The EMBO Journal vol.4 no.12 pp.3167–3171, 1985

Analysis of mouse major urinary protein genes: variation between the exonic sequences of Group 1 genes and a comparison with an active gene outwith Group 1 both suggest that gene conversion has occurred between MUP genes

A.J.Clark[1], A.Chave-Cox, X.Ma and J.O.Bishop

Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK

[1]Present address: A.F.R.C. Animal Breeding Research Organisation, West Mains Road, Edinburgh, UK

Communicated by J.O.Bishop

Here we compare the exonic sequences of four Group 1 mouse major urinary protein (MUP) genes and four Group 1 cDNA sequences. These define seven different nucleotide sequences which differ from each other by 0.35% of bases on average, and which would code for seven different MUP proteins that could probably be resolved physically into at least five classes. The sequences differ at 13 nucleotide positions and at six codons, and although they are closely related their descent cannot be described by a simple series of duplications. We also describe the sequence of another liver cDNA (pMUP15) which has diverged from the Group 1 consensus sequence in 14.6% of bases. The divergence is much greater over exons 1–3 than over exons 4–6, suggesting that an ancestral gene conversion event has occurred. pMUP15 also differs from the Group 1 genes in having a longer signal peptide sequence and a different splice configuration between exons 6 and 7. Unlike the Group 1 sequences, pMUP15 contains a potential N-linked glycosylation site. Other published work has shown that a shorter cDNA clone which is identical over their common sequence to pMUP15 codes for MUP proteins that are unusually large in size and acidic in pI. We show here that mouse urine does indeed contain a glycosylated MUP protein with those properties, presumably the product of the gene that corresponds to pMUP15.

Key words: mouse/major urinary protein genes/variation/gene conversion

Introduction

Major urinary protein (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 submaxillary, sublingual, mammary and lachrymal glands (Hastie *et al.*, 1979; Shaw *et al.*, 1983). The protein is made up of a number of different components, many of which can be resolved on one- or two-dimensional gels. MUPs are under multihormonal control and different components display different patterns of hormonal regulation. For example, testosterone stimulates the expression of some genes more than others in female liver (Clissold *et al.*, 1984) and thyroxine and growth hormone induce the expression of different subsets of MUP genes (Knopf *et al.*, 1983). The mouse genome contains ~35 MUP genes (Bishop *et al.*, 1982), most of which can be classified into two groups (Group 1 and Group 2), each with ~15 members. The Group 1 genes are active, while the Group 2 genes are pseudogenes (Ghazal *et al.*, 1985). In the previous paper (Clark *et al.*, 1985) we presented the full sequence of the transcription unit of a Group 1 gene and a Group 2 gene. Here we present the exonic sequences of three more different Group 1 genes and two Group 1 cDNA clones and compare these with each other and with two published cDNA sequences (Kuhn *et al.*, 1984). The mRNA corresponding to each of these sequences could in principle be translated to give a MUP protein. Seven out of the eight code for slightly different MUP proteins. The nucleotide sequences are closely related, but the pattern of their relationships is complex, suggesting that gene conversion may have played a part in their origin. We also describe a liver cDNA clone, pMUP15, which belongs neither to Group 1 nor to Group 2 and show that in intron 6 the corresponding mRNA is spliced differently from the Group 1 mRNAs. The pMUP15 sequence differs from the Group 1 and Group 2 sequences by ~15 and 17%, respectively. The distribution pattern of the differences along the sequences suggests that an ancient gene conversion event may have occurred between them.

