The EMBO Journal vol.3 no.5 pp.1045–1052, 1984

Structure of mouse major urinary protein genes: different splicing configurations in the 3′-non-coding region

A.John Clark[1]*, Patricia M.Clissold[1,3], Raya Al Shawi[1], Pam Beattie[2] and John Bishop[1]

Departments of [1]Genetics and [2]Molecular Biology, University of Edinburgh, Edinburgh EH9 3JN, UK

[3]Present address: Searle R. & D., P.O. Box 53, High Wycombe, UK

*To whom reprint requests should be sent

Communicated by J.Bishop

The multigene family which codes for the mouse major urinary proteins (MUPs) consists of ~35 genes. Most of these are members of two different groups, Group 1 and Group 2, which can be distinguished by nucleic acid hybridisation. Here we describe the structure of a Group 1 gene and show that two size classes of MUP mRNA which are found in mouse liver result from different splicing events in the 3′-non-coding region and contain different polyadenylation sites. Short mRNA is ~750 nucleotides long, contains six exons, and is the main product of the Group 2 genes. Long mRNA is ~880 nucleotides long, contains seven exons and is the main product of the Group 1 genes. Five exons and part of the sixth are common to long and short mRNA and contain the coding region. This codes for an acidic protein of 180 amino acids containing an 18 residue signal peptide. A comparison of the mouse sequence with a homologous rat α_{2u}-globulin sequence shows that the rate of evolutionary divergence of the two proteins has been high. Silent sites have diverged four times more rapidly than replacement sites, showing that there has been selection against change in the protein sequence.

Key words: major urinary proteins/genes/mRNA/splicing

Introduction

Rodents secrete a family of closely-related proteins known in the rat as the α_{2u}-globulins and in the mouse as the major urinary protein or MUP. MUP is the most abundant product of male mouse liver and MUP mRNA makes up ~5% of liver mRNA on a weight basis (Hastie and Held, 1978; Clissold and Bishop, 1981). MUP mRNA is also found in the lachrymal, salivary and mammary glands of the mouse, and the mRNA levels in the different tissues are variously influenced by androgen, insulin, growth hormone and thyroxine (Hastie *et al.*, 1979; Shaw *et al.*, 1983).

The MUPS are specified by a family of genes with ~35 members (Bishop *et al.*, 1982). Most or all of these are located on mouse chromosome 4 (Bennett *et al.*, 1982; Bishop *et al.*, 1982; Krauter *et al.*, 1982). The 35 genes can be subdivided on the basis of nucleic acid hybridisation experiments into two groups, each with ~15 members, and a small heterogeneous group made up of the remaining genes (Bishop *et al.*, 1982). Here we describe the structure of a Group 1 gene, BS-6 (Clark *et al.*, 1982). Mouse liver contains two main size classes of mRNA (Derman, 1981; Clissold and Bishop, 1982). We show that these result from different splicing configurations in the 3′-non-coding region and contain different polyadenylation sites. The short mRNA contains six exons and the long mRNA contains seven. Most Group 1 mRNAs are of the long type and most or all Group 2 mRNAs are of the short type.

Results

Two size-classes of MUP mRNA

Mouse liver contains two size-classes of MUP mRNA (Derman, 1981; Clissold and Bishop, 1982). Previously we estimated their sizes to be 1.2 and 1 kb (Clissold and Bishop, 1982). We have now re-evaluated these sizes by means of Northern blots using as markers herpes virus thymidine kinase mRNA (1.31 kb, McKnight, 1980) and rabbit β-globin mRNA (0.59 kb, Hardison *et al.*, 1979) and find them to be ~0.90 and 0.77 kb, exclusive of poly(A).

Basic 7-exon structure of a MUP gene

The available amino acid sequence is a partial N-terminal sequence of 36 amino acids, or ~20% of the molecule (Finlayson *et al.*, 1974). Furthermore, we have no full-length cDNA clones. We therefore compared the complete sequence of the MUP gene BS-6 with the published sequences of the exons of an α_{2u}-globulin gene (Unterman *et al.*, 1981; Dolan *et al.*, 1982; Laperche *et al.*, 1983). This comparison showed that the two genes have the same basic 7-exon structure (Figures 1 and 2, exons 1–5, 6a and 7). Exons 2–5 and 6a are identical in length in both cases, while exons 1 and 7 differ in length by a few nucleotides. The overall sequence homology within the exons is 81%. We attempted to compare up to 10 nucleotides of the intron at each splice-junction. The 106 positions which could be compared in this way show one two-nucleotide deletion or insertion and 13 base substitutions. All of the proposed splice-junctions conform to the general consensus (Breathnach and Chambon, 1981).

The gene contains an open 543-nucleotide reading-frame, identical in position and length (excepting a 3-nucleotide deletion) to the reading frame of the α_{2u}-globulin gene. This deletion occurs in the signal peptide region which in BS-6 codes for 18 amino acids but codes for 19 in α_{2u}-globulin genes (Drickamer *et al.*, 1981; Laperche *et al.*, 1983). The gene codes for a mature protein of 162 amino acids with a mol. wt. of 18 730. The sequence of the 36 N-terminal residues shows no discrepancies with the partial N-terminal MUP-1 sequence of Finlayson *et al.* (1974). 66% of the residues are identical to those of the rat protein.

The mRNA cap site has not been determined definitively. We have provisionally placed it 31 nucleotides downstream of the TATA box (Figure 2). This positioning, which is supported by other lines of evidence (see below), places it close to the proposed cap site of α_{2u}-globulin mRNA (Laperche *et al.*, 1983).

Short and long MUP mRNAs contain 6 and 7 exons respectively

The coding region ends at nucleotide 26 of exon 6. Downstream of this position two different sets of splicing and polyadenylation events occur, and these generate the two size-classes of mRNA. This conclusion is partly based on the

© IRL Press Limited, Oxford, England.

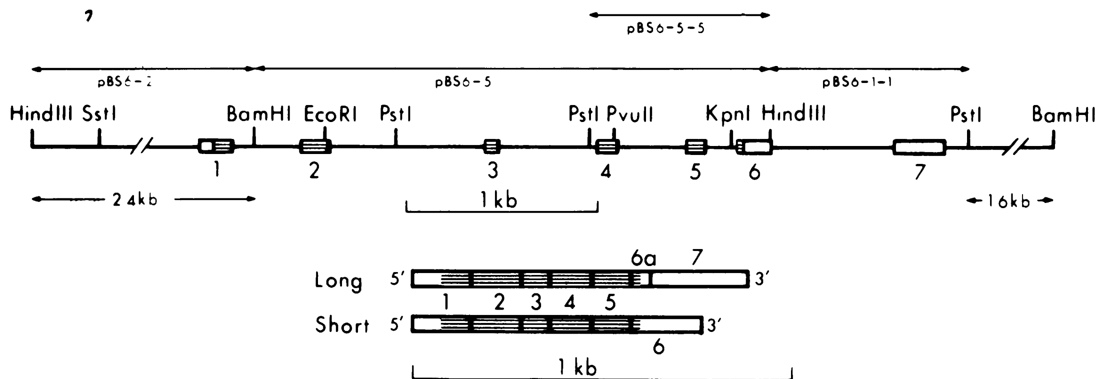

Fig. 1. Structure of MUP gene BS-6. **Above**, the mouse DNA insert of plasmid pBS6-1 showing the size and arrangement of the exons of BS-6. The exons are shown as boxes, and the coding regions are shaded. Four subclones of pBS6-1 are shown. **Below**, the two main MUP mRNAs showing their structural relationship to BS-6.

