Biochem. J. (1985) **225**, 657–663
Printed in Great Britain

A mouse gene family associated with a major submaxillary-gland glycoprotein

A. G. CRAIG, J. J. MULLINS, P. McTURK and W. J. BRAMMAR
Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, U.K.

(*Received 30 July 1984/Accepted 16 October 1984*)

A gene family encoding msp36 (mouse submaxillary protein 36), a major product of the mouse submaxillary gland, was shown by Southern analysis and genomic cloning to contain approx. 10 related genes. Heteroduplex mapping has elucidated the structure of one of the genes. Close physical linkage of four of the genes in this family has been demonstrated by the use of specific oligonucleotide probes.

The submaxillary gland of the mouse produces several physiologically important products, including NGF, EGF, several peptidases, amylase, renin and mucin (Barka, 1980). Clones containing copies of the major submaxillary-gland mRNA species have been described (Windass *et al.*, 1984), one of which was shown to represent a major tissue-specific mRNA that coded for a 16.5 kDa polypeptide. The relatively high abundance of sequences homologous with the recombinant plasmid pSMG751 in a cDNA library for the murine submaxillary gland (17%) (Windass *et al.*, 1984) seemed to indicate a major product of the gland, but the determination of the coding sequence has not revealed the identity of the gene product. For future convenience we have named this protein 'msp36' on the basis of its apparent M_r determined by SDS/polyacrylamide-gel electrophoresis.

In the present paper we show the existence of a tissue-specific gene family related to msp36 and the analysis of the structure of one of these genes. We also demonstrate the close physical linkage of two of the genes with the use of oligonucleotide probes.

Materials and methods

Animals

DBA/2 and C57BL/10 mice were supplied by Bantin and Kingman, Hull, U.K.

Enzymes and reagents

Restriction endonucleases were purchased from Bethesda Research Laboratories (U.K.). All reactions were carried out according to the suppliers' recommendations. Radioactively labelled reagents were supplied by Amersham International. Formamide was purchased from Fluka.

Abbreviations used: msp36, mouse submaxillary protein 36; NGF, nerve growth factor; EGF, epidermal growth factor; cDNA, complementary DNA; 0.5 × SSC, 75 mM-NaCl/7.5 mM-trisodium citrate; DMSO, dimethyl sulphoxide; kb, kilobase.

