Biochem. J. (1987) 245, 281–284 (Printed in Great Britain)

Observations on the structure of two human 7SK pseudogenes and on homologous transcripts in vertebrate species

Peter HUMPHRIES,*§ S. E. Hilary RUSSELL,† Peter McWILLIAM,* Shirley McQUAID,* Colin PEARSON‡ and Marian M. HUMPHRIES*

*Department of Genetics, Lincoln Place Gate, Trinity College, Dublin 2, Ireland, †Department of Oncology, The Queen's University of Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast, Northern Ireland, U.K., and ‡Central Toxicology Laboratory, Imperial Chemical Industries PLC, Alderley Park, Macclesfield, Cheshire SK10 4TJ, U.K.

A comparison of the sequence of two human 7SK RNA pseudogenes, covering approx. 190 and 240 base-pairs of the structural gene, is presented. Both repeated elements are flanked by direct repeats and begin at the 5′ end of the gene. Each terminates approx. 90 base-pairs short of the 3′ end, the latter representing a contiguous sequence and the former carrying an internal deletion of about 40 base-pairs, this region being flanked in the progenitor gene by short repeated sequences. Southern blotting using a human 7SK pseudogene probe illuminated a series of multiple restriction fragments in mammalian genomes, with generally fewer fragments in the genomes of birds and reptiles and a single reactive fragment in DNA from terrapin (*Pseudemys scripta elegans*) and *Xenopus laevis* (South African clawed toad). In the latter case this fragment was only detectable on long exposure under the hybridization stringencies employed. 7SK transcripts were readily detectable in all mammalian, avian, reptilian and amphibian species analysed, although the gene appeared to be expressed at rather low levels in the ovaries of *Xenopus laevis*, possibly accounting for its failure to have become dispersed via 'retroposition' in this species.

INTRODUCTION

Mammalian 4–8S RNAs constitute a variety of species, including the U series (U1–U7), so-called because of their relatively high content of uridylic acid and the 4, 4.5, 5, 5.4, 5.8, 7 and 8S species classified according to their M_r values (Reddy *et al.*, 1981; Gunning *et al.*, 1981). Many of these species are homologous with dispersed repeated elements in genomic DNA, most of which appear to bear properties characteristic of sequences dispersed via 'retroposition' (for a review, see Rogers, 1985). 7S RNA consists of two distinct and abundant species, 7SL and 7SK, and two minor species, 7S-1 and 7S-2 (Reddy *et al.*, 1981). Both the 7SL and 7SK structural genes have recently been cloned (Ullu & Melli, 1982; Ullu *et al.*, 1982; Murphy *et al.*, 1984). The SL gene shares homology with the *Alu* family of dispersed repetitive elements, and the RNA forms part of the signal recognition particle (Walter & Blobel, 1982). The 7SK gene also shares homology with a dispersed pseudogene family and may have a function related to that of the 7SL gene (Murphy *et al.*, 1984). Whereas the *Alu* family of interspersed repeated elements is present at very high copy number, constituting the major 'retroposon' family of the human genome, sequences homologous with 7SK RNA are present at much lower copy number, are much more highly conserved among vertebrate species and are resolvable as a series of discrete bands in Southern transfers of restriction digests of total human DNA (Murphy *et al.*, 1984). The sequences of four dispersed 7S pseudogenes have recently been reported (Murphy *et al.*, 1984). One clone contains the entire sequence of the 7S gene, and three additional clones are truncated, having 30–90 base-pairs missing from the 3′ end. We have also previously determined the sequence of a repeated element of a family arbitrarily termed 'A3' after the recombinant plasmid from which it was derived, which now appears to bear homology with 7SK RNA (Humphries *et al.*, 1983, 1985). The element detects an abundant conserved small transcript in humans and rodents which can (or a portion thereof) be retained on oligo(dT)–cellulose, suggesting that at least some of the molecules are polyadenylated. We have used this element as a probe to isolate and sequence an additional member of the 7SK family of pseudogenes and have compared these sequences with that of the progenitor gene. Both elements are truncated at the 3′ end, and one carries a deletion of approx. 40 base-pairs, the homologous region in the structural gene being terminated by inverted and direct repeats. We have also used the probe to further investigate 7SK transcripts in a wide variety of vertebrate species. Interestingly, transcript levels were relatively low in RNA purified from several *Xenopus laevis* (South African clawed toad) ovary specimens, which could have implications with regard to the dispersal of pseudogenes in this species.