nucleotide sequences of two cloned MUP cDNAs, LVA132 and LVA325. Both lack the 5′ end of the mRNA sequence, but both contain the exon 5 – exon 6 splice. LVA132 contains the splice between exon 6a and exon 7 (nucleotide 46 of exon 6 to nucleotide 1 of exon 7). In contrast, LVA325 continues into the intron between exon 6a and exon 7 and terminates without splicing to exon 7 (Figure 2). The structure of these cDNA clones suggests that the difference between long and short mRNA is that one contains exon 6a and exon 7, and the other exon 6. The evidence presented below strongly reinforces this interpretation.

Short mRNA. Both LVA325 and a second cloned cDNA, LVA318, which is coextensively identical, terminate at nucleotide 166 of exon 6, suggesting that this is the polyadenylation site of the short mRNA. However, these clones were made with A-tailed double-stranded cDNA, with the result that we cannot tell whether or not they contain the 3′-terminal poly(A) of the mRNA. We therefore used S1-mapping to determine the 3′ end of the short mRNA.

The sequence of MUP gene BL-1 is identical to LVA325 and very similar to BS-6 throughout exon 6, and it contains a convenient *Sal*I site in the 3′ part of exon 6, i.e., the part not shared with exon 6a. We constructed a plasmid subclone of BL-1 containing the 3′ region of the gene (exons 4 – 7) together with 1.7 kb of 3′-flanking region (see Figure 3). This was digested with *Sal*I and 3′ end-labelled, and the *Sal*I-*Bam*I fragment was isolated. This was annealed with liver mRNA and challenged with nuclease S1. A marker fragment was prepared by cleaving the labelled fragment with *Hind*III. This marker, 90 nucleotides long, corresponds to nucleotide 172 of exon 6 (Figure 3). The minor fragment, 1 nucleotide shorter (Figure 3), presumably reflects partial failure of the polymerase to complete the blunting of the *Sal*I site. Two clusters of fragments were protected by mRNA from digestion by nuclease S1. These place two mRNA termini at nucleotides 158 – 160 and 176 – 178 of exon 6. No significant fragments correspond to sites close to nucleotide 166 where the 3′ end of LVA325 and LVA318 occurs, suggesting that these cDNA clones were truncated by ~10 nucleotides at some stage in the cloning procedure. Twenty-one nucleotides upstream of the terminus at 176 – 178, BS-6 contains the cleavage signal AATAAA (Proudfoot and Brownlee, 1976; Montell *et al.*, 1983). The corresponding sequence in BL-1 and LVA325 is GATAAA, which is not known to be a cleavage signal. Cleavage signals other than AATAAA have, however, been described (Nevino, 1983). One of these variant signals, ATTAAA (Jung *et al.*, 1980; MacDonald *et al.*, 1980; Hobart *et al.*, 1980; Unterman *et al.*, 1981; Cann *et al.*, 1982) occurs 18 nucleotides upstream of the second terminus located at nucleotides 158 – 160. Thus, cleavage signals are located at appropriate distances upstream of both terminal clusters.

Long mRNA. Nucleotide 46 of exon 6 is spliced to the 251 nucleotide long exon 7. We refer to the 26 nucleotide coding sequence of exon 6 plus the 20 nucleotides up to the splice junction as exon 6a. This is the configuration found in several α_{2u}-globulin cDNA clones (Dolan *et al.*, 1982). The MUP cDNA clone LVA132 contains this splice junction (Figure 2). LVA132 does not continue to the end of exon 7, but ends at nucleotide 171. This may be a case of truncation by nuclease S1 since the end-point is in the middle of an AT-rich region, and no recognised upstream cleavage signal is present.

According to this interpretation, short mRNA contains the 3′ part of exon 6 (that part which is not shared with exon 6a) and lacks exon 7, while long mRNA contains exon 7 but lacks the 3′ part of exon 6 (Figure 1). This model was tested by hybridising Northern blots of liver mRNA with selected DNA probes (Figure 4). Plasmid pBS6-1-1 carries a region to the 3′ side of exon 6 which includes exon 7. This plasmid hybridised to long but not to short mRNA (Figure 4, lane 2). The 86 nucleotide *Sal*I-*Hind*III fragment (Figures 2 and 4) was prepared from a subclone of BL-1. This probe is specific for the 3′ part of exon 6 and hybridised with short but not with long mRNA (lane 3). Plasmid pBS6-5-5, which spans the entire region containing exons 4, 5 and 6, hybridised with both long and short mRNA (lane 1).

From the sequence data, the main processed transcription products of the MUP genes, long mRNA and the two forms of short mRNA, are 878, 759 and 741 nucleotides long respectively, in agreement with the 900 and 770 nucleotide lengths estimated by gel electrophoresis of the mRNA.

DNA-mRNA heteroduplexes contain exons 1 – 5 and 7

Heteroduplexes between hybrid-selected MUP mRNA and pBS6-1, a plasmid subclone of BS-1 (Figure 1), were examined in the electron microscope. About half of the molecules that were analysed contained the same structure (Figure 5). This agrees with the proposed structure of long mRNA (Figure 1), given that the short (46-nucleotide) exon 6a had not formed a heteroduplex. The estimated sizes of exons 1 – 5