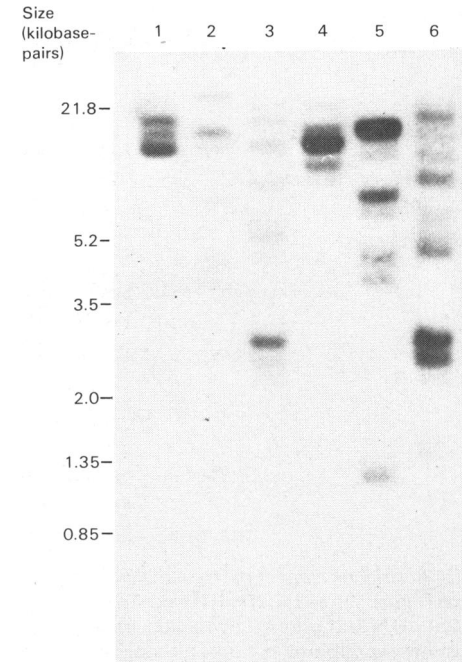
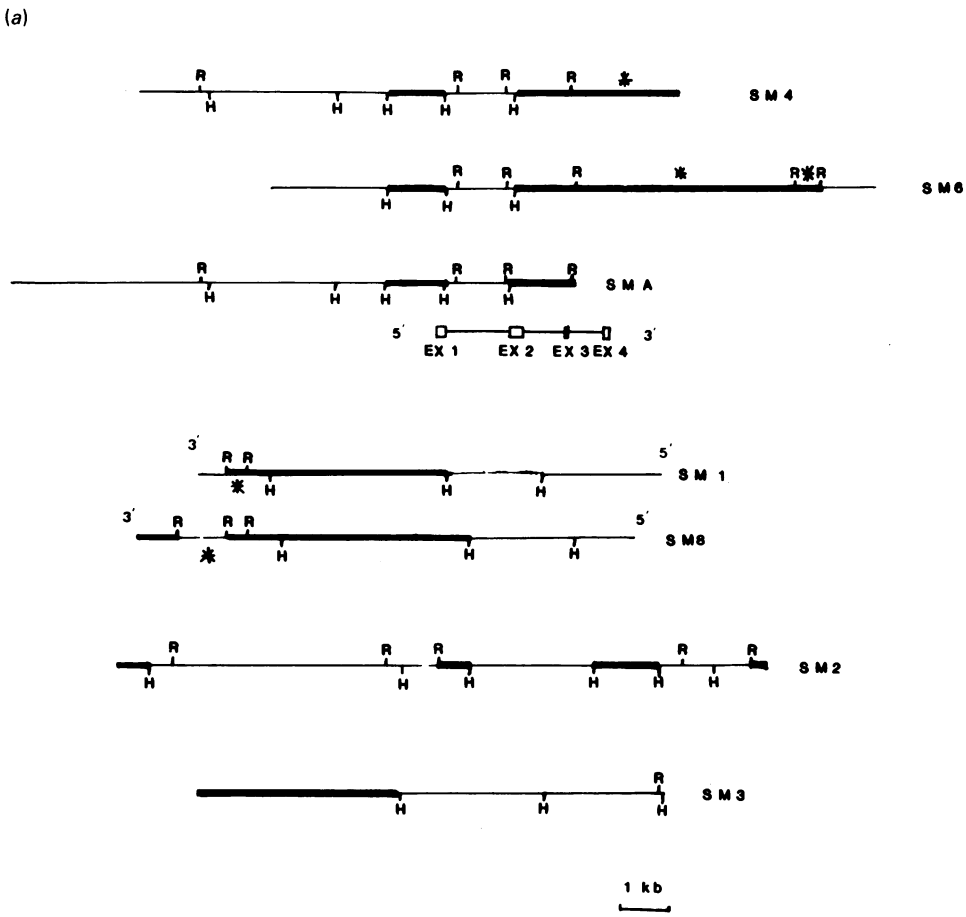


Fig. 1. *Southern analysis of murine genomic DNA with pSMG751 cDNA*
Tracks 1–3 contain DNA from C57BL/10 mice and tracks 4–6 from DBA/2 mice, each track containing 10 μg of DNA. The DNA species were digested with three enzymes: tracks 1 and 4, *Bam*HI; tracks 2 and 5, *Hin*dIII; tracks 3 and 6, *Eco*RI.



(b)

```
                    170         180         190         200
                     .           .           .           .
SMG751:   ACTGCTGATT CTGCTGGTGA AAATACTGGC TCTGAAACTC
          ********** ********** ********** **********
SM6/0.5:  ACTGCTGATT CTGCTGGTGA AAATACTGGC TCTGAAACTC

                    210         220
SMG751:   AGGCAGATAG TACTGATCAA
          *****   ** **********
SM6/0.5:  AGGCA---AG TACTGATCAA
```

Fig. 2. *(a) Restriction maps of the recombinant phage, and (b) partial sequence of the 0.5 kb EcoRI fragment of SM6 (SM6/0.5)* (*a*) This Figure shows the *Hin*dIII and *Eco*RI sites in the recombinant phage inserts. Regions which hybridize to the SMG751 cDNA are shown by thicker lines. Some of the inserts have been orientated (see Fig. 3*c*). SM4, SM6 and SMA show overlapping regions, although a *Hin*dIII site is missing in SM6. The structure of the gene thought to be represented by these phages is shown, derived from heteroduplex analysis (Fig. 5). The data for SM3, SM2 and SM1 are not shown (for the other phage, see Fig. 3). For SM3 the cDNA hybridizes to a 4.5 kb *Hin*dIII fragment at the left end of the insert, which includes 490 bases of the left arm of the vector. This fragment is not homologous with the 3′-fragment probe. In a *Hin*dIII/*Eco*RI double digest of SM2 DNA, the cDNA hybridizes to four bands of 2.2 kb (including 2 kb of the vector), 0.76 kb, 1.3 kb and 0.85 kb (containing 0.5 kb of vector). As found for SM3, the 3′-probe does not hybridize to the insert of SM2. The region of relatedness to the cDNA in SM1 is defined by a 3.6 kb *Hin*dIII fragment and a 0.42 kb *Eco*RI fragment. The 3′-probe gives a positive result with a 2.0 kb *Hin*dIII fragment, which is shown to the left of the 3.6 kb *Hin*dIII and contains 0.5 kb of the left arm of the vector. (Fragments that

General techniques

The methods for manipulation of phages, preparation, restriction and ligation of DNA and gel electrophoresis of DNA fragments were described previously (Burt & Brammar, 1982).