Results

Comparison of the sequences of different Group 1 MUP genes

In the previous paper (Clark *et al.*, 1985) we described the sequence of a complete Group 1 gene and a complete Group 2 gene. We have also determined the exonic sequences of three other MUP genes (isolated as genomic clones from BALB/c DNA libraries, Clark *et al*, 1982) and of three MUP cDNA clones isolated from a BALB/c liver cDNA library (Figure 1). The four genes had all been classified as belonging to Group 1 by DNA-RNA hybridisation methods. The sequencing data confirm their close similarity, and show that each one codes for a different MUP protein (Tables I and III). Two of the three cDNA clones (pMUP8 and pMUP11) correspond to transcripts of Group 1 genes. Both are incomplete copies of the mRNA, with a deficiency at the 5′ end. pMUP8 is 670 and pMUP11 714 nucleotides long while the length of the long form of MUP mRNA is 879 nucleotides (Clark *et al.*, 1984a). Gene BL1 and pMUP8 are identical over their common length. pMUP8 may therefore correspond to the mRNA transcribed from BL1. The coding sequence of pMUP11 corresponds to a fifth MUP protein. All of the Group 1 sequences are very closely related. The average divergence from the Group 1 consensus sequence is 0.35%. The five Group 1 sequences which are different at the DNA level also differ from each other at the protein level by at least one amino acid.

Table I shows the differences between the regions of seven MUP genes that code for mRNA. Five of these are described above. Two more are cDNA clones isolated and sequenced by Kuhn *et al.* (1984). Where possible, the part of intron 6 that is present in short Group 1 mRNA is included, with intron residue numbers. Variation is observed at a total of 19 different sites, of which 14 are present in long-form mRNA. At 18 sites only two alternative nucleotides are found. At 13 of these sites the minority nucleotide is present in only one of the seven genes, at four sites it is present in two genes and at one site it is present in three. The nucleotide differences listed in Table I affect 10 codons (Table II). Four are silent differences while six are re-

