), it is likely that the additional factor(s), supplied by the nuclear extract, is (are) specific for the transcription of the 7SK RNA gene. We have not defined the nature of the transcripts which derive from the 3' flanking region of the 7SK gene. However, our results show that these transcripts are polymerase III dependent and are reduced in size after removal of the sequences flanking the 5' end of the 7SK gene (Figure 6b, lanes 7 and 8). Since the Alu sequence at the 3' end of the 7SK gene contains a promoter for polymerase III, we think that Alu DNA may be responsible for the synthesis of these RNA species. The size of the RNAs is consistent with transcripts initiated at the 5' end of the Alu DNA and terminated at stretches of Ts present in the antisense strand of the DNA flanking the 7SK gene at the 5' end. In the absence of the 5' end flanking region these transcripts may terminate in the adjacent plasmid sequences, generating a smaller transcript. Although this transcription would proceed in the opposite direction to the transcription of the 7SK RNA gene, we have found no evidence of interference with the synthesis of 7SK RNA, probably due to the more efficient transcription of the 7SK gene compared to the transcription of the Alu sequence.

Promoter regions of the 7SK gene

The genes transcribed by polymerase III contain an internal promoter consisting of two conserved sequences either separate or adjacent to one another. One of these sequences is highly conserved and common to all genes (box A), while the second sequence is variable. Two different consensus sequences have been described for this element: one found in transfer RNA genes (box B) and the other found in 5S RNA genes (box C) (22,27).
The internal promoters are usually sufficient but upstream sequences can also be required for efficient transcription. Although the 7SK gene contains a

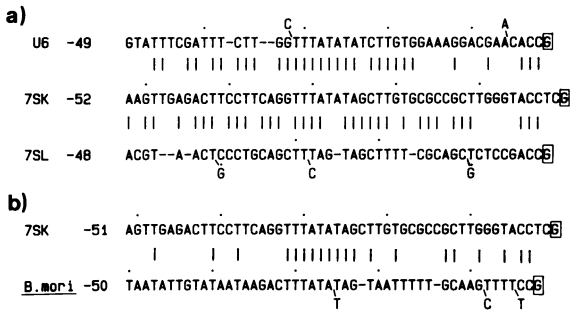


Figure 7
Comparison of the 5' flanking region of the 7SK RNA gene with a) U6 and 7SL RNA gene sequences, and b) Bombyx mori alanine tRNA gene sequences. The human U6 RNA gene (1), the 30.1 human 7SL RNA gene (2) and the Bombyx mori alanine tRNA gene (3) are compared to the 7SK33 gene. Homology is indicated by a line and the first base of each transcript is boxed. Every 10th base from the start site is indicated by a dot above the base. Gaps have been introduced and deletions made to maximize the homology. Gaps are indicated by a dash and the deleted bases are shown above or below the line of the sequence.

possible box A, located approximately 70bp downstream from the start of transcription, it does not have an obvious box B or box C (10). In addition we find that upstream sequences are necessary for "in vitro" transcription of the gene. A comparison with other genes which require external sequences for transcription (Bombyx mori alanine tRNA gene) (3) or modulation of transcription (human 7SL gene (2)), reveals interesting structural homologies (Figure 7). Furthermore, homologies are found to the upstream sequences of the human U6 RNA gene which is polymerase III dependent (1). This gene also contains a box A and no box B or box C but the role of upstream sequences in transcription is not yet known. Figure 7 shows the homology between the upstream region of the 7SK RNA gene and those of the human genes for U6 and 7SL, and with the upstream region of the Bombyx mori alanine tRNA gene. The homology between the 7SK and the two human genes is more extensive than that observed between the 7SK and the Bombyx mori gene. The region common to the four genes is strikingly similar to the polymerase II TATA box in sequence and position (4). This motif is conserved in several polymerase III dependent genes of Bombyx mori (3), and in the 5S RNA genes of several eucaryotes (28). The existence of the TATA box in possible promoter regions of genes

transcribed by polymerase III can be viewed as an evolutionary link between bacterial promoters and eucaryotic polymerase II and polymerase III promoters. The existence of common sites may imply coordinate regulation of the 7SK, 7SL and U6 RNA genes.

ACKNOWLEDGEMENTS

We would like to thank Cristina Grimaudo for help with the restriction mapping, Giorgio Corsi for preparation of the figures and Susan Taddeo for typing the manuscript. We are grateful to T. Pedersen for data prior to publication.

*To whom correspondence should be addressed

REFERENCES

1. Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) Proc. Natl. Acad. Sci. USA (in press).
2. Ullu, E. and Weiner, A.M. (1985) Nature 318, 371-374.
3. Larson, D., Bradford-Wilcox, J., Young, L.S. and Sprague, K.V. (1983) Proc. Natl. Acad. SCi. USA 80, 3416-3420.
4. Corden, J., Waslylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, D. and Chambon, P. (1980) Science 209, 1406-1414.
5. Kramer, A., Keller, W., Appel, B. and Luhrmann, R. (1984) Cell 38, 299-307.
6. Strub, K., Galli, G., Busslinger, M. and Birnstiel, M.L. (1984) EMBO J. 3, 2801-2807.
7. Walter, P. and Blobel, G. (1982) Nature 299, 691-698.
8. Zieve, G., Benecke, B. and Penman, S. (1977) Biochemistry 16, 4520-4525.
9. Ullu, E., Esposito, V. and Melli, M. (1982) J. Mol. Biol. 161, 195-201.
10. Murphy, S., Altruda, F., Ullu, E., Tripodi, M., Silengo, L. and Melli, M. (1984) J. Mol. Biol. 177, 575-590.
11. Lawn, R.M., Fritch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) Cell 15, 1157-1174.
12. Bensi, G., Raugei, G., Klenfenz, H. and Cortese, R. (1985) EMBL J. 4, 119-126.
13. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
14. Dente, L., Sollazzo, M.,Baldari, C., Cesareni, G. and Cortese, R. (1985) in DNA Cloning Vol. 1, Glover, D.M. Ed., IRL Press.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucl. Acids Res. 11, 1475-1489.
17. Peterson, R.C., Doering, J.L. and Brown, D.D. (1980) Cell 20, 131-147.
18. Tranquilla, T.A., Cortese, R., Melton, D. and Smith, J.D. (1982) Nucl. Acids Res. 10, 7919-7934.

19. Maniatis, T., Fritsch, E.F; and Sambrook, J. (1982) in Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Press, N.Y.
20. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res. 11, 1645)1655.
21. England, T.E. and Uhlenbeck, C.C. (1978) Nature 275, 560-561.
22. Ciliberto, G., Cstagnoli, L. and Cortese, R. (1983) in Current Topics in Developmental Biology, Vol. 18, pp. 59-88. Academic Press, N.Y.
23. Bogenhagen, D. and Brown, D. (1981) Cell 24, 261-270.
24. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Nature 294, 626-631.
25. Miller, J.R., Cartwright, E.M., Brownlee, G.G., Federoff, N.V. and Brown, D.D. (1978) Cell 13, 717-725.
26. Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983) Science 222, 740-748.
27. Ciliberto, G., Raugei, G., Costanzo, F., Dente, L. and Cortese, R. (1983) Cell 32, 725-733.
28. Rubacha, A., Sumner III, W., Richter, L. and Beckingham, K. (1984) Nucl. Acids Res. 12, 8193-8207.
29. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.