Southern blotting of mouse DNA

DNA was prepared as described by Barrie *et al.* (1981) and probed by the method of Jeffreys *et al.* (1980), the final wash being carried out in 0.5 × SSC. Nick-translation was performed by the method of Jeffreys *et al.* (1980), with the cDNA fragment from pSMG751 as the template.

Isolation of genomic clones

DNA was prepared as described by Barrie *et al.* (1981) and the genomic library constructed as described by Mullins *et al.* (1982) by using the vector λL47.1.

Size-selected (10–15kb) fragments of DBA/2 DNA generated by partial digestion with endonuclease *Sau*3A were ligated into the replacement vector λL47.1 (Loenen & Brammar, 1980) and recombinants were recovered by packaging *in vitro* (Jeffreys *et al.*, 1981). The phage library was screened by a modification of the procedure of Benton & Davis (1977), by using pSMG751 as a probe, with the λL47.1 vector and a λL47.1 recombinant containing bacterial DNA as negative controls.

After the purification cycles ten isolates were recovered and DNA was prepared from each.

Southern blotting of genomic clones

DNA from the λ recombinants was digested with a range of restriction endonucleases and treated as described above, the final wash being carried out in 0.3 × SSC.

When oligomers were used as probes, 100ng of the oligonucleotide was end-labelled (Maniatis *et al.*, 1982) and the hybridization performed overnight in 3 × SSC/0.1% SDS/alkaline-denatured salmon sperm DNA (50μg/ml) at 30°C (these conditions were also used for the pre-hybridization treatment of the nitrocellulose filter for 4–5h).

After hybridization the filter was washed for 15min at room temperature with two changes of 3 × SSC/0.1% SDS followed by washing in the same solution at 30°C for 30min. Then the filter was dried and autoradiographed.

Northern blotting

RNA was prepared from the tissues of 11-week-old male DBA/2 mice by the method of Noyes *et al.* (1979) and separated by electrophoresis after denaturation with glyoxal and DMSO (Maniatis *et al.*, 1982).

The RNA was transferred on to nitrocellulose and hybridization (with nick-translated pSMG751 cDNA) was performed as described by Thomas (1980).

Electron microscopy

Heteroduplexes (SM6 and pSMG751/*Bam*HI) were formed as described by Durham *et al.* (1980). Grids were processed (Davis *et al.*, 1971) and hybrids observed in a Jeol 100CX microscope. The DNA species of pBR322 and M13mp8 were used as size standards for measurements from prints using a Kontron digitizing tablet and computer.

Oligonucleotide synthesis

Oligonucleotides (16-mers) were synthesized using the filter paper bound technique of Matthes *et al.* (1984) and purified on 20% polyacrylamide/urea gels.

Results

Hybridization of pSMG751 to mouse genomic DNA

Genomic DNA prepared from DBA/2 and C57BL/10 mice was digested to completion with restriction endonucleases *Eco*RI, *Hin*dIII and *Bam*HI. The digests were electrophoresed on a 0.8% agarose gel and treated as described by Jeffreys *et al.* (1980), nick-translated pSMG751 cDNA being used as a probe (Fig. 1). Multiple bands can be seen in each track, indicating a complex genetic structure such as a multi-gene family, homologous with pSMG751.