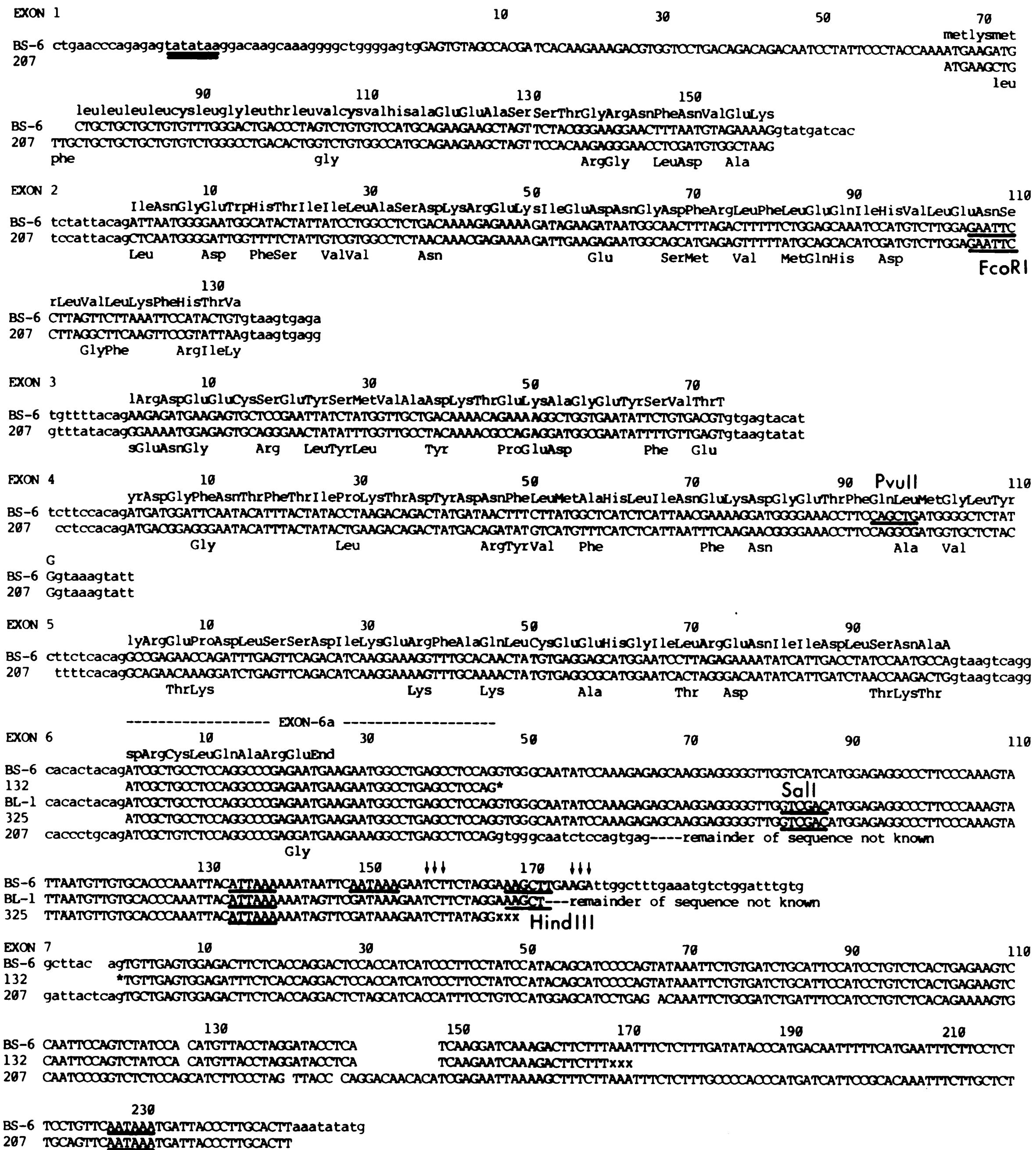

Fig. 2. MUP gene sequence. BS-6 and BL-1 (exon 6 only) are two Group 1 MUP genes (Clark *et al.*, 1982). 207 is a rat α_{2u}-globulin sequence from Dolan *et al.* (1982) and Laperche *et al.* (1983). (LVA)132 and (LVA)325 are two cloned MUP cDNAs (Clissold and Bishop, 1981). The rat and mouse sequences are aligned, and the exons of the mouse sequence are separately numbered. Exons are shown in upper case and flanking regions in lower case lettering, except that the bifunctional 3′ part of exon 6 is shown in upper case lettering. The exon 6a-exon 7 splice junctions are shown by *. The TATA box is double underlined and polyadenylation signals are underlined singly; ↓, 3′ termini of short mRNA (Figure 3); xxx, 3′ ends of LVA132 and LVA325. The amino acid sequence of BS-6 is shown together with those 207 amino acids that differ from BS-6. The signal peptide amino acids are given lower case initials.

and 7, based on the analysis of 22 molecules, are listed in Table I.

The remaining 28 molecules examined include four that are also consistent with the structure of long mRNA, given that neither exon 5 nor 6a had formed a heteroduplex, thus making a D-shaped structure. Two additional molecules may be heteroduplexes containing short mRNA. Neither of these groups is large enough to warrant detailed analysis. We expect the majority of molecules to contain long mRNA because the majority of liver MUP mRNA is of the long type (Figure 4, lane 1).

Nuclease S1 mapping confirms the 6 and 7 exon structures

Plasmid subclones of MUP genes were nick-translated and then repaired with DNA ligase to ensure that they contained long unbroken DNA sequences. Four restriction fragments were isolated from one plasmid and the others were linearised. Each fragment or plasmid was annealed with male mouse liver mRNA and treated with nuclease S1, and the protected fragments were sized.

The four restriction fragments of pBS6-1 span BS-6 from the *Sst*I site 1.7 kb upstream of the cap site to the *Pst*I site downstream of exon 7 (Figure 6). Two linearised plasmid

subclones overlap the fragments in the 3′-terminal region containing exons 4, 5, 6 and 7. Together, the fragments and subclones define five regions of BS-6 (Figure 6). Two plasmid subclones of a second Group 1 MUP gene, BL-1, span the 3′-terminal region.

The results of these experiments are summarised in Figure 6. Region I yielded two major nuclease-resistant fragments, 160 and 132 nucleotides in length (Figure 7, lane 2′), which presumably correspond to exons 1 and 2 (Table I). A minor 150 nucleotide fragment was also observed. This minor fragment could have been produced in any one of several different ways. It could represent a genuine alternative exon generated by a different cap-site or splice junction, or an additional exon present in a minority of MUP genes. Alternatively, it may be an artefact caused by nuclease S1 attack, either on a mismatched heteroduplex (between BS-6 and the transcription product of another MUP gene) or on an AT-rich heteroduplex region. In fact, there is an AT-rich region located ~150 nucleotides downstream of the proposed exon 1 cap site, which tends to support the last explanation. Region II yielded a single resistant fragment, 68 nucleotides in length, in agreement with the length of exon 3. Region III yielded a single fragment, ~100 nucleotides in length, which we take to contain exons 4 and 5 (111 and 102 nucleotides, Table I).