MATERIALS AND METHODS

Recombinants containing 7SK pseudogenes

The recombinant pA3.1, containing a single 7SK sequence, has previously been sequenced (Humphries *et al.*, 1985). A second recombinant bearing homology with this probe was isolated from a plasmid library of human genomic segments and sequenced according to the di-deoxy method of Sanger *et al.* (1977).

§ To whom correspondence and reprint requests should be addressed.

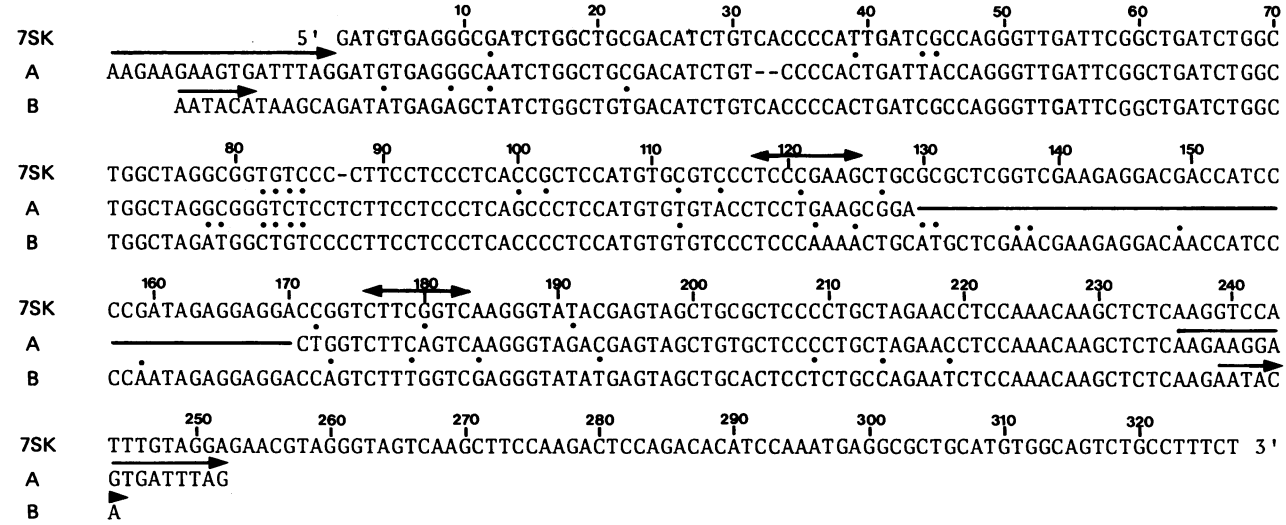

Fig. 1. Sequence of two 7SK pseudogenes (A and B) in alignment with the structural gene 7SK

Point substitutions are indicated by dots, and the internal deletion in pseudogene A by a solid line. Terminal direct repeats and the inverted/direct repeat flanking the deletion are indicated by arrows.