The *Bam*HI tracks clearly show the presence of a restriction site polymorphism between the inbred lines DBA/2 and C57BL/10. A high-M_r band of approx. 20kb in track 1 is not seen in track 4, in which two lower-M_r bands of 11.8kb and 8.8kb appear. This genomic difference is not reflected in the level of the homologous mRNA in the submaxillary glands (A. G. Craig, unpublished work).

hybridize to the 3'-fragment probe are marked with an asterisk.) (*b*) The sequence is compared with that of pSMG751 cDNA and indicates the position of the fragment sequence starting towards the 5' end of the cDNA (160–170) and extending in a 3' direction. When related to the msp36 gene structure, derived from the heteroduplex data, this would indicate that the *Eco*RI fragment contains sequences from both exon 1 ('EX1') and exon 2 ('EX2'). The symbol '—' indicates ambiguous reading of the sequencing gel.

Analysis of a DBA-2 genomic library

The recombinant phage DNA species were mapped with several restriction enzymes including *Eco*RI and *Hin*dIII (Fig. 2). The restriction maps fall into four to five distinct groups, giving a lower limit for the number of genes present in the family.

Analysis of the recombinant phage DNA by Southern blotting with both full-length cDNA and a 3′ fragment (containing 111 bases of 3′ coding sequence and all of the 3′ non-coding region) as probes (Figs. 3*a*, 3*b* and 3*c*) was carried out. This indicated, in combination with the heteroduplex data (see below), that two genes in this family are closely linked and in some cases enabled the genomic inserts to be orientated. The orientation of the inserts of λSM4 and λSM6 was shown by hybridization of the 3′-probe to 3.3 kb (containing 2.3 kb of the right arm of the vector) and 4.4 kb *Eco*RI fragments respectively (marked with an asterisk on Fig. 2*a*). The 0.5 kb *Eco*RI fragment within λSM6 also hybridizes to the 3′-probe. In the case of λSM8, the 3.5 kb (+0.49 kb of the vector left-arm) *Hin*dIII fragment was found to share homology with the 3′-probe.

Regions that hybridized to the full-length cDNA probe were found along the whole length of two of the inserts (Fig. 2; SM2 and SM6). In the case of SM6 an important finding is that the 0.5 kb *Eco*RI fragment is homologous, in part, with the cDNA. This means that a span of at least 7.2 kb of the insert is covered by the cDNA.

The 0.5 kb *Eco*RI fragment has been partially sequenced, using the chain-termination procedure of Sanger *et al.* (1977), by subcloning this fragment into the plasmid vector pUC13 (Messing & Vieira, 1982) and sequencing directly from the plasmid (Wallace *et al.*, 1981). The data (see Fig. 2*b*) show that the 0.5 kb *Eco*RI fragment contains sequences from both exon 1 and exon 2. Further information derived from hybridization of 5′- and 3′-specific oligomers to Southern blots of recombination phage DNA digests strengthens the evidence for two closely linked genes within SM6 (see below).

Northern blotting and hybridization of murine RNA species

Since the identity of the pSMG751-related gene product remained obscure, we decided to investigate whether the corresponding mRNA was tissue-specific. Preparations of RNA from several tissues were analysed by Northern blotting with a labelled cDNA probe.

The autoradiograph (Fig. 4) shows that the mRNA is a major product of the submaxillary gland and also that the sublingual gland appears to express msp36 at a low level; however, because of the latter gland's proximity to the submaxillary, this observation may be attributable, at least in part, to tissue contamination. The mRNA is not present in significant amounts in the other tissues tested. Prolonged exposure of the autoradiograph reveals immature, precursor RNA species (Fig. 4,