Region IV is defined by the difference between pBS6-5-5 and the 0.75-kb *Pst*I-*Kpn*I fragment (region III) and also by the difference between the 1.3-kb *Kpn*I-*Pst*I fragment and

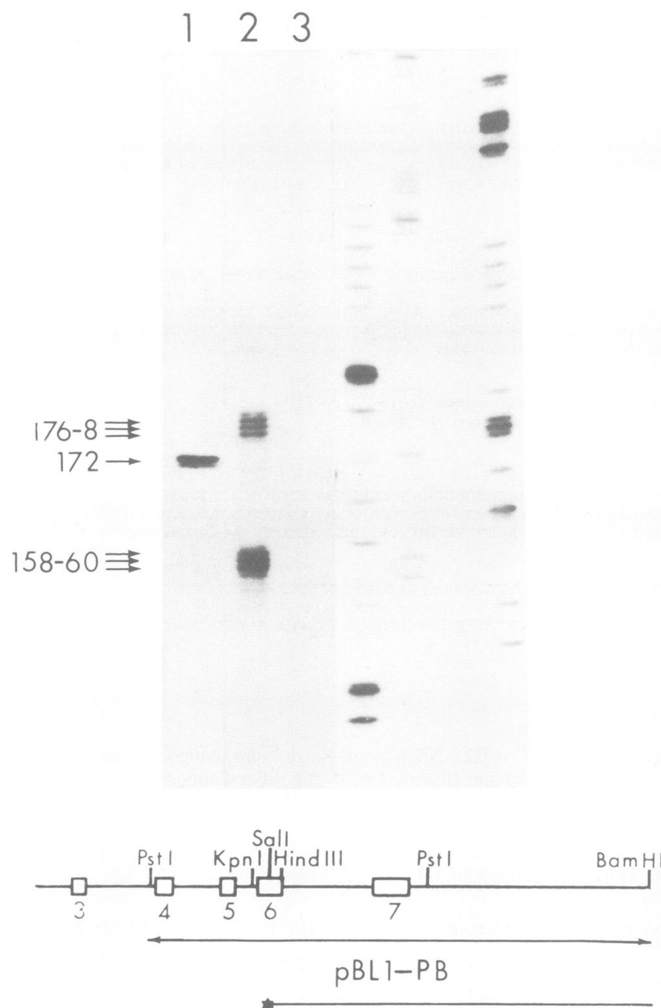

Fig. 3. Location of the 3′ ends of small mRNA. **Below:** A partial map of MUP gene BL-1, showing the portion cloned in plasmid pBL1-PB and the end-labelled probe used in the experiment. **Above:** Acrylamide gel autoradiograph. **Lane 1**, 90-base *Sal*I-*Hind*III fragment of the probe. **Lane 2**, fragments protected from nuclease S1 digestion by annealing with liver mRNA. **Lane 3**, control, annealed with tRNA. Remaining lanes, sequencing ladder. The numbers at the left hand side show the distances of the ends of the fragments from the 3′-terminal A-residue of the filled-in *Sal*I site.

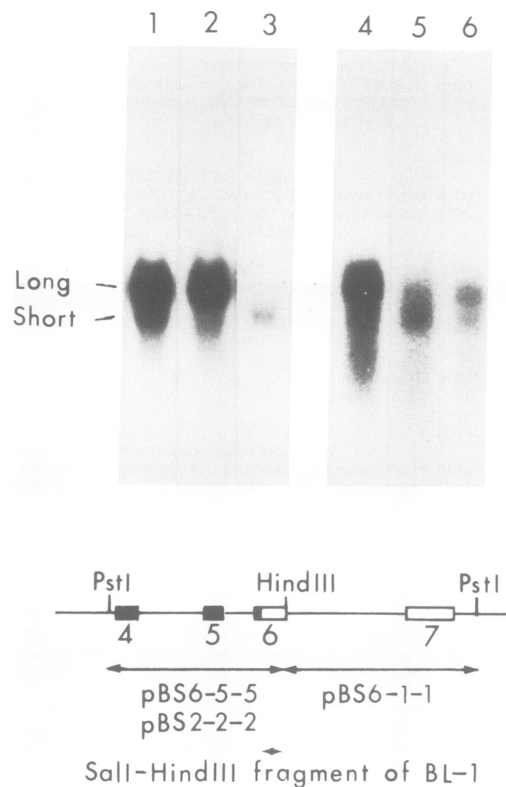

Fig. 4. Northern blots of male mouse liver mRNA probed for MUP sequences. **Lanes 1–3:** 1 μg total poly(A) mRNA probed with (1) pBS6-5-5; (2) pBS6-1-1; (3) *Sal*I-*Hind*III fragment of BL-1. Final wash 0.5 x SET, 70°C. **Lanes 4–6: 4**, 1 μg mRNA probed with pBS6-5-5 (Group 1), exposed 2 d. **5**, 2 μg mRNA probed with pBS2-2-2 (Group 2), exposed 14 d. Final wash 0.2 x SET, 70°C. **6**, same filter as 5 probed again with pBS2-2-2 and washed finally in 0.5 x SET.

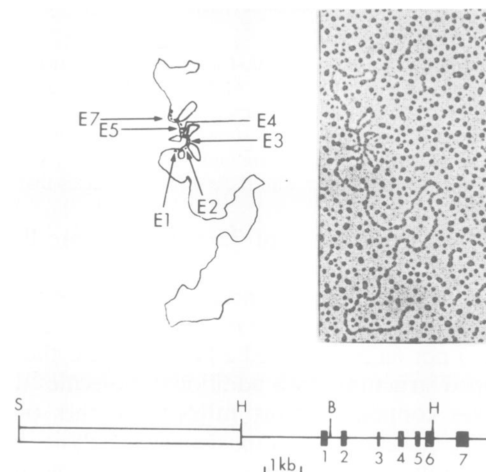

Fig. 5. Heteroduplex between pBS6-1 and hybrid-selected MUP mRNA. **Below**, a diagram of pBS6-1 shown cleaved at the plasmid *Sal*I site. Filled boxes, exons; open boxes, plasmid DNA; H, *Hind*III; B, *Bam*HI; S, *Sal*I.

pBS6-1-1 (region V). It contains most of exon 6 (166 nucleotides) plus 28 nucleotides on the 5′ side. Two fragments, 170 and 45 nucleotides long, were generated by pBS6-5-5 but not by the 0.75-kb *Pst*I-*Kpn*I fragment. A 40-nucleotide fragment was generated by the 1.3-kb *Kpn*I-*Pst*I fragment but not by pBS6-1-1. Presumably the 170-nucleotide fragment was present in the products of the *Kpn*I-*Pst*I fragment but obscured by the 175-nucleotide fragment which is generated in region V. Thus region IV, which is mainly exon 6, generated two fragments of ~45 and 170 nucleotides. Presumably the 45-nucleotide fragment represents exon 6a, and the 170-nucleotide fragment represents the two forms of exon 6, the longer of the two truncated by five nucleotides through cleavage at the *Hind*III site.

Table I. Sizes of BS-6 exons (lengths in nucleotides)

Exon	Sequence data		Electron microscope	Nuclease S1[c]
	Long mRNA	Short mRNA		
1	161[a]		159 (26)[b]	160, 130, 150
2	134		164 (47)	
3	74		125 (35)	68
4	111		128 (24)	105
5	102		186 (51)	
6a	46			
6		159, 177		45, 170
7	250		267 (84)	265, 175, 200

[a]Exons 1–5 are common to long and short mRNA.
[b]Mean and standard deviation.
[c]The sets correspond to the five regions shown in Figure 5.