© IRL Press Limited, Oxford, England

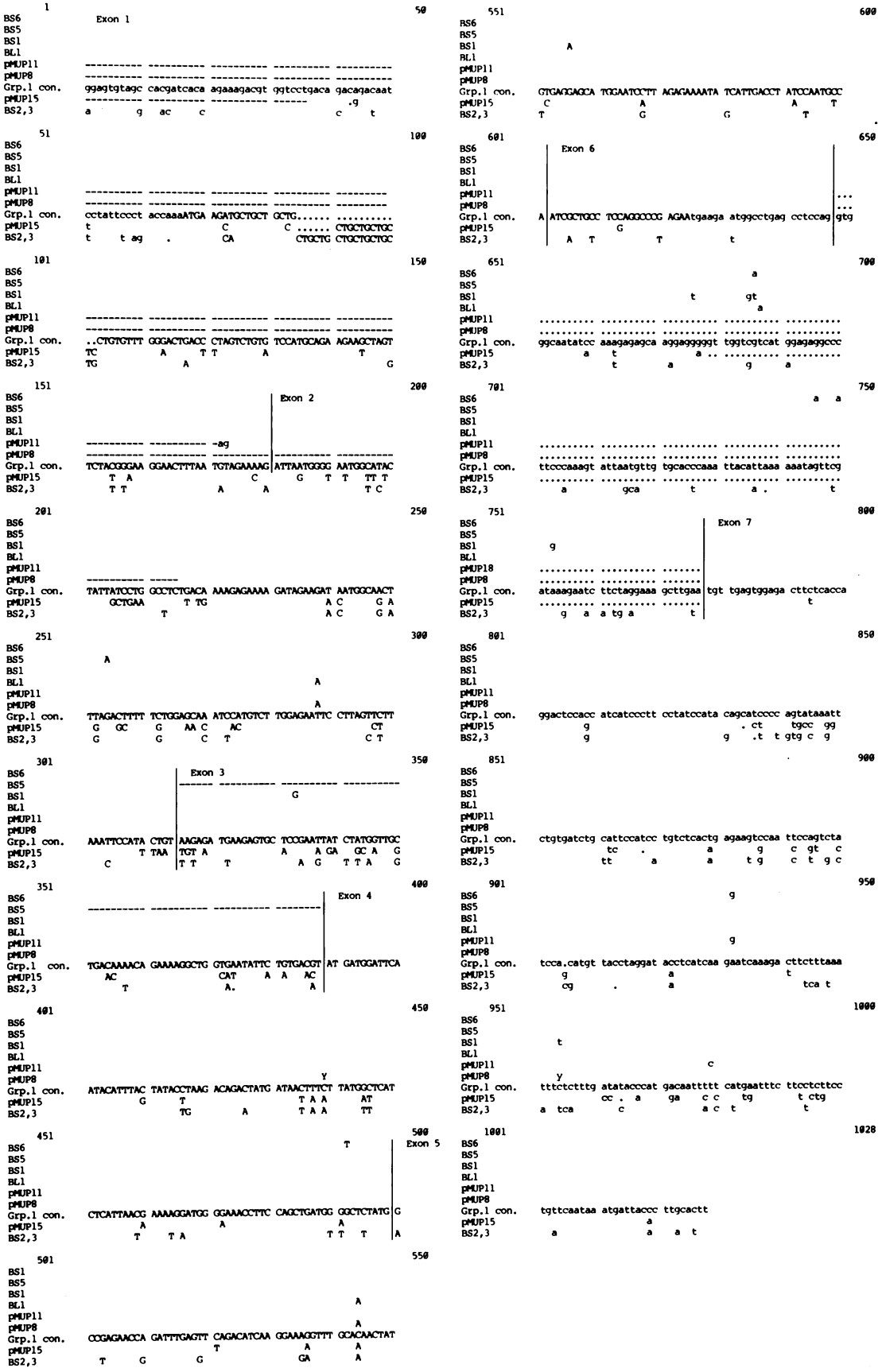

Fig. 1. Sequence comparison of exonic MUP sequences. The exonic sequences of MUP genes cloned in λ phages (BS6, etc.) and MUP cDNA clones isolated from a BALB/c liver cDNA library (pMUP11, etc.) were aligned. The first six of these are Group 1 genes and their consensus sequence (Group 1-Con) is shown immediately underneath and is compard with pMUP15 and BS2,3. The vertical lines delineate the exons.

Table I. Differences between Group 1 MUP genes

	Coding region											3′-non-coding			5′ part of intron 6				
	1	1	1	1	2	2	2	3	4	5	5	7	7	8					1
	0	5	5	9	3	7	8	0	7	2	3	2	7	3	2	3	3	4	0
	8	4	5	2	6	0	6	5	4	6	8	3	5	1	9	8	9	0	6
BS6	C	G	T	G	G	T	T	C	T	C	G	G	C	T	G	C	A	T	A
BS5	C	G	T	G	A	T	T	/	G	C	G	A	C	T	G	C	G	T	A
BS1	C	G	T	G	G	T	T	G	G	C	A	A	T	T	T	G	T	T	G
BL1	C	G	T	G	G	A	T	C	G	A	G	A	C	T	G	C	G	A	A
MUP11	/	A	G	G	G	T	T	C	G	C	G	G	C	C	/	/	/	/	/
1057	G	G	T	A	G	T	G	C	G	C	A	A	T	T	/	/	/	/	/
499	C	G	T	A	G	A	T	C	G	C	G	A	T	T	/	/	/	/	/
				*		*				*		*	*						

The numbers at the top are the residue numbers of the long form of Group 1 mRNA (coding region and 3′-non-coding region), or residue numbers within intron 6 (5′ part of intron 6). 1057 and 499 are from Kuhn *et al.* (1984). / signifies not known. The positions marked * are those at which a single minority nucleotide is present in more than one sequence.

Table II. Codon and amino acid changes at the 10 positions at which Group 1 MUP genes are known to vary

Nucleotide	Amino acid	Consensus codon		Rarer codon		
108	−5	CTG	Leu	CTA	Leu	Silent
154/5	11	GTA	Val	AGA	Arg	
192	24	GTC	Val	GTG	Val	Silent
236	39	AGA	Arg	AAA	Lys	
270	50	AAT	Asn	AAA	Lys	
286	56	TTC	Phe	GTC	Val	
305	62	TCC	Ser	TCG	Ser	Silent
474	118	GGG	Gly	GGT	Gly	Silent
526	136	CAA	Gln	AAA	Lys	
538	140	GAG	Glu	AAG	Lys	

placement differences. Of these, five could be expected to produce a significant difference in the charge of the protein.

