The EMBO Journal vol.6 no.3 pp.567-575, 1987

Differentially expressed bovine cytokeratin genes. Analysis of gene linkage and evolutionary conservation of 5'-upstream sequences

Manfred Blessing[1,2], Hanswalter Zentgraf[3] and José L.Jorcano[1,2]

[1]Division of Membrane Biology and Biochemistry, Institute of Cell and Tumor Biology, German Cancer Research Center, [2]Center of Molecular Biology, University of Heidelberg, and [3]Institute of Virus Research, German Cancer Research Center, D-6900 Heidelberg, FRG

Communicated by Werner W.Franke

Cytokeratins are a family of ~20 polypeptides which form the intermediate-sized filaments (IFs) characteristic of epithelial cells. They are synthesized co-ordinately as 'pairs' consisting of one representative from each of the two cytokeratin subfamilies, i.e. the acidic (type I) and the more basic (type II) polypeptides, in cell type-specific combinations. We have isolated and characterized the genes coding for four bovine cytokeratins of the basic (type II) subfamily, i.e. cytokeratins Ib, III, IV and 6*, by Southern blot hybridization, hybridization-selection-translation experiments, heteroduplex mapping, and partial sequencing of the exons coding for the hypervariable carboxy-terminal 'tail' regions of the proteins and the 3'-non-translated ends of the mRNAs which are distinct for the inividual cytokeratin polypeptides. Limited 'chromosomal walk' experiments demonstrated that the genes are organized into two tandems, i.e. 6*→Ib and III→IV, in which they are separated by ~11 kb. RNA analysis by Northern and dot blots show that both genes of the III→IV tandem are co-expressed in some bovine tissues (muzzle epidermis, hoof pad and tongue mucosa) and cultured cells (BMGE+H) but that in other tissues, cornea for example, only the gene encoding III is expressed. Unexpectedly, the genes linked in the tandem 6*→Ib are not co-expressed in any of the tissues examined. mRNA from gene 6* has been found in tongue mucosa but in none of the other cell lines and tissues examined, whereas mRNA for cytokeratin Ib is expressed in cornea and muzzle epidermis but not in, for example, tongue mucosa and in the epidermis of the heel pad. The latter observation indicates that the genes of the same tandem can be separately expressed not only in relation to tissue differentiation but that in addition body site-specific regulatory mechanisms of cytokeratin gene expression exist, at least for epidermis. Sequence analysis reveals that the 5'-upstream regions of the genes coding for individual cytokeratins are remarkably conserved between different species (bovine, murine, human), indicating that they might play a role in the regulation of these genes. However, these regions do not exhibit extensive sequence homology when the various, differentially expressed bovine genes are compared, except for a symmetrical motif AAPuCCAAA located in a defined region upstream of the TATA box in all genes sequenced so far coding for bovine, human and murine epidermal cytokeratins as well as in the gene coding for involucrin, another epidermally expressed protein but not in the other cytokeratin and IF protein genes so far known. We discuss the linkage of this multigene family and the possible relevance of the 5'-flanking sequences to the regulation of its spatial and temporal differential expression.

Key words: cytokeratin genes/intermediate filament proteins/multigene families/clustered genes/regulation of gene expression

Introduction

The proteins of the cytoplasmic intermediate-sized filaments (IFs; 8-12 nm diameter) represent, together with the proteins of the nuclear lamina, a large and complex multigene family whose members are differentially expressed according to different pathways of cell differentiation. The principles of the structural organization of these proteins are largely known (for references see Geisler and Weber, 1982; Fuchs and Hanukoglu, 1983; Weber and Geisler, 1984; Steinert *et al.*, 1985; McKeon *et al.*, 1986). The genes coding for IF proteins also share a similar intron-exon pattern, which suggests an evolution from a common ancestral gene (Quax *et al.*, 1983, 1985; Lehnert *et al.*, 1984; Marchuk *et al.*, 1984; Balcarek and Cowan, 1985; Johnston *et al.*, 1985; Krieg *et al.*, 1985; Rieger *et al.*, 1985; Tyner *et al.*, 1985; Bader *et al.*, 1986; for the exceptional case of a neurofilament protein see Lewis and Cowan, 1986).

The expression of the members of this multigene family during embryonic development and cell differentiation is regulated in such a way that different IF proteins are synthesized in different cell types. For example, cytokeratins are found in epithelia, neurofilament proteins primarily in neurons, glial filament protein in astrocytes, desmin in various myogenic cells, and vimentin in cells of mesenchymal origin and certain other cell types (for reviews see Franke *et al.*, 1982; Lazarides, 1982; Osborn and Weber, 1983).