Region V generated three fragments, 260, 175 and 200 nucleotides in length, in order of decreasing quantity. The 260 nucleotide fragment presumably corresponds to exon 7 (Table I). The other fragments could represent additional exons, or result from mismatching between exon 7 of BS-6 and the homologous regions of mRNA molecules transcribed from other MUP genes, or from nuclease S1 attack on AT-rich regions of the exon 7 heteroduplex. In fact, AT-rich regions are found ~175 and 200 nucleotides from the 5′ end of exon 7 (Figure 2). Nuclease S1 attack on these regions could have generated the minor fragments and also the 3′ end of cDNA clone pLVA132 (Figure 2). On the other hand, Unterman *et al.* (1981) have demonstrated that the 3′ end of some α_{2u}-globulin mRNA molecules is at a position 5 nucleotides downstream of the 3′ end of LVA132. These relate, however, to a known upstream cleavage signal (AATTAAA). The corresponding sequence in pLVA132 is AATCAAA (Figure 2), which is not known to be a cleavage signal. Thus it

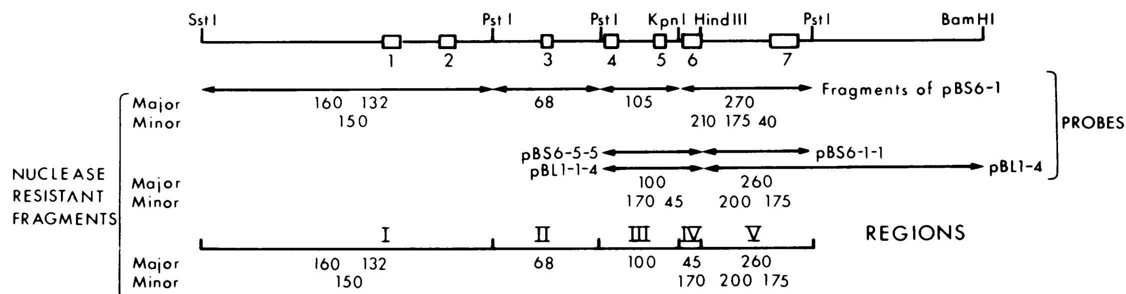

Fig. 6. Assignment of protected DNA fragments to sequenced exons. The different probes shown were labelled, annealed with male liver poly(A) mRNA, and challenged with nuclease S1. The sizes of the protected DNA fragments (see Figure 7) are shown beneath the corresponding probes. The five regions defined by the probes and the sizes of the protected fragments assigned to each are shown on the bottom line.

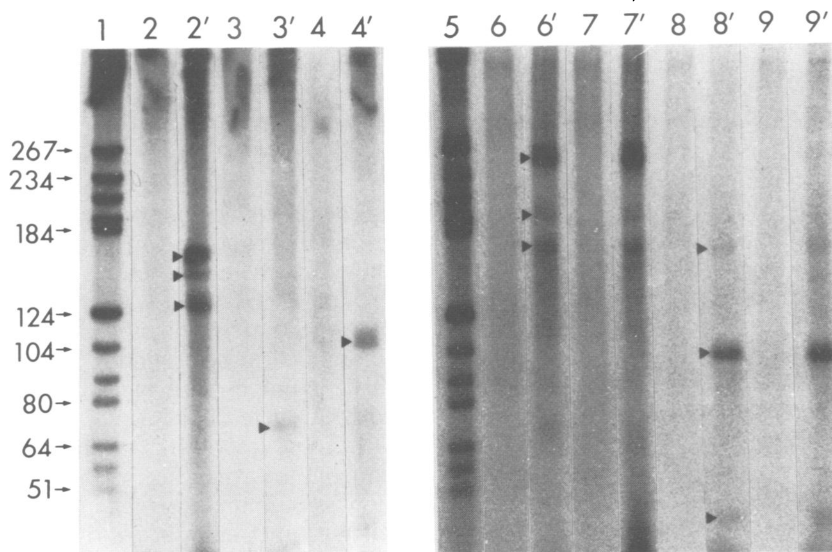

Fig. 7. Probe fragments protected by MUP mRNA. **Tracks 1** and **5**, pBR322 x *Hae*III markers. Other tracks – probes annealed with *Drosophila* rRNA (2 etc.) or mouse liver poly(A) mRNA (2′ etc.). **2**, Region I probe; **3**, Region II probe; **4**, Region III probe; **6**, pBS6-1-1 and **7**, pBL1-4 (both Region V); **8**, pBS6-5-5 and **9**, pBL1-1-4 (both Regions III + IV).

is likely that the 3′ end of pLVA132 was produced by nuclease S1 digestion of a full-length exon 7. However, this argument relates specifically to pLVA132 and does not eliminate the possibility that the 170 and 200 nucleotide fragments are true variants of exon 7. They may possibly represent products of MUP genes which have true cleavage signals near these positions.

Similar subclones of two different MUP genes gave identical results. These contain regions III + IV and region V (Figure 7). This result reflects the close similarity of the Group 1 MUP genes (Clark *et al.*, 1982; Bishop *et al.*, 1982), but does not preclude the possibility that other members of the family may differ to the small extent entailed by the possession of alternative cleavage sites.

Group 2 MUP genes mainly generate short mRNA

Subclone pBS6-5-5 and pBS2-2-2, the homologous subclone of a Group 2 gene (BS-2), were originally used to identify Group 1 and Group 2 genes (Bishop *et al.*, 1982). The Group 1 probe gives a much stronger hybridisation signal with long mRNA than with short (Figure 4, lane 4). Thus, as reported before (Clissold and Bishop, 1982), Group 1 genes mainly generate long mRNA. That they also generate some short mRNA is shown by the short-form Group 1 cDNA clones pLVA318 and pLVA325. The Group 2 probe, on the other hand, gives a stronger hybridisation signal with short mRNA (Figure 4, lane 5). The signal strengths show that liver contains much more long mRNA than short and that all or nearly all of the Group 2 transcripts in liver are short.

Note the relative hybridisation of long and short mRNA with the pBS6-5-5 probe in Figure 4, lanes 1 and 4, and with the pBS2-2-2 probe, lanes 5 and 6. These differences are due to the different salt concentrations used in the final wash. Group 1 x Group 2 heteroduplexes are quite stable in 0.5 x SET at 70°C, but melt in 0.2 x SET (Clark *et al.*, 1982). Consequently, the Group 1 probe (pBS6-5-5) shows relatively more hybridisation with short mRNA in lane 1 than in lane 4 and the Group 2 probe (pBS2-2-2) shows more with long mRNA in lane 6 than in lane 5.

Discussion

A common structure for MUP genes

All of the available evidence indicates that BS-6 is a true gene, i.e., one that is transcribed and translated, rather than a pseudogene. It contains a correctly positioned TATA box and 3′ termination signals. All of the designated donor and acceptor splice sites conform to the established consensus sequences, and in particular, all obey the GT-AG rule (Breathnach and Chambon, 1981). Finally, the gene contains a 181-codon open reading frame specifying a sequence which by all the available criteria corresponds to a MUP.