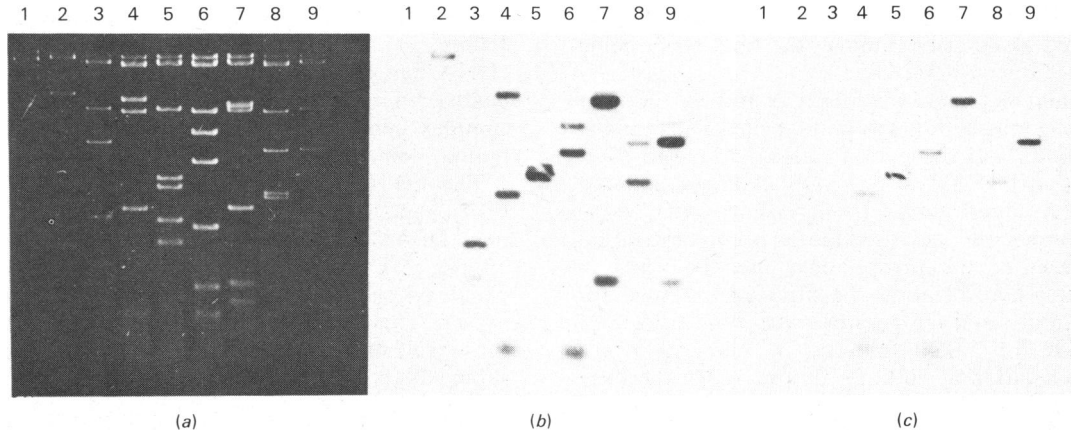


Fig. 3. *Physical mapping of recombinant phage containing msp36-related regions*
Tracks 1–9 are equivalent throughout all three parts of the Figure. They contain the following: track 1, SM3/*Xba*I (incomplete digest); track 2, SMA-*Bam*HI; track 3, SMA/*Hin*dIII; track 4, SM8/*Eco*RI; track 5, SM8/*Hin*dIII; track 6, SM6/*Eco*RI; track 7, SM6/*Hin*dIII; track 8, SM4/*Eco*RI; track 9, SM4/*Hin*dIII. (*a*) Represents the DNA bands in the gel after staining with ethidium bromide and u.v. irradiation. (*b*) Shows the same gel after transfer on to nitrocellulose and hybridization to ^{32}P-labelled SMG751 cDNA. In (*c*) the same nitrocellulose filter has been washed and hybridized with a 3′ cDNA fragment derived from a *Hin*cII digest of pSMG751.

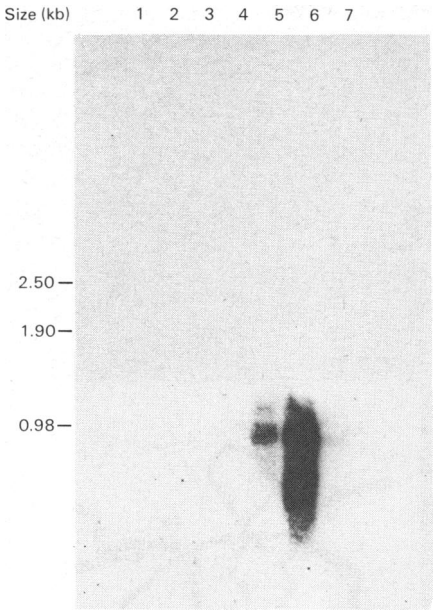


Fig. 4. *Northern analysis of a selection of murine tissues* A 20μg portion of the total RNA from various tissues was run per track. The tissues were: track 1, heart; track 2, skeletal muscle; track 3, liver; track 4, lung; track 5, sublingual gland; track 6, submaxillary gland; track 7, kidney. The RNA species were hybridized with SMG751 cDNA. The size markers (not shown) were prokaryotic 70S RNA, calf liver tRNA and eukaryotic rRNA. The sizes shown correspond to some of the msp36-related RNA precursors.

track 6), the sizes of which are approx. 3.8, 2.5, 1.7 and 0.98kb (see below). Other bands are seen at 2.1, 1.9 and 0.86kb, which may also be due to related RNA species, but it is not possible at present to assign these to a gene structure.

Heteroduplex analysis of a pSMG751-related gene

Heteroduplexes of the genomic DNA of SM6 and linearized pSMG751 DNA (using *Bam*HI, which cuts within the pAT153 vector) showed three intron deletion loops, along with duplex formation (Fig. 5). By measuring the lengths of the phage vector arms and the pAT153 vector arms it is possible to deduce the relative orientation of the gene. The lengths of the four exons and three introns are indicated in Fig. 5. These are consistent with the lengths of the RNA precursors (Fig. 4) and suggest that the order of splicing of the introns at 5′ to 3′.

Hybridization of specific oligonucleotides to recombinant phage DNA

Synthetic DNA oligomers were prepared (Matthes *et al.*, 1984) homologous with 16-base regions in the 5′ and 3′ ends of the predicted coding segment of the msp36 cDNA. (5′-probe: 5′-GGAACTTCATGGTGAC-3′; 3′-Probe: 5′-ACGTTGTAAGAAATT-3′.)