The epithelial cytokeratins constitute a family of ~20 polypeptides (e.g. 19 polypeptides have been distinguished in human tissues; cf. Moll *et al.*, 1982). On the basis of biochemical, immunological and nucleic acid data, they have been divided into two subfamilies: the basic (or type II) and the acidic (or type I) cytokeratins (Fuchs *et al.*, 1981; Schiller *et al.*, 1982; Crewther *et al.*, 1983; Hanukoglu and Fuchs, 1983; Sun *et al.*, 1984). Heterotypic complexes of basic and acidic cytokeratins, which appear as tetramers containing two molecules of each type, are the structural subunits of cytokeratin IFs (Franke *et al.*, 1983; Quinlan *et al.*, 1984; Woods and Inglis, 1984; Parry *et al.*, 1985; Fraser *et al.*, 1986). 'Expression pairs' of certain basic and acidic cytokeratin polypeptides have been identified in the various types of epithelial cells, and the different epithelia are characterized by the combinations of cytokeratin pairs which they synthesize (Moll *et al.*, 1982; Sun *et al.*, 1984; Cooper *et al.*, 1985; Quinlan *et al.*, 1985). In addition, a different set of eight (four each of the type I and the type II subfamily) 'hard' α-keratins have been identified which are specifically expressed, as products of distinct mRNAs, in the hair-forming cells ('trichocytes') of follicles of human and bovine hairs as well as sheep wool (Crewther *et al.*, 1980; Rogers, 1985; Heid *et al.*, 1986).

© IRL Press Limited, Oxford, England

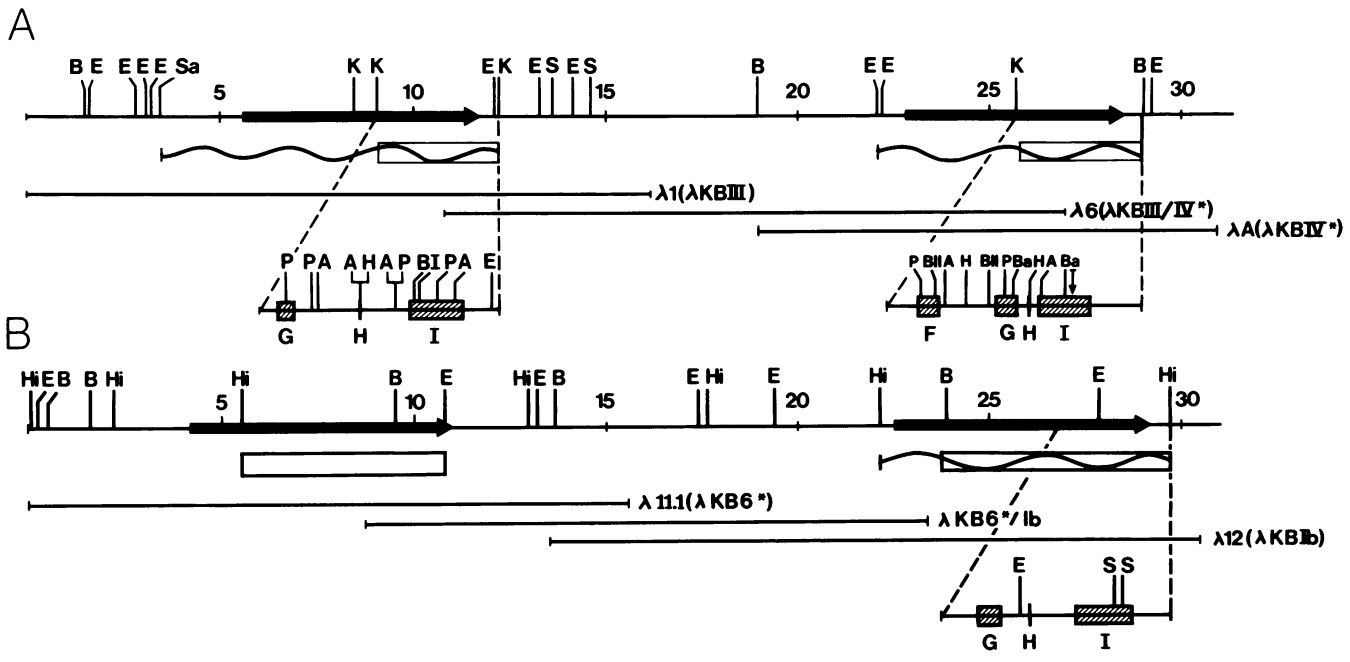

Fig. 1. Maps of the loci containing the genes coding for cytokeratins III and IV (**A**) and 6* and Ib (**B**). The position and direction of transcription of the genes is indicated by the horizontal arrows. The horizontal lines followed by the names of the clones represent the inserts of the recombinant phages used to establish these maps. The open rectangles indicate restriction fragments hybridizing with the corresponding cDNA clones (see text) and the wavy lines are regions hybridizing with calf muzzle epidermis mRNA. The expanded restriction fragments have been sequenced. The filled rectangles denote the positions of exons F, G, H and I coding for the carboxy end of the polypeptide and the 3'-untranslated part of the mRNA. The restriction enzymes used were: *Ava*II (A), *Bam*HI (B), *Bgl*I (BI), *Bgl*II (BII), *Ban*II (Ba), *Eco*RI (E), *Hinc*II (H), *Hin*dIII (Hi), *Kpn*I (K), *Pst*I (P), *Sma*I (S) and *Sac*II (Sa). Lengths are given in kb.