Of the 162 amino acids of the mature protein, 23 are basic and 34 acidic, in accordance with the low pI values of the MUPs (Clissold and Bishop, 1982). The very high (35%) content of charged residues is consistent with the very considerable charge heterogeneity of the MUPs (Hainey and Bishop, 1982; Clissold and Bishop, 1982). Unlike the α_{2u}-globulins, MUP BS-6 does not contain a site of attachment for N-acetylglucosamine. Szoka and Paigen (1978) showed that this is a general property of the MUPs.

It is likely that the 3.9-kb seven exon structure (Figure 1) is representative of all Group 1 MUP genes. Within this group precise restriction site homologies occur throughout the gene and extend several kilobases into both flanking regions (Clark *et al.*, 1982; Bishop *et al.*, 1982). Between Group 1 and Group 2 genes, equally precise homologies occur from the *Bam*I site in the first intron through the gene to the 3′ flanking region. Thus the structure of this region is likely to be the same in the Group 2 genes.

The analysis of a member of a multigene family by methods based on heteroduplexes made between the gene and a mixed population of mRNAs, might be extremely difficult in principle due to the formation of mismatched heteroduplexes with transcripts of other genes of the family. However, our EM and S1 nuclease experiments gave results which in most instances can be interpreted unambiguously in relation to the nucleotide sequence, showing that mismatching was not usually a significant factor. We observed only three nuclease-protected fragments which are not predicted by the data shown in Figure 2, all of them minor. Two of these were found in region V (3′) and one in region I (5′). All three can be explained by nuclease S1 attack on intra-exonic AT-rich regions recognisable in the BS-6 sequence (although this is not necessarily the correct explanation of their origin).

This situation would occur if the heterologous mRNAs are so divergent that no heteroduplexes are formed, or if they are so alike that the behaviour of the mismatched heteroduplexes is very similar or identical to that of the perfectly matched heteroduplexes. The close restriction site homology of the MUP genes strongly favours the latter interpretation. Furthermore, S1 mapping of two Group 1 genes (BS-6 and BL-1) in the region containing exons 4 – 7 gave identical results. Thus these exonic regions, at least of the Group 1 MUP genes from which most of the liver MUP mRNA is transcribed, must be very similar.

Alternative splicing configurations

We have shown that there are two main forms of MUP mRNA. Short mRNA contains exons 1 – 6, while long mRNA contains exons 1 – 5, a short 5′ region of exon 6 (exon 6a) and exon 7. The greater part of liver MUP mRNA is long mRNA, and nearly all of this is transcribed from Group 1 genes. That several genes are transcribed is shown by the multiplicity of proteins synthesised in a cell-free system in response to MUP mRNA (Szoka *et al.*, 1980; Clissold and Bishop, 1982; Shaw *et al.*, 1983). The mRNA transcribed mainly from Group 2 genes is short mRNA. Note, however, that the correlation of long mRNA with Group 1 genes and short mRNA with Group 2 genes is not perfect. Two Group 1 cDNA clones are known to correspond to long mRNA but two are also known to correspond to short.

The preponderance of long mRNA among Group 1 transcripts and short mRNA among Group 2 transcripts does show that there is at least a very strong tendency for the individual genes to be transcribed either in one mode or the other. As yet we do not know whether any MUP gene is transcribed in both modes. It may be possible to approach this question by studying MUP expression in cells transformed with single cloned genes.

In some cases different mRNA configurations are produced from the same gene by the use of alternative polyadenylation sites and 3′-terminal splice junctions. IgM and IgD mRNAs are transcribed in this way from the immunoglobulin heavy chain gene (Maki *et al.*, 1981). Membrane-bound and secreted forms of IgM with different C-terminal ends are coded for by mRNAs produced in the same way (Alt *et al.*, 1982; Early *et al.*, 1980). Other examples of the same phenomenon

lead to the synthesis of different forms of growth hormone (De Noto *et al.*, 1981), calcitonin (Rosenfeld *et al.*, 1982), γ-fibrinogen (Crabtree and Kant, 1982) and αA-crystallin (King and Piatigorsky, 1983). In all of these cases the different splicing and termination events generate mRNA molecules that code for variants of a protein. In the case of the MUP genes, on the other hand, the alternative splicing configurations are confined to the 3'-non-coding region. In some instances, mRNAs of different lengths that code for the same protein are due to variation in the length of the final exon. Examples are the chicken X gene (Hellig *et al.*, 1980), the mouse α-amylase (Tosi *et al.*, 1981) and dihydrofolate reductase (Setzer *et al.*, 1982) genes, and the rat α_{2u}-globulin genes (Unterman *et al.*, 1981). The two forms of short MUP mRNA are another example of this phenomenon.

In at least some cases, transcription proceeds beyond the polyadenylation site (Hofer and Darnell, 1981). Both splicing patterns (Treisman *et al.*, 1982, 1983) and cleavage/polyadenylation sites (Montell *et al.*, 1983) can be altered by single-base mutational changes. Thus, it is perhaps surprising that alternative splicing configurations in 3'-non-coding regions are not more common. The rarity of the phenomenon raises the possibility that the 3'-non-coding regions of genes play a hitherto unsuspected role in gene expression. In this connection we note the remarkable heterogeneity of the MUP and α_{2u}-globulin gene families in tissue-specific expression and response to hormonal induction (Shaw *et al.*, 1983; Roy *et al.*, 1982; Roy, 1983; Laperche *et al.*, 1983; Vandoren *et al.*, 1983).

Evolution of rodent urinary proteins

We can compare the sequence of BS-6 with that of the rat α_{2u}-globulin gene 207 over exons 1 – 5, 6a and 7. Differences occur in 154 out of 806 common positions (18.9%) and in addition six deletion or insertion events are proposed to maximise homology. One such event is located in the signal peptide region (3 nucleotides) and five in exon 7 (four of one nucleotide and one of 10 nucleotides). The two longest deletions or insertions are bounded by 3-base direct repeats (Efstratiadis *et al.*, 1980). In the coding region 108/543 (19.9%) of bases and 61/180 (33.9%) of amino acids differ. This agrees with the degree of homology deduced from the stability of DNA-RNA hybrids between MUP cDNA and α_{2u}-globulin mRNA (Hastie *et al.*, 1979).

The corrected frequencies of silent site and replacement site differences (Perler *et al.*, 1980) are 91% and 23% respectively. Taking the time since the mouse-rat divergence to be 30 M years, the unit evolutionary periods (time to produce 1% sequence divergence) are 0.33 M years and 1.31 M years respectively. These represent very high rates of divergence (see, e.g., Perler *et al.*, 1980; Efstratiadis *et al.*, 1980). Nevertheless, they do show the effect of selective pressure against evolutionary change in the sequence of the proteins. Thus, although the function of the rodent urinary proteins is not known, the relative conservation of replacement sites makes it clear that they do have a function which is sufficiently important to exert selective pressure against change.