Fig. 6 shows the regions that contain sequences related to the 5′- and 3′-specific oligonucleotides. As expected, the 5′-oligomer labels the 1.2kb *Hin*dIII fragment but also hybridizes to the 0.5kb *Eco*RI fragment. The 3′-oligomer hybridizes to a 1.4kb *Eco*RI/*Bgl*II fragment that may contain more than one 3′-related region. These findings are consistent with a 'tail-to-tail' arrangement of the msp36-related genes on SM6.

Other data (A. G. Craig, unpublished work) indicate that there are two closely linked genes within the insert of λSM2. The 5′-probe hybridizes to a 1.3kb *Hin*dIII fragment that is bounded on both sides by other regions that are homologous with the cDNA. The 3′-probe does not recognize any sequences within this recombinant phage, which implies that the two genes present are in a 'head-to-head' conformation with two 5′-ends in the 1.3kb *Hin*dIII segment.

Discussion

Although the physiological role of the major submaxillary protein, msp36, has not been determined, the primary structure indicates that it is a secretory glycoprotein (Windass *et al.*, 1984) and it may be involved in salivary lubrication. A major secretory glycoprotein with many of the characteristics of msp36 has been previously isolated from mouse submaxillary glands (Nieuw Amerongen *et al.*, 1978). Sequences homologous with the msp36 coding sequence are present in human DNA (A. G. Craig, unpublished work).

The observations presented indicate that the major mouse submaxillary product, msp36, is encoded by one or more genes that are part of a multigene family. The expression of the msp36 gene family is tissue-specific and has previously been shown not to be androgen-inducible.

It is difficult to estimate accurately the number of genes in this family, though at least five different classes of recombinant phages have been isolated, judged by the restriction maps of their DNA species and the differing degrees of homology with one cDNA of pSMG751 (as shown by different responses to stringent washing with 0.3 × SSC–0.05 × SSC, after hybridization to the recombinant phage DNA species). A different exon pattern found by R-loop analysis on a subfragment of one