Clone pMUP15

The third MUP cDNA clone, pMUP15, is longer than the two Group 1 cDNA clones. It extends 30 bp into the 5′-untranslated region of exon 1 while terminating at the same 3′ position. It also contains two insertions relative to the Group 1 sequences (see below). At the nucleotide level pMUP15 is considerably diverged from both the Group 1 consensus sequence (14.6%) and from the Group 2 gene BS2,3 (17.4%) (Table IV). It has an open reading frame 186 amino acids long which begins at the ATG homologous to the Group 1 and Group 2 genes and terminates at the homologous TGA stop codon. The signal peptide region is 66 nucleotides long, 12 nucleotides [four CTG (Leu) codons] longer than that of the Group 1 genes. The region corresponding to the mature protein is the same length as in the Group 1 genes. The divergence of pMUP15 and the Group 1 consensus is reflected at the protein level. The two sequences differ by 56/182 amino acids (30.7%).

The sequences of both pMUP8 and pMUP11 show that corresponding RNA transcripts were spliced to yield a 45-bp exon 6. The same pattern of splicing is seen in three other Group 1 MUP cDNA clones (Clark *et al.*, 1984a; Kuhn *et al.*, 1984). S1 nuclease protection experiments and Northern blot analyses

Table III. Amino acid patterns at the six known variant sites of Group 1 MUP genes

	1	1				
	1	3	5	5	3	4
	1	9	0	6	6	0
BS1 }	V	R	N	F	Q	K
1057 }	V	R	N	V	Q	K
BL1	V	R	K	F	K	E
MUP11	R	R	N	F	Q	E
499	V	R	K	F	Q	E
BS6	V	R	N	F	Q	R
BS5	V	K	N	/	Q	E

The patterns are arranged in order of charge, with those likely to produce a more positive protein at the top. The numbers at the top are the residue numbers of mature Group 1 MUP protein. Amino acid 56 of BS5 is not known.

Table IV. Divergence between 5′ and 3′ regions of various MUP genes

Comparison	% Divergence		
	Total sequence	Exons 1−3	Exons 4−7
Group 1-Con. versus pMUP15	14.6	20.4	11.1
BS2,3 versus pMUP15	17.4	22.2	14.4
Group 1-Con. versus BS2,3	13.0	12.2	13.5
Consensus versus Group 1-Con.	4.5	4.9	4.3
Consensus versus pMUP15	8.9	15.2	5.2
Consensus versus BS2,3	7.2	6.8	7.7

Group 1-Con. refers to the consensus sequence of the six Group 1 genes sequenced. Consensus refers to the consensus sequence of the comparison between Group 1-Con., MUP15 and BS2,3.

confirm that this is the prevalent mode of splicing of the Group 1 MUP gene transcripts (Clark *et al.*, 1984a). In contrast the transcript of pMUP15 has been spliced to yield a 76-bp exon 6. In principle these different splicing patterns could be due to differences in the region of the Group 1 donor site, in the MUP15 donor site (within intron 6 of the Group 1 gene), or in the common acceptor site, or it could be due to some combination of such differences (Busslinger *et al.*, 1981; Fukumaki *et al.*, 1982; Felber *et al.*, 1982; Treisman *et al.*, 1982, 1983). The gene corresponding to MUP15 has not yet been isolated, so that a complete comparison of the introns is not possible. However, examination of the homologous sequences in the Group 1 genes shows that the splice point that generated pMUP15 occurred 31 bp downstream to that which generated the two Group 1 cDNA clones. In the Group 1 consensus sequence (Figure 1) the sequence immediately 3′ to this point is GTTGGT, which is quite close to the donor site consensus GTARGY. Presumably this sequence, or a closely related counterpart, exists in the MUP15 gene and acts as a functional donor site. However, since there is no divergence between pMUP15 and the Group 1 consensus sequence in the region of the donor site which generates the shorter (45 bp) exon 6, the precursor of pMUP15 mRNA may itself be spliced in two configurations.