The wide divergence of the mouse and rat genes is in contrast to the marked similarity of the different genes within each species. For example, based on homology between 11 intragenic restriction sites, six MUP genes (four Group 1 and two Group 2) differ on average in 5% of nucleotides (Bishop *et al.*, 1982). Thus inter-species divergence has gone hand-in-hand with intra-species conservation. This suggests that the gene family is in flux, with loss of some members balanced by duplication (by whatever means) of others, or alternatively that intra-species homogeneity is maintained by gene conversion (Slightom *et al.*, 1980; Shen *et al.*, 1981).

Materials and methods

Cloned DNA

The propagation of bacteriophage and plasmid clones and the isolation of DNA were carried out as described before (Clissold and Bishop, 1982; Clark *et al.*, 1982; Bishop *et al.*, 1982).

mRNA

Poly(A) RNA was isolated from the endoplasmic reticulum of male BALB/c mice as described (Clissold and Bishop, 1981) but omitting the sucrose gradient step. MUP mRNA was hybrid-selected using pLVA325 DNA bound to DBM-paper discs (20 μg/11 mm diameter disc, Stark and Williams, 1979). 20 μl of a solution containing 450 mM NaCl, 45 mM Na_3-citrate, 0.1% SDS, 50% formamide, 2 μg poly(A) and 10 μg poly(A) RNA was annealed with each disc for 12 h at 55°C, washed twice (15 min each) in 150 mM Na_3-citrate, 0.1% SDS at 45°C, once in 2 mM EDTA, 0.1% SDS for 15 min at 22°C, and once in 2 mM EDTA for 1 min at 45°C. The hybridised mRNA was eluted with 100 μl H_2O per disc for 1 min at 90°C and twice ethanol-precipitated with 1 μg of guinea-pig tRNA carrier. The final RNA precipitate was dried and dissolved in H_2O.

Electron microscopy

Plasmid pBS6-1 was cleaved at the plasmid *Sal*I site and 5 – 10 μg was mixed with 10 – 20 μg of hybrid-selected MUP mRNA in 120 μl of 70% formamide, 0.4 M NaCl, 0.1 M Hepes, 10 mM EDTA, pH 7.8 (Chow *et al.*, 1977) and sealed in a capillary. This was kept at 85°C for 5 min to denature the DNA and then annealed for 5 h at 63°C. It was immediately flushed into 210 μl of hyperphase to give a final composition of 45% formamide, 90 μg/ml cytochrome c, 90 mM Tris, 9 mM EDTA, pH 8.5 and spread on a hypophase containing 15% formamide, 10 mM Tris, 1 mM EDTA, pH 8.5. Grids were stained in 10^{-4} M uranyl acetate and rotary shadowed with Pt. The internal size standards were pAT153 and M13 DNA.

DNA sequencing

Briefly, restriction fragments of plasmids pBS6-2, pBS6-5, pBS6-1-1, pBL1-1-4, LVA132 and LVA325 were cloned into M13 mp 7, 8 or 9 and sequenced by the dideoxynucleotide method essentially as described (Sanger *et al.*, 1977; Anderson *et al.*, 1980). 91% of the BS6-2 sequence shown was sequenced in both strands. Full details will be published later.

S1 nuclease mapping

End-labelled DNA (Figure 3). Plasmid pBL1-PB was digested with *Sal*I and 3' end-labelled with [^{32}P]dCTP and DNA polymerase I large fragment. The DNA was recovered by phenol and chloroform extraction, precipitation with ethanol and digestion with *Bam*HI. The 2.5 kb *Sal*I-*Bam*HI fragment was recovered by agarose electrophoresis and 700 ng was mixed with 20 μg of male mouse liver mRNA in 20 μl of 80% formamide, 0.4 M NaCl, 1 mM EDTA, 40 mM Pipes, pH 7.4, denatured for 15 min at 82°C and annealed for 3 h at 43°C. Samples were diluted with 0.3 ml of S1 buffer (0.28 M NaCl, 30 mM Na-acetate, pH 4.6, 4.5 mM $ZnCl_2$, 5% glycerol, 20 μg/ml denatured salmon sperm DNA, 400 U/ml S1 nuclease) and incubated for 30 min at 37°C. 50 μl of 4 M ammonium acetate, 0.1 M EDTA was added, and the samples were recovered by phenol extraction and ethanol precipitation and run on a sequencing gel.

Uniformly-labelled DNA (Figure 6). Plasmid DNA was nicked in a controlled way as described (Bishop and Davies, 1980) except that a series of identical DNA samples was nicked with serial 2-fold dilutions of DNAase I (highest concentration, 20 ng/ml). After 7 min at 20°C the reactions were stopped by adding EDTA to 5 mM. 0.5 μg of each digest was run on an agarose gel. The point at which 50% of the closed circular DNA was converted to open circles was taken to represent one nick per two strands and the sample most closely approximating 1 nick per 1.5 kb was chosen by making the assumption that the number of nicks made is linearly proportional to the input of DNAase. This DNA sample was recovered by phenol and chloroform extraction and precipitation with ethanol. About 200 nucleotides at each nick were replaced by using DNA polymerase I in the presence of [^{32}P]dCTP and the reaction was stopped by adding ATP to 0.1 mM and T4 DNA ligase. After a few minutes no further incorporation of [^{32}P]dCTP occurred due to ligation at the nicks. The DNA was again recovered with phenol, chloroform and ethanol and digested appropriately with restriction enzymes. In some cases labelled restriction fragments were recovered after agarose electrophoresis (Yung *et al.*, 1979). In all cases samples were run on a denaturing formamide gel to ensure that at least 50% of the single strands (linearised plasmid DNA or

isolated fragment) were intact.

About 50 ng of DNA (10^5 c.p.m.) was annealed with 1 μg of total male liver poly(A) RNA or *Drosophila* rRNA in a volume of 12 μl as described above (electron microscopy) and treated with S1 nuclease in 100 μl of S1 buffer for 30 min at 37°C. 200 nmol of EDTA, 2.5 μg of *Drosophila* rRNA, and 2 vol of ethanol were added. Samples were kept overnight at −20°C, collected by centrifugation and analysed using 7% acrylamide denaturing formamide gels (Maniatis *et al.*, 1975).

Northern blots

Samples containing 0.5−2 μg of poly(A) RNA from male BALB/c liver endoplasmic reticulum were run on 1.4% agarose gels containing formaldehyde and transferred to nitrocellulose filters as described (Clissold and Bishop, 1982). The filters were annealed with nick-translated plasmid DNA probes as described (Clissold and Bishop, 1982). The final wash at 68°C was either with 0.5 x SET or 0.2 x SET (1 x SET is 150 mM NaCl, 30 mM Tris, 1 mM EDTA, pH 8).