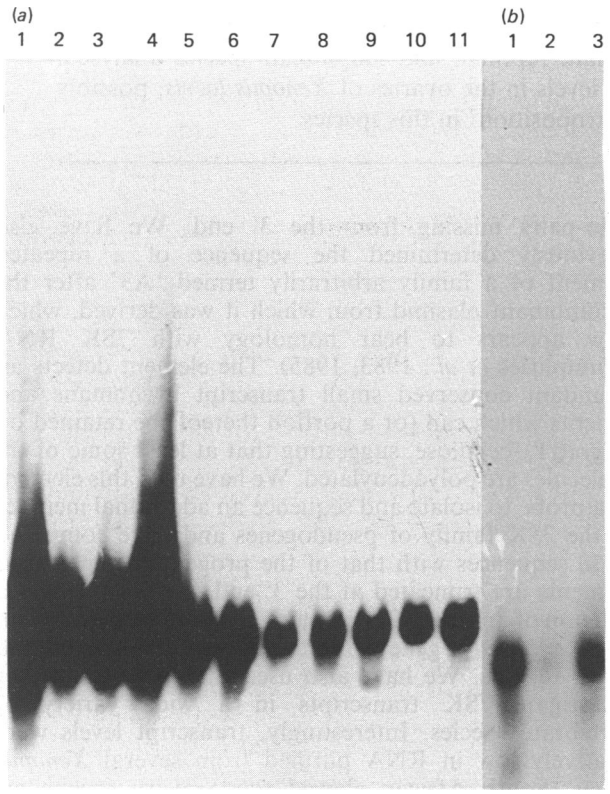

Fig. 2. Analysis of transcripts homologous with 7SK pseudogene A

Samples (10 μg) of total RNA, prepared as described in the text, were electrophoresed on 1.5%-agarose/Mops/formaldehyde gels and blotted on to nylon membrane. Blots were hybridized with nick-translated element-A probe (Fig. 1) and exposed to X-ray film at −70 °C for 24 h. Gel (*a*) lanes 1–11 show autoradiographic signals from Bob White and Japanese quails, Cape Barren goose (*Cereopsis novae-hollaniae*), pigeon, snake, terrapin, toad (*Xenopus laevis*) liver, frog, skink, salamander and green lizard respectively. Gel (*b*), lanes 1–3, are signals from *Xenopus laevis* liver, ovary and (again) liver respectively. RNAs extracted from all other animals produced similar signals. In all cases, signals obtained with RNA from heart and liver were the same.

Nucleic acids

Hearts and livers were dissected from Japanese and Bob White quail, pigeon (*Columba livia*), terrapin (*Pseudemys scripta elegans*), dice snake (*Natrix piscator*), green lizard (*Lacerta viridis*), *Xenopus laevis*, common frog (*Rana temporaria*), fire-bellied newt, spotted salamander (*Salamandra maculosa*), rat, mouse and guinea pig. Tissues were pulverized individually in liquid N_2 and the DNA extracted by treatment with proteinase K and SDS, followed by phenolic extraction and ethanol precipitation. All DNAs were of high M_r, as judged by analysis of undigested samples in 1%-agarose gels. DNAs were also isolated directly from peripheral-blood samples taken from geese and chickens and from white-cell buffy coats isolated by low-speed centrifugation of human and orang-outang (*Pongo pygmaeus*) peripheral blood.

In parallel, total RNA was purified by the guanidinium isothiocyanate/hot-phenol procedure (Maniatis *et al.*, 1982) from similar tissues. Total RNA from *Xenopus* ovaries was very kindly provided by Dr. Bob Old, Department of Biological Sciences, University of Warwick, Coventry, U.K.

Conditions employed for Southern and Northern blotting have been previously described (Humphries *et al.*, 1985). In the case of Northern blots, equality of transfer of RNA was confirmed by hybridization of the filters to an 18S-ribosomal-DNA probe (a plasmid containing a 1.9-kilobase insert from the mouse 18S-ribosomal-RNA gene, kindly supplied by Dr. N. Arnheim, State University, New York, NY, U.S.A.).