Several genes coding for individual cytokeratin polypeptides have been cloned with a view to understanding the mechanisms governing the control of their cell type-specific expression (Lehnert *et al.*, 1984; Marchuk *et al.*, 1984; Johnson *et al.*, 1985; Krieg *et al.*, 1985; Rieger *et al.*, 1985; Tyner *et al.*, 1985; Bader *et al.*, 1986). We have specifically examined whether the cytokeratin genes are clustered and arranged in patterns that might give a clue to the mechanisms that co-ordinate the expression of groups of genes in a cell type-specific manner.

Here we show that several bovine genes encoding cytokeratins of the basic subfamily are closely linked in tandems, and we discuss the significance of this topological arrangement as well as the importance of the 5'-flanking sequences for the differential expression of the cytokeratin genes.

Results

Linkage of the genes encoding cytokeratins III and IV

We have recently isolated and characterized cDNA clones coding for the major bovine epidermal cytokeratins Ia, Ib, III, IV, VIb and VII (Jorcano *et al.*, 1984a,b,c) and have partially characterized the corresponding cloned genes by means of heteroduplex and S1 nuclease mapping as well as nucleotide sequence analysis (Lehnert *et al.*, 1984; Rieger *et al.*, 1985).

From a bovine genomic library we have isolated six recombinant λ phages which hybridized strongly with the cDNA clone pKBIII coding for cytokeratin III. As expected, DNA isolated from the genomic recombinants reacted strongly in Southern hybridizations with plasmid pKBIII, and the six clones turned out to be overlapping fragments covering a continuous region of ~27 kb (results not shown; for a detailed characterization of these genomic clones see Blessing, 1985). Clones λ1 and λ6 were selected for further studies because they encompassed the whole locus (Figure 1A).

Hybridization with the cDNA clone allowed us to map the position of the 3'-end of the cytokeratin III gene in a 3.1-kb *Kpn*I fragment of clone λ1 (Figure 1A). Rehybridization of the same filters with radioactively labelled bovine muzzle epidermal poly(A)$^+$ RNA localized the 5'-end of the gene in a *Sac*II/*Kpn*I fragment at the left of the 3'-end in clone λ1 and indicated that the right end of clone λ6 also contained sequences complementary to RNA expressed in calf muzzel epidermis (Figure 1A). Since the mRNA isolated from this tissue is particularly enriched in sequences coding for cytokeratins (Kreis *et al.*, 1983; Jorcano *et al.*, 1984a) this result suggested that the cloned locus might contain two closely adjacent cytokeratin genes.

To test this hypothesis, DNA from clones λ1 and λ6 was bound to nitrocellulose filters, hybridized with muzzle epidermis poly(A)$^+$ RNA and the selected RNAs were translated *in vitro*. As shown in Figure 2, clone λ1 selected mRNA coding for cytokeratin III whereas clone λ6 selected mRNAs encoding for cytokeratins III and IV, suggesting that the genes encoding these two cytokeratins might be linked. Therefore, the genomic library was again screened, using as a probe the *Eco*RI/*Kpn*I fragment of λ6 assumed to hybridize with the mRNA encoding cytokeratin IV. The positive clones were ordered according to their restriction maps and clone λA, i.e. the one extending the most into the desired direction, was selected for further studies (Figure 1A).

Restriction map analysis showed (Figure 1A; for a detailed characterization of the clones see Blessing, 1985) that clones λA and λ6 were overlapping. This was corroborated by sequencing 300 bp of the two clones on either side of the common *Kpn*I site, i.e. a region located in an intron (see Figure 1A) as sequences of introns of cytokeratin genes are known to be highly divergent (cf. Johnson *et al.*, 1985; Rieger *et al.*, 1985; Tyner *et al.*, 1985). On Southern blot hybridization analysis, we noted that a *Kpn*I/*Bam*HI fragment of clone λA (3.3 kb), which is only partially present in clone λ6, hybridized with plasmid pKBIV coding

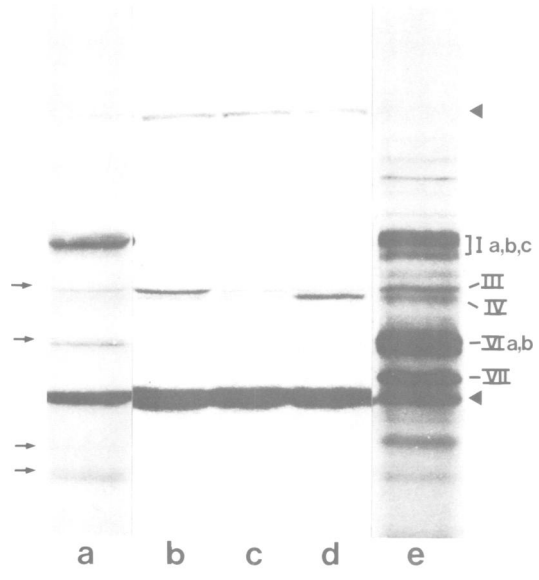

Fig. 2. Identification of genomic clones for bovine epidermal cytokeratins by hybidization−selection−translation using total poly(A)$^+$ RNA from bovine muzzle epidermis. 20 μg of DNA from the different genomic clones were bound to nitrocellulose filters and hybridized with 50 μg of mRNA. The selected RNAs were translated *in vitro* and the synthesized, ^{35}S-methionine labelled polypeptides were analyzed by SDS−polyacrylamide gel electrophoresis. The figure shows a fluorography of the gel. (**a**) Clone λ12; (**b**) clone λl; (**c**) clone λ6; (**d**) clone λA; (**e**) 1 μg of cow muzzle mRNA translated *in vitro* as control. Roman numerals denote major cytokeratin polypeptides (cf. Schiller *et al.*, 1982; Jorcano *et al.*, 1984a). The arrowheads denote endogenous components of the *in vitro* translation system, and the arrows in (a) point to degradation products of cytokeratin Ib (cf. Jorcano *et al.*, 1984a). Lanes b, c and d were exposed four times longer than a and e.