Acknowledgements

We are grateful for the assistance of Melville Richardson, Avril Duncan and Morag Robertson, and to Dr Norman Maitland for the gift of HSV-infected CV-1 cells. This work was supported by the Cancer Research Campaign and the Medical Research Council.

References

Alt,F.W., Rosenberg,N., Casanova,R.J., Thomas,E. and Baltimore,D. (1982) *Nature*, **296**, 325-331.
Anderson,S., Gait,M., Mayol,L. and Young,I.G. (1980) *Nucleic Acids Res.*, **8**, 1731-1745.
Bennett,K., Lalley,P., Barth,R. and Hastie,N. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1220-1224.
Bishop,J.O. and Davies,J.A. (1980) *Mol. Gen. Genet.*, **179**, 583-580.
Bishop,J.O., Clark,A.J., Clissold,P.M., Hainey,S. and Francke,U. (1982) *EMBO J.*, **1**, 615-620.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Cann,G.M., Zaritsky,A. and Koshland,M.E. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6656-6660.
Chow,L.T., Roberts,J.M., Lewis,J.B. and Broker,T.R. (1977) *Cell*, **11**, 819-836.
Clark,A.J., Clissold,P.M. and Bishop,J.O. (1982) *Gene*, **18**, 221-230.
Clissold,P.M. and Bishop,J.O. (1981) *Gene*, **15**, 225-235.
Clissold,P.M. and Bishop,J.O. (1982) *Gene*, **18**, 211-220.
Crabtree,G.R. and Kant,J.A. (1982) *Cell*, **31**, 159-166.
De Noto,F.M., Moore,D.T. and Goodman,H.M. (1981) *Nucleic Acids Res.*, **9**, 3719-3730.
Derman,E. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5425-5429.
Dolan,K.P., Unterman,R., McLaughlin,M., Nakhasi,H.L., Lynch,K.R. and Feigelsen,P. (1982) *J. Biol. Chem.*, **257**, 13527-13543.
Drickamer,K., Kwoh,J.J. and Kurtz,D.T. (1981) *J. Biol. Chem.*, **256**, 3634-3636.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **30**, 313-319.
Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., De Rich,J.K., Forget,G.G., Weissman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.S. and Proudfoot, N.J. (1980) *Cell*, **21**, 653-668.
Finlayson,J.S., Potter,M., Shinnick,C.S. and Smithies,O. (1974) *Biochem. Genet.*, **11**, 325-335.
Hainey,S. and Bishop,J.O. (1982) *Genet. Res. Camb.*, **39**, 31-39.
Hardison,R.C., Butler,E.T.,III, Lacy,E., Maniatis,T., Rosenthal,N. and Efstratiadis,A. (1979) *Cell*, **18**, 1285-1297.
Hastie,N. and Held,W. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1217-1221.
Hastie,N., Held,W. and Toole,J. (1979) *Cell*, **17**, 449-457.
Hellig,R., Perrin,F., Gannon,F., Mandel,J.L. and Chambon,P. (1980) *Cell*, **20**, 625-637.
Hobart,P., Crawford,R., Shen,L.P., Pictet,R. and Rutter,W.J. (1980) *Nature*, **288**, 137-141.
Hofer,E. and Darnell,J.E.,Jr. (1981) *Cell*, **23**, 585-593.
Jung,A., Sippel,A.E., Grez,M. and Schutz,G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5759-5763.
King,C.R. and Piatigorsky,J. (1983) *Cell*, **32**, 707-712.
Krauter,K., Leinwald,L., D'Eustachio,P., Ruddle,F. and Darnell,J. (1982) *J. Cell Biol.*, **94**, 414-417.
Laperche,Y., Lynch,K.R., Dolan,K.P. and Feigelson,P. (1983) *Cell*, **32**, 453-460.
MacDonald,R.J., Crerar,M.M., Swain,W.F., Pictet,R.L., Thomas,G. and Rutter,W.J. (1980) *Nature*, **287**, 117-122.
Maki,R., Roeder,W., Traunecker,A., Sidman,C., Wahl,M., Raschke,W. and Tonegawa,S. (1981) *Cell*, **24**, 353-365.
Maniatis,T., Jeffrey,A. and van de Sande,H. (1975) *Biochemistry*, **14**, 3787-3794.
McKnight,S.L. (1980) *Nucleic Acids Res.*, **8**, 5949-5963.
Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1938) *Nature*, **305**, 600-605.
Nevins,J.R. (1983) *Annu. Rev. Biochem.*, **52**, 441-466.
Perler,F., Efstratiadis,A., Lomedico,P., Gilbert,W., Kolodner,R. and Dodgson,J. (1980) *Cell*, **20**, 555-565.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-314.
Rosenfeld,M.G., Lin,C.R., Amara,S.G., Stolarsky,L., Roos,B.A., Ong,E.S. and Evans,R.M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1717-1721.
Roy,A.K. (1983) in Oppenheimer,J.H. and Samuels,H.H. (eds.), *Molecular Basis of Thyroid Action*, Academic Press, NY, pp. 213-243.
Roy,A.K., Chatterjee,B., Demyan,W.F., Nath,T.S. and Motwani,N.M. (1982) *J. Biol. Chem.*, **257**, 7834-7838.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Setzer,D.R., McGrogan,M. and Schimke,R.T. (1982) *J. Biol. Chem.*, **257**, 5143-5147.
Shaw,P.H., Held,W.A. and Hastie,N.D. (1983) *Cell*, **32**, 755-761.
Shen,S.-H., Slightom,J.L. and Smithies,O. (1981) *Cell*, **26**, 191-203.
Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell*, **21**, 627-638.
Stark,G.R. and Williams,J.G. (1979) *Nucleic Acids Res.*, **6**, 195-204.
Szoka,P.R. and Paigen,K. (1978) *Genetics*, **90**, 597-612.
Szoka,P.R., Gallagher,J.F. and Held,W.A. (1980) *J. Biol. Chem.*, **255**, 1367-1373.
Tosi,M., Young,R.A., Hagenbuchle,O. and Schibler,U. (1981) *Nucleic Acids Res.*, **9**, 2313-2323.
Treisman,R., Proudfoot,N.J., Shander,M. and Maniatis,T. (1982) *Cell*, **29**, 903-911.
Treisman,R., Orkin,S.H. and Maniatis,T. (1983) *Nature*, **302**, 591-596.
Unterman,R.D., Lynch,K.R., Nakhasi,H.L., Dolan,K.P., Hamilton,J.W., Cohn,D.V. and Feigelson,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3478-3482.
Vandoren,G., Mertens,B., Heyns,W., Van Baelen,H., Rombauts,W. and Verhoeven,G. (1983) *Eur. J. Biochem.*, **134**, 175-181.
Yung,R.C.A., Lis,J. and Wu,R. (1979) *Methods Enzymol.*, **68**, 176-182.