RESULTS AND DISCUSSION

The sequences of two 7SK pseudogenes, in alignment with the sequence of 7SK cDNA (Murphy *et al.*, 1984) are depicted in Fig. 1. Element A has 17 single base-pair substitutions, a deletion of two base-pairs and a point addition, as well as a deletion of 41 base-pairs spanning nucleotides 130–171 of the 7SK gene. Element B has 26

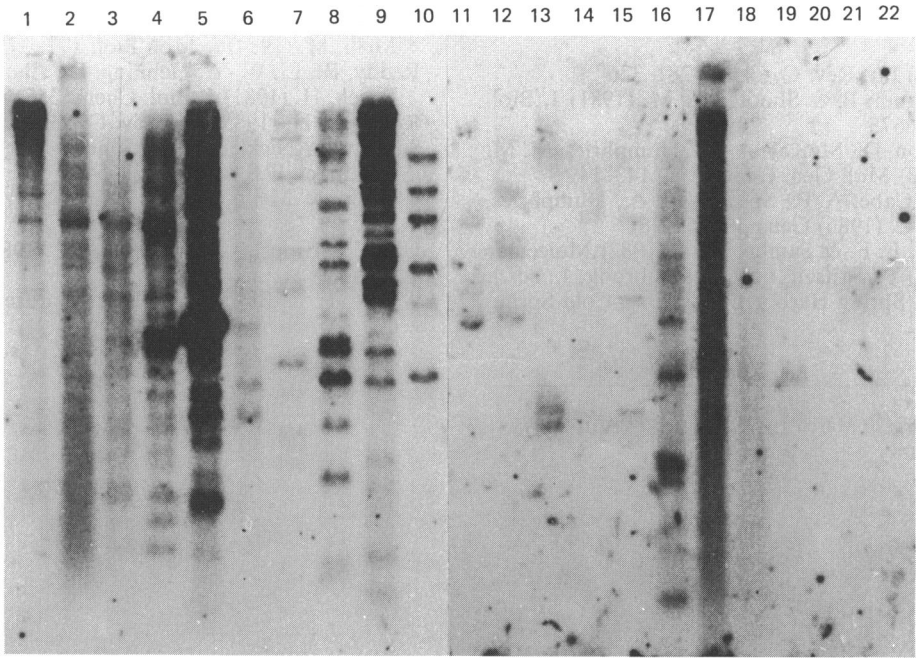

Fig. 3. Hybridization of 7SK pseudogene A (Fig. 1) to digested vertebrate DNAs

DNA purified from human (Caucasian, Chinese, African), orang-outang, rabbit, guinea-pig, rat, mouse, dog, goat, goose, Japanese quail, Bob White quail, pigeon, chick, dice snake, green lizard, skink, terrapin, salamander, newt and toad (*Xenopus laevis*) was digested with *Eco*RI restriction nuclease, separated by electrophoresis in a 1% -agarose gel (lanes 1–22 respectively), blot-transferred on to nitrocellulose and hybridized with a ^{32}P-labelled segment of human DNA containing 7SK pseudogene A (Fig. 1). The filter was washed to a stringency of 0.2 × SSC (SSC is 0.15 M-NaCl/0.015 M-sodium citrate) at 65 °C and exposed to X-ray film for 24 h. In this experiment, digestion of the human DNA appeared incomplete, possibly due to inhibition of the restriction nuclease by traces of heparin in this particular sample.

substitutions and one point addition. Both elements show approx. 6% divergence from the sequence of 7SK RNA, estimates from other pseudogenes ranging from 2 to 11% (Murphy *et al.*, 1984). The sequence of both pseudogenes begins at the 5′ end of the 7SK gene and both elements terminate at nucleotide 238, about 90 base-pairs short of the 3′ terminus. The elements are flanked by 17 and 6 base-pair direct repeats, indicative of the amplification of target sequences results from integration into genomic DNA.