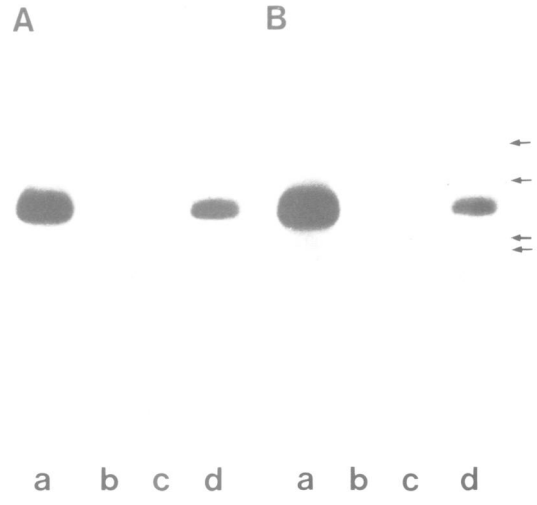

Fig. 3. Northern blot comparison of cDNA clone pKBIV and genomic clone λKBIV*. 20 μg of total cellular RNA isolated from bovine tongue mucosa (**a**), skeletal muscle (**b**), cornea (**c**) and muzzle epidermis (**d**) were glyoxylated, divided into two portions of 10 μg each and electrophoresed on the same 1.2% agarose gel. After blotting to a nitrocellulose filter a half of the filter (**A**) was hybridized with a ^{32}P-labelled anti-sense RNA probe synthesized from the species 3′-end of cDNA clone pkBIV; the other half of the filter (**B**) was hybridized with the same type of probe synthesized from genomic clone λKBIV*. The arrows point to the positions of *Escherichia coli* (16S and 23S) and *Xenopus laevis* (18S and 28S) rRNAs used as markers. Note that the mRNA hybridizing to the two probes has the same length and is expressed in the same tissues.

for keratin IV (Figure 1A). Hybridization with end-labelled muzzle mRNA established the polarity and approximate length of the gene as presented in Figure 1A. Using bovine muzzle epidermal poly(A)$^+$ RNA in hybrid-selection−translation experiments, clone λA selected exclusively mRNA coding for cytokeratin IV (Figure 2).

Although the region of the mRNA coding for the central α-helical rod domain (for terminology see Giesler and Weber, 1982; Weber and Geisler, 1984) of different basic (type II) cytokeratins can be relatively similar in sequence, allowing cross-hybridization at reduced stringency (Hanukoglu and Fuchs, 1983; Kim *et al.*, 1983; Roop *et al.*, 1983; Jorcano *et al.*, 1984b,c), the sequences of the 3′-end region of our cDNA clones, comprising the hypervariable cytokeratin 'tails' and the 3′-untranslated portion of the mRNA, are highly divergent and specific for each cytokeratin mRNA (Jorcano *et al.*, 1984b,c). For this reason, and in order to identify fully the two adjacent genes, the *Kpn*I/*Kpn*I fragment of clone λ1 hybridizing with plasmid pKBIII and the *Kpn*I/*Bam*HI fragment of clone λA hybridizing with plasmid pKBIV were subcloned in pUC18, mapped by restriction analysis, and the fragments hybridizing with the cDNA clones were sequenced (Figure 1A).

In the case of the genomic clone λ1, the sequences of exons G (encoding the end of the α-helical cytokeratin domain), H and I (containing the tail region of the polypeptides and the untranslated part of the mRNA) were identical with the corresponding regions of the cDNA clone pKBIII (results not shown; cf. Blessing, 1985; for intron−exon designations see Lehnert *et al.*, 1984). Thus, clone λ1 contains the gene coding for the bovine cytokeratin III and will be designated as λKBIII in future. In agreement with these data, heteroduplex analysis of this clone has shown (Lehnert *et al.*, 1984) that it contains the gene encoding cytokeratin III plus several kilobases of flanking sequences and that the gene conforms to the pattern of nine exons (A−I) interrupted by eight introns as it is characteristic of all type II epidermal cytokeratin genes so far analyzed (Lehnert *et al.*, 1984; Johnson *et al.*, 1985; Tyner *et al.*, 1985).