Minor MUP proteins are glycosylated

Szoka and Paigen (1978) failed to detect glycosylation of MUP proteins, and indeed the known sequences of the Group 1 genes, which probably code for the bulk of MUP, do not contain an N-linked glycosylation site. However, Kuhn *et al.* (1984) found a potential glycosylation site in the sequence of p199, a cDNA clone that is less complete than MUP15, but identical to it over

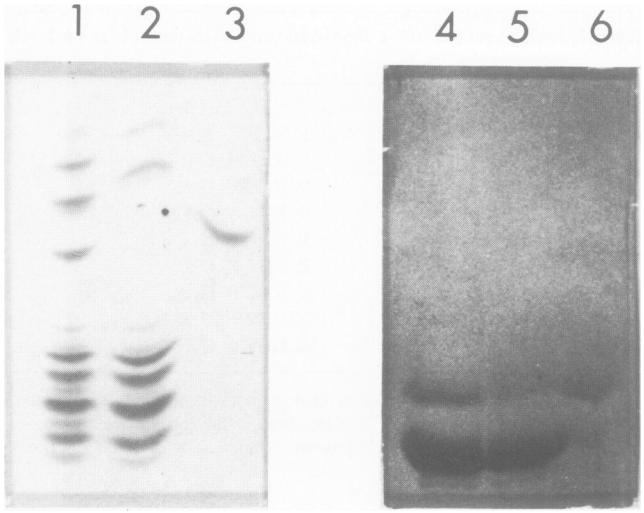

Fig. 2. Glycosylation of a minor MUP component. Purified MUP was fractionated on a column of Con-A-Sepharose. The fractions were resolved by IEF (1–3) and SDS-PAGE (4–6) and stained with Coomassie blue. **Tracks 1** and **4**, fractionated; **2** and **5**, unbound, **3** and **6**, bound to Con-A-Sepharose.

their common length. By hybrid-selection followed by cell-free translation of the hybridised mRNA in the presence of the dog pancreas membrane fraction to effect post-translational modification, Kuhn *et al.* (1984) showed that p199 is homologous in sequence to mRNA that codes for a group of four proteins that migrate more slowly through an SDS gel than most of the MUPs and are also more negative in charge. That is, they show the characteristics expected of glycosylated MUPs.

Our interest in MUP15 led us to ask directly whether mouse urine contains glycosylated MUP proteins, and whether these are the proteins identified by Kuhn *et al.* (1984) as coded for by a gene or a small gene family which is homologous to p199 = MUP15. The MUP proteins were isolated from the urine of both BALB/c and C57BL mice and fractionated on Con-A-Sepharose columns. Both the fractions and the unfractionated MUP proteins were resolved on SDS gels and also separately by isoelectric focusing, and the gels were stained with a general protein dye. Urine from both strains contained a minor fraction of MUP protein that binds specifically to Con-A-Sepharose, and which migrates more slowly on SDS gels and is more negatively charged than the bulk of the MUP protein (Figure 2, only BALB/c shown). Only the band that was retained by the Con-A-Sepharose was stained with a thymol-H_2SO_4 stain specific for glycosylated proteins (data not shown).