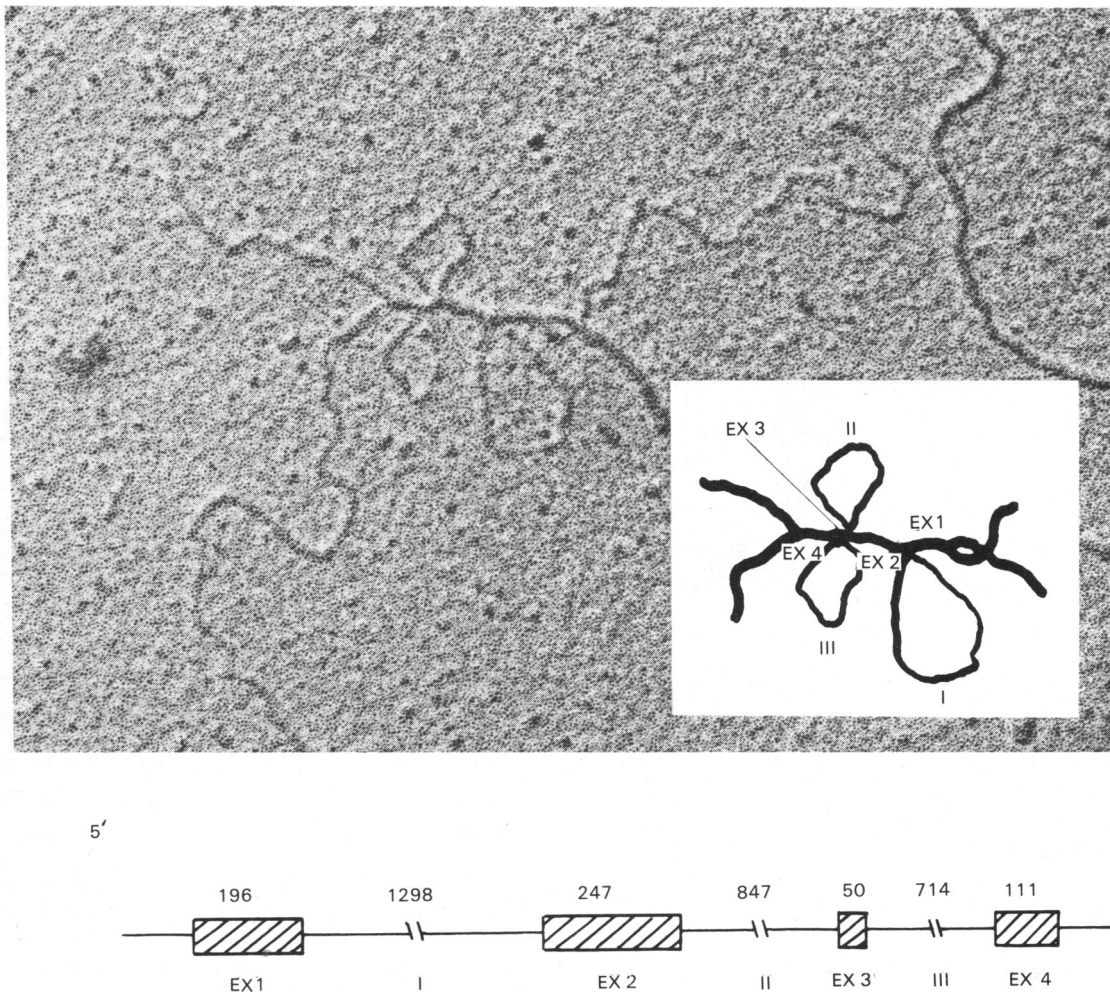


Fig. 5. *Heteroduplex analysis of an msp36-related gene*
Heteroduplex DNA molecules were formed between SM6 and pSMG751/*Bam*HI. The line drawing gives the sizes of the introns (I, II and III) and exons (EX 1, 2, 3 and 4) (shown in base-pairs).

of the recombinant phage DNA species may be due to fusion of some of the exons.

In at least four cases the individual genes are closely linked, since more than one gene is present on the same recombinant phage. Close linkage of this nature has been reported previously for the mouse embryonic β-globin genes (Hill *et al.*, 1984). This observation, together with the Southern-transfer analysis of genomic DNA, suggest there are about eight to ten genes in the family. There is no information on whether any of these might be pseudogenes. Sequencing of cDNA species prepared from submaxillary polyadenylated RNA needs to be carried out to determine how many genes of the family are transcriptionally active in the gland.

The structure of one of the msp36-related genes has been derived from heteroduplex mapping. This approach, allied to the use of specific DNA oligomers derived from the cDNA sequence presents a way of dissecting the genomic DNA contained within the recombinant phage. This is particularly relevant for this gene family, since it seems likely that the form of the gene shown by heteroduplexing is not common to the other members.

Other mouse gene families, for example the β-globin family, the mouse major urinary proteins (Bishop *et al.*, 1982) and the histocompatability locus (Hood *et al.*, 1982) have most, if not all, of their members on one chromosome. As yet it is not known if the genes corresponding to msp36 are