In the genomic clone λA, exons F (not contained in the cDNA clone pKBIV) to I were sequenced. Exons G and H and the portion of exon I containing the carboxy end of the polypeptide plus the 70 bp following the stop codon (arrow to exon I, Figure 1A) were identical with the corresponding regions of cDNA clone pKBIV. The rest of the 3′-untranslated region of exon I had only 84% sequence identity with the cDNA. As clone λA codes for a polypeptide sequence identical with the part of cytokeratin IV so far sequenced (Jorcano *et al.*, 1984c), including the specific, hypervariable carboxy end and as the mRNA encoded by this clone is indistinguishable, in size and pattern of expression, from cytokeratin IV mRNA (Figure 3) we designate this clone as λKBIV*. A detailed study of this genomic clone showing the existence of two genes encoding mRNAs for the same polypeptide, i.e. cytokeratin IV, but different in their 3′-untranslated region, will be published elsewhere (M.Blessing and J.L.Jorcano, in preparation).

Heteroduplex analysis (Figure 4) shows that gene λBKIV* also displays the exon−intron pattern common to genes of the basic (type II) cytokeratin subfamily.

*Linkage of genes coding for cytokeratins Ib and 6**

When the genomic library was screened with a cDNA clone encoding cytokeratin Ib (pKBIb; for details see Jorcano *et al.*, 1984a,c) four overlapping positive genomic recombinants were found which were ordered according to their restriction maps. The position of the 3′-end and the length of the gene was map-

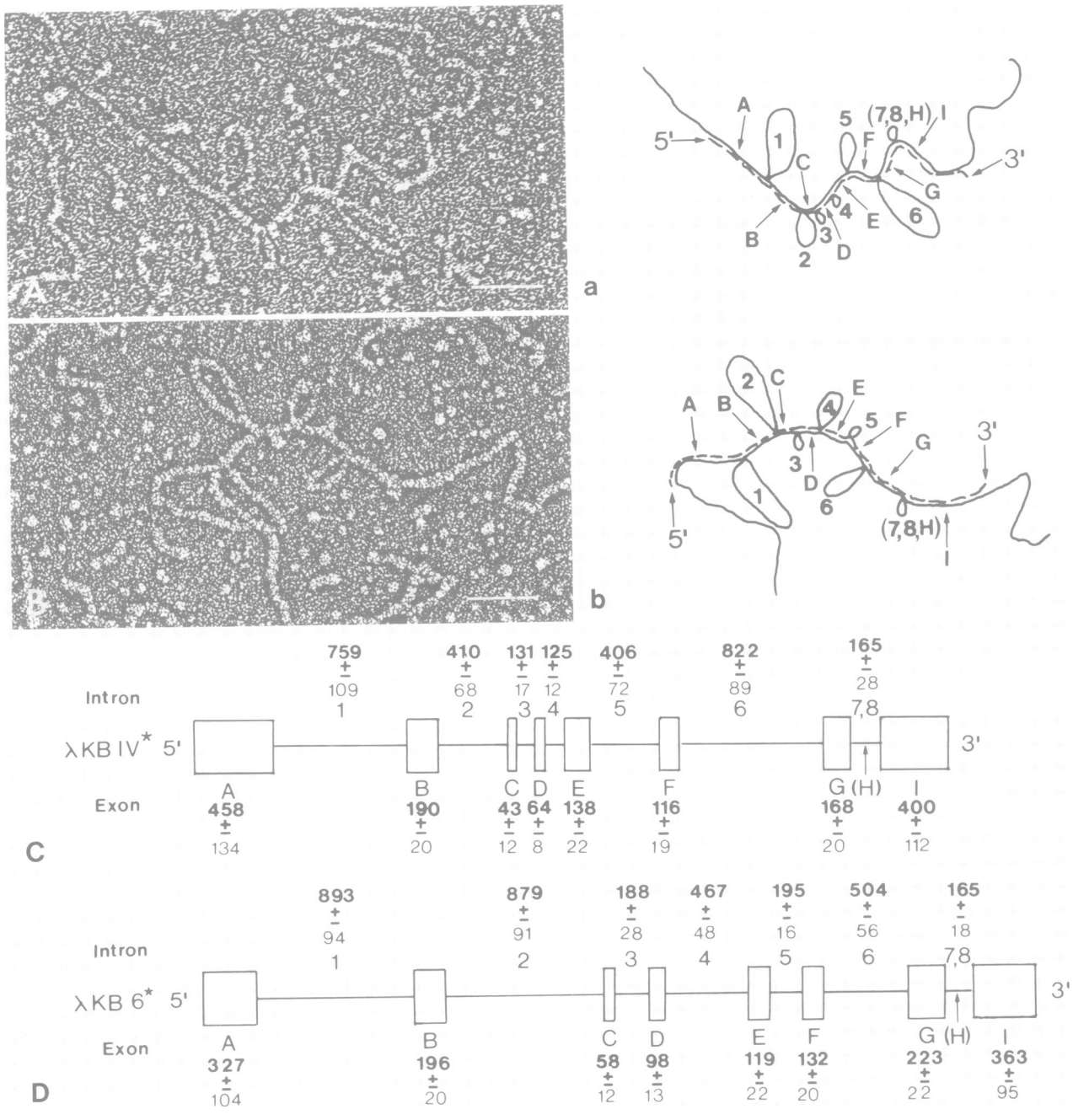

Fig. 4. Characterization of the intron−exon structure of genomic clones by heteroduplex analysis. Electron micrographs showing hybrid molecules between tongue mucosa RNA and DNA from clone λKBIV* (A) and clone λKB6* (**B**). (**a**) and (**b**) are the corresponding interpretative drawings; DNA is represented by a continuous line and RNA by an interrupted line. Exons are represented by capital letters and introns by arabic numerals. (**C**) and (**D**) are schematic representations of the intron−exon structure of the genes as derived from the heteroduplexes. Mean values of lengths (in basepairs) and standard derivations are indicated. The very small (35 bp) exon H cannot be detected by this caim of analysis in these genes due to the stringent hybridization conditions used in this experiment (cf. Lehnert *et al.*, 1984) but its presence has been confirmed in clone λKBIV* by sequence analysis (see Figure 1A).