Discussion

Group 1 genes are active

All the available evidence indicates that BS6 is a true gene in that it appears to have all the DNA sequences necessary for correct transcription, processing and translation. The BS6 promoter region is active in fibroblasts in association with the SV40 enhancer (P.Ghazal and J.O.Bishop, unpublished experiments). The translation of BS6 and three other Group 1 genes, BS1, BS5 and BL1, would yield acidic proteins of the size of MUP. A number of lines of evidence show that Group 1 genes are the most abundantly expressed in both BABL/c and C57 livers. (i) 6/8 independently isolated cDNA clones correspond very closely (<0.5% divergence) to the Group 1 consensus sequence (Clark *et al.*, 1984a; Kuhn *et al.*, 1984; and this paper). (ii) Under stringent washing conditions Group 1 genes hybridise considerably more strongly to end-labelled liver mRNA than do other isolated MUP genes (Clark *et al.*, 1982). (iii) From MUP mRNA hybrid-selection-translation experiments and signal intensities in Northern blot hybridisations, Kuhn *et al.* (1984) estimated that the ratio of Group 1 sequences to p199-type (MUP15-type) sequences in total C57BL liver mRNA is ~10. Similarly, Clark *et al.* (1984a) found the ratio of Group 1 to Group 2 mRNA in BALB/c liver mRNA to be ~10.

About half of the 35 MUP genes are pseudogenes (Ghazal *et al.*, 1985). Of the active genes, one or two are identical or very similar in sequence to MUP15 (Kuhn *et al.*, 1984). Thus out of the total of ~18 active genes we have extensive data on the coding sequences of eight, including about half of the Group 1 genes.

Group 1 genes may be related through gene conversion

The seven Group 1 sequences contain a non-random distribution of differences at 19 sites. At 13 sites one sequence differs from all the others. We have argued that the 45-kb units that contain the Group 1 genes are closely related to a common ancestor (Clark *et al.*, 1984b; Ghazal *et al.*, 1985). If so, these single-incidence differences may be explained as relatively recent mutation events. The three different bases present at another site are also consistent with independent mutational changes. At the remaining five sites two different nucleotides are each present in more than one gene (see Table I). The distribution of changes at these five sites may be used in an attempt to establish a pedigree of relationships between the genes. The inter-relationship of the entire set can be explained only by supposing that multiple mutations to the same base have occurred at several of the five sites. However, there is no evident reason why this should have occurred. Two of the sites are in the 3′ non-coding region, and the others correspond to Leu→Leu, Asn→Lys and Glu→Lys, respectively. Furthermore, no simple series of recombination events between any two postulated ancestral combinations of the two variants at each site can explain the observed sequences. The most likely explanation of these relationships would seem to involve gene conversion superimposed on gene duplication.

Group 1 MUP genes code for different protein products

Five of the six Group 1 sequences described here code for proteins that differ from one another by at least one amino acid. Two of the sequences described by Knopf *et al.* (1983) code for two additional Group 1 proteins. The greatest pair-wise difference is three amino acids. The pattern of differences present in the seven protein variants is shown in Table III, where the proteins are listed in order of probable net charge. It is likely that these proteins could be resolved by sensitive methods into at least five components. In fact the two-dimensional gels of Knopf *et al.* (1983) and Kuhn *et al.* (1984) show that the main size class of MUP protein resolves into five charge components.

pMUP15 may have been generated by an ancient gene conversion event

In contrast to the differences between the Group 1 consensus and BS2,3, those between pMUP15 and the Group 1 consensus and those between pMUP15 and BS2,3 are distributed non-uniformly along the sequence. In the latter comparisons the first three exons are considerably more diverged than the last four (Table IV). One possible explanation for this is that the 3′ regions of MUP genes are under stronger selective constraints than the 5′ regions and are not as free to diverge. If this were the case, the ratio