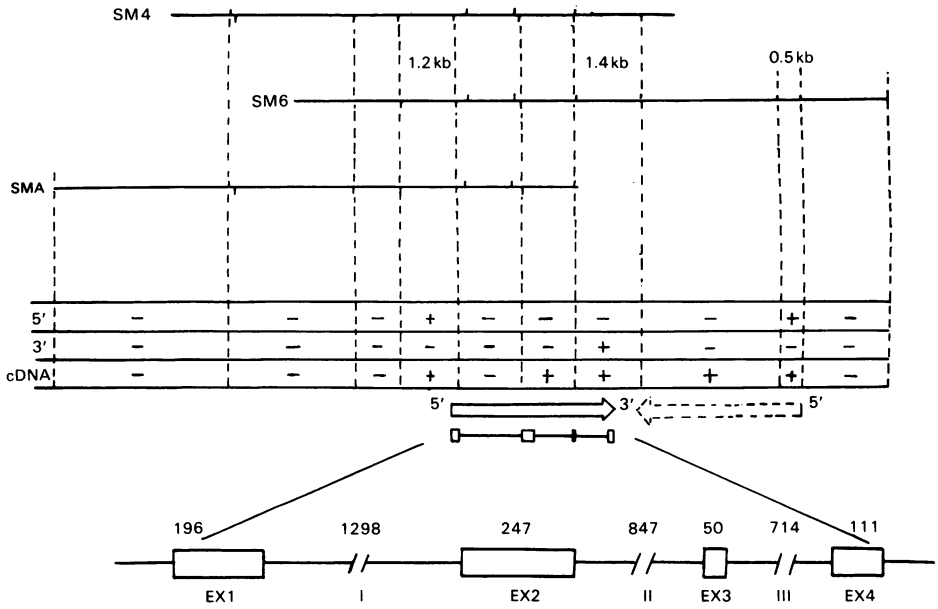


Fig. 6. *Oligonucleotides analysis of SM6.*
The regions which hybridize to the oligomers are indicated with a '+' sign. The DNA inserts have been subdivided by digestion with *Hin*dIII, *Eco*RI and *Bgl*II. The known msp36 gene (from heteroduplex mapping) has been drawn beneath the solid arrow, which indicates the direction of its transcription. The position and orientation of the predicted second gene is indicated by the dotted arrow.

located on the same chromosome, although the clear linkage of some of the genes suggests that they may be closely grouped.

We are grateful to Dr. John Windass, Dr. Linda Beecroft, Dr. David Burt, Professor Adrian Allen and colleagues in Leicester for valuable discussion, to Dr. Daniel Brenner and Mr. John Kyte for oligonucleotide synthesis, to Mr. Ian Riddell for photography and to the Science and Engineering Research Council for financial support.

References

Barka, T. (1980) *J. Histochem. Cytochem.* **28**, 836–859
Barrie, P. A., Jeffreys, A. J. & Scott, A. F. (1981) *J. Mol. Biol.* **149**, 319–336
Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182
Bishop, J. O., Clark, A. J., Clissold, P. M., Hainey, S. & Franke, U. (1982) *EMBO J.* **1**, 615–620
Burt, D. W. & Brammar, W. J. (1982) *Mol. Gen. Genet.* **185**, 462–467
Davis, R. W., Simon, M. M. & Davidson, N. (1971) *Methods Enzymol.* **21**, 413–428
Durham, D. M., Perrin, F., Gannon, F. & Palmiter, R. D. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6511–6515
Hill, A., Hardies, S. C., Phillips, S. J., Davis, M. G., Hutchison, C. A. & Edgell, M. H. (1984) *J. Biol. Chem.* **259**, 3739–3747
Hood, L., Steinmetz, M. & Goodenow, R. (1982) *Cell* **28**, 685–687
Jeffreys, A. J., Wilson, V., Wood, D., Simons, J. P., Kay, R. M. & Williams, J. G. (1980) *Cell* **21**, 555–564
Jeffreys, A. J., Barrie, P. A., Harris, S., Fawcett, D. H., Nugent, Z. J. & Boyd, A. C. (1981) *J. Mol. Biol.* **156**, 487–503
Loenen, W. A. M. & Brammar, W. J. (1980) *Gene* **10**, 249–259
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Matthes, H. W. D., Zenke, W. M., Grundstrom, T., Staub, A., Wintzerith, M. & Chambon, P. (1984) *EMBO J.* **3**, 801–805
Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
Mullins, J. J., Burt, D. W., Windass, J. D., McTurk, P., George, H. & Brammar, W. J. (1982) *EMBO J.* **1**, 1461–1466
Nieuw Amerongen, A. V., Aarsman, M. E. G., Vreugdenhil, A. P. & Roukema, P. A. (1978) *Biochim. Biophys. Acta* **534**, 26–37
Noyes, B., Mevarech, M., Stein, R. & Agarnal, K. L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1770–1774
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205
Wallace, R. B., Johnson, M. J., Suggs, S. V., Miyoshi, K., Bhatt, R. & Itakura, K. (1981) *Gene* **16**, 21–26
Windass, J. D., Mullins, J. J., Beecroft, L. J., George, H., Meacock, P. A., Williams, B. R. G. & Brammar, W. J. (1984) *Nucleic Acids Res.* **12**, 1361–1376