ped by hybridizing various restriction fragments of the genomic clones with radioactively labeled pKBIb and muzzle epidermal mRNA (Figure 1B). Clone λ12, containing the complete coding region and several kb of flanking sequences on either side, was selected for further studies. In hybrid-selection−translation experiments using cow muzzle epidermal poly(A)$^+$ mRNA, this clone reacted exclusively with mRNA encoding cytokeratin Ib (Figure 2). The identity of clone λ12 was again established by sequencing exons G, H and I, the latter containing the hypervariable — and hence diagnostic — carboxy-terminal portion and the 3′-untranslated region (Figure 1B). Since the sequence of this region was identical with that of the cDNA clone pKBIb, we concluded that clone λ12 contains the gene encoding cytokeratin Ib and designated it λKBIb. We have previously shown that this gene exhibits the pattern of nine exons−eight introns that is characteristic of epidermal type II cytokeratin genes (Lehnert *et al.*, 1984).

To find out whether λKBIb was flanked by other cytokeratin genes, the 4.2-kb *Eco*RI 5′-fragment was subcloned in pUC8 and used for re-screening the genomic library. The five clones obtained were ordered according to their restriction maps, and the clone extending the most into the 5′-direction (designated

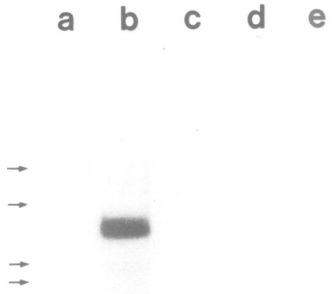

Fig. 5. Characterization of clone λ11.1 by Northern blot analysis. 10 μg of total cellular RNA isolated from BMGE-H cells (**a**), tongue mucosa (**b**), skeletal muscle (**c**), cornea (**d**) and muzzle epidermis (**e**) were glyoxylated, separated on 1.2% agarose gels and transferred to a nitrocellulose filter. The filter was hybridized with a radioactive RNA probe synthesized from the *Eco*RI site of the 5.2-kb *Eco*RI/*Hin*dIII fragment of clone λ11.1 subcloned in the pTZ18R vector (Pharmacia). The filter was exposed for 1 h. The genomic clone hybridized exclusively with tongue mucosa RNA even after longer exposure times. Arrows denote rRNA markers as in Figure 3.

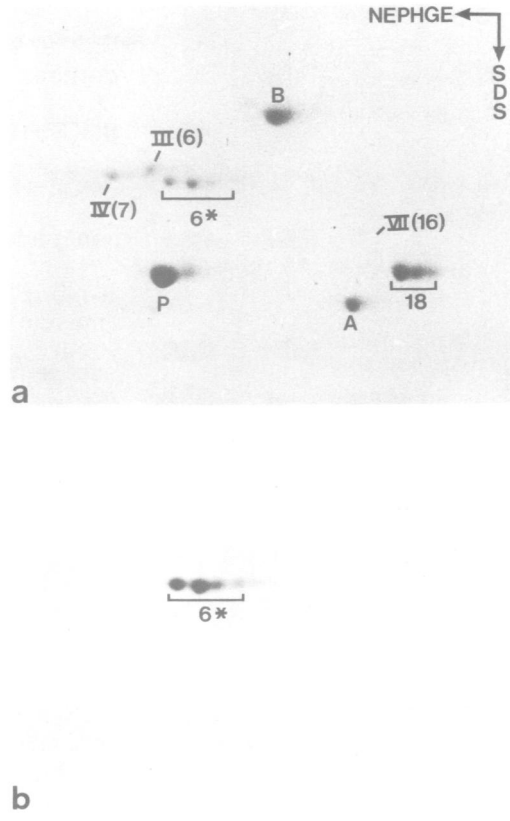

Fig. 6. Identification of clone λ11.1 by two-dimensional gel electrophoresis analysis of hybrid-selection−translation experiments. 10 μg of DNA from clone λ11.1 were bound to a nitrocellulose filter and hybridized with 250 μg of tongue mucosa total RNA. The *in vitro* translation assays obtained with the hybridized RNA were complemented with purified tongue mucosa cytokeratins and subjected to two-dimensional gel electrophoresis (NEPHGE, non-equilibrium pH gradient in the first dimension; SDS, electrophoresis in the presence of SDS used in the second dimension), followed by fluorography. (**a**) Coomassie blue-stained gel showing the tongue mucosa keratins (Roman and arabic numerals; see Schiller, 1985) as well as bovine serum albumin (B), phosphoglycerokinase (P) and α-actin (A) added as markers. (**b**) Fluorogram of the same gel demonstrating that clone λ11.1 codes for cytokeratin 6*.