of replacemnet site to silent site differences would be greater in exons 1−3 than in exons 4−6. The corrected frequencies of replacement site and silent site differences (Perler *et al.*, 1980) are 14.4% and 11.6%, respectively, a ratio of 1:2. The frequencies over exons 1−3 are 27.5% and 22.4%, a ratio of 1.2 and over exons 4−6 they are 9.8% and 8.9%, a ratio of 1.1. Thus the disproportionate conservation of exons 4−6 of pMUP15 and the Group 1 consensus seems not to be due to selection. Another possible explanation for the non-uniform divergence between pMUP15 and other MUP genes is gene conversion. The fact that pMUP15 is equally divergent from the Group 1 consensus and BS2,3 suggests that the 3′ region of a gene ancestral to both the Group 1 and Group 2 genes may have converted the homologous regions of MUP15 (or *vice versa*) at about the same time as the onset of divergence of the Group 1 and Group 2 genes. Alternatively, the 5′ regions of MUP15 or the Group 1/Group 2 ancestor may have been converted by a more distantly related MUP gene. The two different regions of divergence define a notional junction between exons 1−3 and exons 4−7, with twice as much divergence to the 5′ side as to the 3′ side of the junction (Table IV). Genomic clones corresponding to pMUP15 are not yet available so that it is not possible to determine precisely the positions of the boundaries of the proposed converted region. However, given that the difference in degree of divergence occurs between exons 3 and 4, a boundary must occur in intron 3. In both BS6 and BS2,3 this intron contains a long stretch of the repeated dinucleotide GT (Clark *et al.*, 1985), which has been found at the boundaries of putative regions of gene conversion (Proudfoot and Maniatis, 1980; Shen *et al.*, 1981).

Materials and methods

Cloned DNA

The isolation of MUP genomic clones and subclones is described in Clark *et al.* (1982, 1984b) and Bishop *et al.* (1982). pMUP11 and pMUP15 were isolated from a cDNA library prepared by using poly(A)$^+$ RNA from female mouse (BALB/c) liver as a template for oligo(dT)-primed DNA synthesis by reverse transcriptase. Single-stranded cDNA >1 kb in size was isolated by polyacrylamide gel electrophoresis and made double-stranded with DNA polymerase I and reverse transcriptase. After treatment with nuclease S1, fragments >1 kb were again isolated by polyacrylamide gel electrophoresis, treated with DNA polymerase I and cloned into *Sma*I-cleaved pUC8. White colonies were isolated into microtitre plates and screened with mouse DNA isolated from the genomic subclones pBS6-5-5 and pBS2-2-2 (Bishop *et al.*, 1982) by the method of Gergen *et al.* (1979). The methods used differed only in minor ways from standard procedures (e.g., Maniatis *et al.*, 1982) and will be published in detail elsewhere (A.Chave-Cox, unpublished data). The propagation of bacteriophage and plasmid clones and the isolation of DNA were carried out as described (Clissold and Bishop, 1982; Clark *et al.*, 1982; Bishop *et al.*, 1982).

DNA sequencing

The exonic sequences of BS1, BS5 and BL1 were obtained by sequencing from nearby restriction sites using subclones generated in M13mp8 and M13mp9 and M13tg130 and M13tg131 (Kieny *et al.*, 1984). The cDNA inserts were excised from pUC8 at the polylinker, recloned into M13mp8 and sequenced by the progressive method of Hong (1982).

Fractionation of MUP with Con-A-Sepharose

Urine was collected from 8- to 10-week-old mice by bladder massage, dialysed overnight against distilled water, and fractionated on a column of Sephadex G-100 developed with 0.2 M NH_4HCO_3. Fractions containing MUP were lyophilised and dissolved in 50 mM Tris-HCl, 500 mM NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 6.0. 2 ml (20 mg) of protein was passed over a 7 ml column of Con-A-Sepharose-4B (flow-rate 3.6 ml/h) which was thoroughly washed with the same buffer, and the bound fraction was eluted with the same, containing 0.17 M Na_2-tetraborate. The agarose IEF gel (170 mm running length × 110 × 3 mm) contained 0.74% agarose, 8.75% sorbitol, 3.1% pH 4−6 ampholines and 0.8% pH 3−10 ampholines (both Pharmacia). The electrode solutions were 0.5 M H_2SO_4 and 1 N NaOH. Focusing was for 90 min at 1000 V and 30 min at 1500 V. Acrylamide IEF was performed on 5% preformed plates (LKB Pagplate 1804-111) containing 2% pH 4−5 ampholines. The electrode solutions were 1 M H_3PO_4 and 1 M glycine. Focusing was for 3 h at 1400 V. The SDS-PAGE separating gels contained 15% acrylamide, and the spacers 6%, and the gels were run for 6−8 h at 0.12 mA per mm^2 cross-sectional area. Glycosylated proteins were stained by washing polyacrylamide gels twice (2 h each) in 10 volumes of 25% isopropanol, 10% acetic acid, for 2 h in the same solution containing 0.2% thymol, and for 2.5 h at 35°C in 80% H_2SO_4, 20% ethanol.