The 41-base-pair deletion in element A is immediately flanked in the progenitor gene by the sequence TCCCGAAG·CTTCGGTC, representing a hexanucleotide inverted repeat flanked by direct repeats (TC) of 2 base-pairs. Such structures may have facilitated in the excision or integration of this segment (for a review, see Finegan, 1985). The analysis of many more 7S pseudogenes would be required, however, in order to determine the prevalence of deletions within this region.

Previous Northern-blotting data (Ullu *et al.*, 1982) have indicated that a 7SK probe detects a single abundant homologous 7S transcript in RNA from HeLa cells, mouse L-cells and chick embryos, and a transcript of relatively weaker intensity and of noticeably higher mobility in RNA from *Xenopus laevis* blood cells. We have examined RNA from a wide variety of vertebrates, and have found an abundant 7SK transcript in all species (Fig. 2). However, whereas the transcript is plentiful in RNA from *Xenopus laevis* heart and liver and of the same size as that detected in other species, it was noticeably lower in two RNA preparations from *Xenopus* ovaries (Fig. 2). Levels of transcript may be affected by unknown parameters, and we have no evidence at present to indicate the generality, or otherwise, of this phenomenon. However, decreased transcription in ovaries would have interesting theoretical implications with regard to the dispersal of 7SK pseudogenes, since, without active transcription in germ cells, the gene would be unable to disperse via 'retroposition', even in spite of its being abundantly expressed in other tissues. Examination of DNA sequences homologous with 7SK genes indicates a dispersed pattern of multiple bands in mammals, with proportionally fewer bands in birds and reptiles (Fig. 3), although DNA from the green lizard produced a broad autoradiographic smear superimposed upon a pattern of discrete bands. Such smearing may indicate insufficient DNA purity. DNA from terrapin produced a single reactive band, and at this exposure no signals were detectable from salamander, newt or *Xenopus laevis*. In the case of the latter three species, two independent DNA preparations were analysed. On long exposure, however, a weak single band of hybridization was detected in DNA from *Xenopus*, suggesting a single genomic locus for the 7SK gene. Thus the transcriptional properties and genomic organization of the 7SK gene in *Xenopus* are entirely compatible with a 'retroposon' model for its dispersal in other vertebrate genomes.

This work was supported by the Medical Research Council of Ireland and in the United Kingdom by Imperial Chemical Industries PLC. We thank Dr. Cormac O'Carroll for the provision of animals and Dr. Bob Old for generous gifts of *Xenopus* RNA.

REFERENCES

Finegan, D. J. (1985) Int. Rev. Cytol. **93**, 281–326
Gunning, P. W., Beguin, P. & Shooter, E. M. (1981) J. Biol. Chem. **256**, 6670–6675
Humphries, P., Barton, D., McKay, A. M., Humphries, M. M. & Carritt, B. (1983) Mol. Gen. Genet. **190**, 143–149
Humphries, P., MacCabe, A. P., Spencer, R. A., Humphries, M. M.& Pearson, C. (1985) Gene **39**, 255–261
Maniatis, T. Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning (Maniatis, T., Fritsch, E. F. & Sambrook, J., eds.), pp. 194–195, Cold Spring Harbor Laboratory, Cold Spring Harbour, NY
Murphy, S., Altruda, F., Ullu, E., Tripodi, M., Silengo, L. & Melli, M. (1984) J. Mol. Biol. **177**, 575–590
Reddy, R., Li, W.-Y., Henning, D., Choi, Y. C., Nohga, K. & Busch, H. (1981) J. Biol. Chem. **256**, 8452–8457
Rogers, J. H. (1985) Int. Rev. Cytol. **93**, 187–279
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Ullu, E. & Melli, M. (1982) Nucleic Acids Res. **10**, 2209–2223
Ullu, E., Esposito, V. & Melli, M. (1982) J. Mol. Biol. **161**, 195–201
Walter, P. & Blobel, G. (1982) Nature (London) **299**, 691–698

Received 6 January 1987/20 March 1987; accepted 14 April 1987