λ11.1) was selected for further studies (Figure 1B).

In Southern hybridization experiments, the central *Eco*RI/*Hin*dIII fragment of 5.2 kb of λ11.1 (Figure 1B) hybridized, although not strongly, with cDNA clones encoding the basic cytokeratins Ia, Ib, III, IV and 8 (Jorcano *et al.*, 1984c; Magin *et al.*, 1986), but it did not hybridize with the cDNA encoding the acidic cytokeratins VI and VII (Jorcano *et al.*, 1984a,b) and the small cytokeratin of M_r 43 700 (Bader *et al.*, 1986) which is equivalent to human cytokeratin no. 19 (cf. Moll *et al.*, 1982). This result (not shown) indicated that clone λ11.1 carried a gene coding for a basic cytokeratin that was different from all those we had cloned so far.

The *Eco*RI/*Hin*dIII fragment hybridizing with the cDNAs was subcloned in plasmids pTZ18R and pTZ19R, i.e. vectors containing the T7 RNA polymerase promoter in opposite orientations and thus allowing transcription of the DNA insert in the sense or the anti-sense direction. Total RNA extracted from several bovine tissues and cultured cell lines was hybridized, under stringent conditions, with radioactively labelled RNA synthesized from the pTZ plasmids in both orientations. Only in RNA isolated from the mucosa of the tongue did we recognize a band, corresponding to a mRNA of 2.4 kb, that hybridized strongly with the probe synthesized from pTZ18R, i.e. in the *Eco*RI→*Hin*dIII direction (Figure 5). This result established the polarity of the gene as presented in Figure 1B and identified a tissue in which this gene was abundantly expressed.

For the final identification of the gene, DNA from the pTZ18R subclone carrying the genomic *Eco*RI/*Hin*dIII fragment was bound to nitrocellullose filters and allowed to hybridize with total RNA extracted from tongue mucosa. The selected RNA was released, translated *in vitro* and the ^{35}S-labelled polypeptides were co-electrophoresed, in two-dimensional gel electrophoresis, with cytokeratins extracted from the same tissue. The result of this experiment (Figure 6) identifies the gene contained in clone λ11.1 as that coding for the cytokeratin of M_r 59 000 designated 6* by Schiller (1985; see also Schmid *et al.*, 1983; Cooper and Sun, 1986). Therefore, clone λ11.1 has been redesignated λKB6*. Heteroduplex analysis of λKB6* with tongue mucosa RNA (Figure 4) demonstrated that the gene has an intron − exon pattern similar to that of λKBIV* and allowed us to define the ends of the gene which has the same orientation as that encoding cytokeratin Ib, with a separation of ~11 kb (Figure 1B).

Linked cytokeratin genes are co-expressed in some cells but differentially expressed in others

For studies of the expression of these two pairs of linked basic (type II) cytokeratin genes we applied total RNA from bovine skeletal muscle (a non-epithelial tissue), bladder urothelium, several stratified epithelia (cornea, tongue, muzzle epidermis, heel pad epidermis) and two epithelial cell lines derived from mammary gland (BMGE+H and BMGE−H) to nitrocellulose filters and hybridized them with anti-sense RNA probes from the specific 3′-ends of the four genes at high stringency, i.e. conditions allowing differentiation between sequences sharing as much as 85% homology.

Although the two members of the tandem KBIII→KBIV* are co-expressed in the three types of stratified epithelial tissues ex-

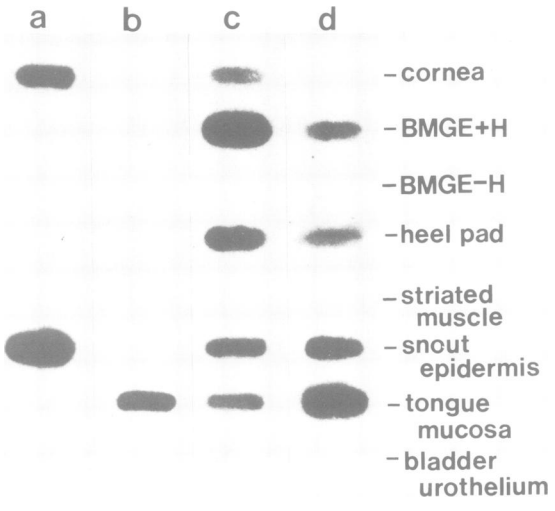

Fig. 7. Expression patterns of linked cytokeratin genes. 10 μg of total cellular RNA extracted from the tissues and cell lines indicated in the figure were bound to nitrocellulose filters and hybridized with radioactively labelled anti-sense RNA probes derived from the genomic clones λKBIb (**a**), λKB6* (**b**), λKBIII (**c**) and λKBIV* (**d**).

amined (tongue mucosa, muzzle and hoof heel pad epidermis) and in BMGE+H cells, another stratified epithelium, cornea, synthesizes only mRNA from KBIII (Figure 7). The members of the pair KBIb→KB6* are also independently regulated. Transcripts encoding KBIb were recognized only in cornea and muzzle epidermis but not in any of the other cell types examined, including heel pad epidermis. In contrast, mRNA encoding cytokeratin 6* was absent from cornea and epidermis but was exclusively found in tongue mucosa where it was present in high concentrations. These results show that there is no obligatory correlation between gene linkage and cell type-specific expression as neighbouring genes are co-expressed in some tissues, are differentially expressed in others or can exhibit totally different patterns of expression.