Acknowledgements

We are grateful to Morag Robertson and Melville Richardson for technical assistance and to the MRC and the Cancer Research Campaign for financial support.

References

Bishop,J.O., Clark,A.J., Clissold,P.M., Hainey,S. and Francke,U. (1982) *EMBO J.*, **1**, 615-620.
Bishop,J.O., Selman,G.G., Hickman,J., Black,L., Saunders,R.D.P. and Clark, A.J. (1985) *Mol. Cell Biol.*, **5**, 1591-1600.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Busslinger,M., Moschonas,N. and Flavell,R.A. (1981) *Cell*, **27**, 289-298.
Clark,A.J., Clissold,P.M. and Bishop,J.O. (1982) *Gene*, **18**, 221-230.
Clark,A.J., Clissold,P.M., Al-Shawi,R., Beattie,P. and Bishop,J.O. (1984a) *EMBO J.*, **3**, 1045-1052.
Clark,A.J., Hickman,J. and Bishop,J.O. (1984b) *EMBO J.*, **3**, 2055-2064.
Clark,A.J., Ghazal,P., Bingham,R., Barrett,D. and Bishop,J.O. (1985) *EMBO J.*, **4**, 3159-3165.
Clissold,P.M. and Bishop,J.O. (1981) *Gene*, **15**, 225-235.
Clissold,P.M. and Bishop,J.O. (1982) *Gene*, **18**, 211-220.
Clissold,P.M., Hainey,S. and Bishop,J.O. (1984) *Biochem. Genet.*, **22**, 379-387.
Felber,B.K., Orkin,S.H. and Hamer,D.H. (1982) *Cell*, **29**, 895-902.
Fukumaki,Y., Ghosh,P.K., Benz,E.J., Jr., Reddy,V.B., Lebowitz,P., Forget,B. and Weissman,S. (1982) *Cell*, **28**, 585-593.
Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.
Ghazal,P., Clark,A.J. and Bishop,J.O. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4182-4185.
Hastie,N.D. and Held,W.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 414-417.
Hastie,N.D., Held,W.A. and Toole,J.J. (1979) *Cell*, **17**, 449-457.
Hong,G.F. (1982) *J. Mol. Biol.*, **158**, 539-549.
Knopf,J.L., Gallagher,J.A. and Held,W.A. (1983) *Mol. Cell. Biol.*, **3**, 2232-2240.
Kuhn,N.J., Woodworth-Gutai,M., Gross,K.W. and Held,W.A. (1984) *Nucleic Acids Res.*, **12**, 6073-6090.
Maniatias,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Perler,F., Efstratiadis,A., Lomedico,P., Gilbert,W., Kolodner,R. and Dodgson, J. (1980) *Cell*, **20**, 555-565.
Proudfoot,N.J. and Maniatis,T. (1980) *Cell*, **21**, 537-544.
Shaw,P.H., Held,W.A. and Hastie,N.D. (1983) *Cell*, **32**, 755-761.
Shen,S., Slightom,J.L. and Smithies,O. (1981) *Cell*, **26**, 191-203.
Szoka,P.R. and Paigen,K. (1978) *Genetics*, **90**, 597-612.
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.

Received on 27 June 1985; revised on 16 September 1985