Comparison of the 5'-flanking regions of basic (type II) cytokeratin genes

The genes encoding the basic (type II) cytokeratins Ia, Ib and IV and those encoding the acidic (type I) cytokeratin VIb are coexpressed in muzzle epidermis (this study; see also Schiller *et al.*, 1982; Jorcano *et al.*, 1984a; Cooper and Sun, 1986). However, when the sequences upstream of the protein coding region of these genes (cf. Rieger *et al.*, 1985; M.Blessing, H.Zentgraf and J.L.Jorcano, unpublished results) were compared they did not reveal conspicuous homology. The only genes showing some degree of homology in this region were those encoding Ia and Ib (Figure 8A) which exhibit 54% sequence identity (69% sequence homology) in the 250-bp region preceding the initiation codon. Besides the TATA box, only two other conserved motifs were detected in these two genes. The first motif is 11 nucleotides long, starting at position 27 (Figure 8A), 112 bp upstream of the TATA box. The second is located at position 66 (Figure 8A), 70 bp upstream of the TATA box and comprises 12 nucleotides. Interestingly, this last homology box includes, in the gene encoding cytokeratin Ia, the symmetrical sequence AAACCAAA which is found, with very few deviations from the consensus form AAPuCCAAA, in the vinicity of the TATA box of all bovine, human and murine epidermal cytokeratin genes of both subfamilies so far sequenced (Table I). A similar box is found upstream of the TATA box in the gene coding for involucrin

Table I. DNA sequence motif found in the 5'-upstream region of epidermal genes

Gene	Sequence	Position (from 'TATA')
KB Ia	A A A C C A A A	−65
KB Ib	A A A C C c A A	−69
KB IV*	A A A C C A A A	−130
KB VIb[a]	A A A C C A A A	−269
	A A A C C A A A	−132
	A A G C C A A A	−81
KH 6[b]	A A G C g A A A	−132
KH 1[c]	A A G C C A A A	−67
KH14[d]	A A G C C c A A A	−146
KM 59K[e]	A A G C C A A g	−76
H involucrin[f]	A A G t C A A A	−410
Consensus	A A R C C A A A	

K, cytokeratin; B, bovine; H, human; M, mouse; R, purine; lower case letters indicate differences from the consensus sequence.
[a]Rieger *et al.*, 1985.
[b]Tyner *et al.*, 1985.
[c]Johnson *et al.*, 1985.
[d]Marchuk *et al.*, 1984.
[e]Krieg *et al.*, 1985.
[f]Eckert and Green, 1986.

(Eckert and Green, 1986; see Table I) which is also expressed in epidermis. This sequence motif is missing, however, in the available sequences of genes coding for non-epidermal cytokeratins such as human cytokeratin 17 (Chaudhury *et al.*, 1986), bovine cytokeratin of M_r 43 700 equivalent to human cytokeratin 19 (Bader *et al.*, 1986) and murine cytokeratin A ('endo A'; Vasseur *et al.*, 1985) which is equivalent to human cytokeratin 8 (cf. Schiller *et al.*, 1982).

On the basis of their high degree of sequence homology and cell type-specific patterns of expression, we have previously concluded that the human cytokeratin no. 6 and the bovine cytokeratin IV are equivalent cytokeratins in the two species, and similarly the murine epidermal cytokeratin of M_r 59 000 is equivalent to bovine cytokeratin VIb (Jorcano *et al.*, 1984b,c). Figure 8B shows that this close relationship is not restricted to the coding portions of the gene but extends to the 5'-flanking regions. In the genes encoding human cytokeratin no. 6 and bovine cytokeratin IV the 350 nucleotides preceding the start codon exhibit 84% sequence homology (72% identity) and, most strikingly, there is a stretch of 90 nucleotides (positions 185−275 in Figure 8B) ending at the TATA box that has remained almost unchanged between the two species. A similar degree of homology (81% sequence identity, 88% homology; Figure 8C) is seen in the 175 nucleotides preceding the start codon of the iso-types bovine cytokeratin VIb and mouse cytokeratin of M_r 59 000.

Discussion

Our results show that several bovine genes coding for basic cytokeratins are closely linked and arranged in the same orientation. The identification of the genes is based on hybrid-selection−translation experiments performed under stringent conditions and, where possible (in three of the four genes studied), by comparison of the sequence of the 3'-ends of the genes with the corresponding cDNAs. As the regions coding for the non-α-helical 'tail' of the protein and the 3'-untranslated portions of the mRNAs (exons H and I) are known to be highly divergent and therefore characteristic of each cytokeratin our criteria permit an unequivocal identification of the genes (for bovine type

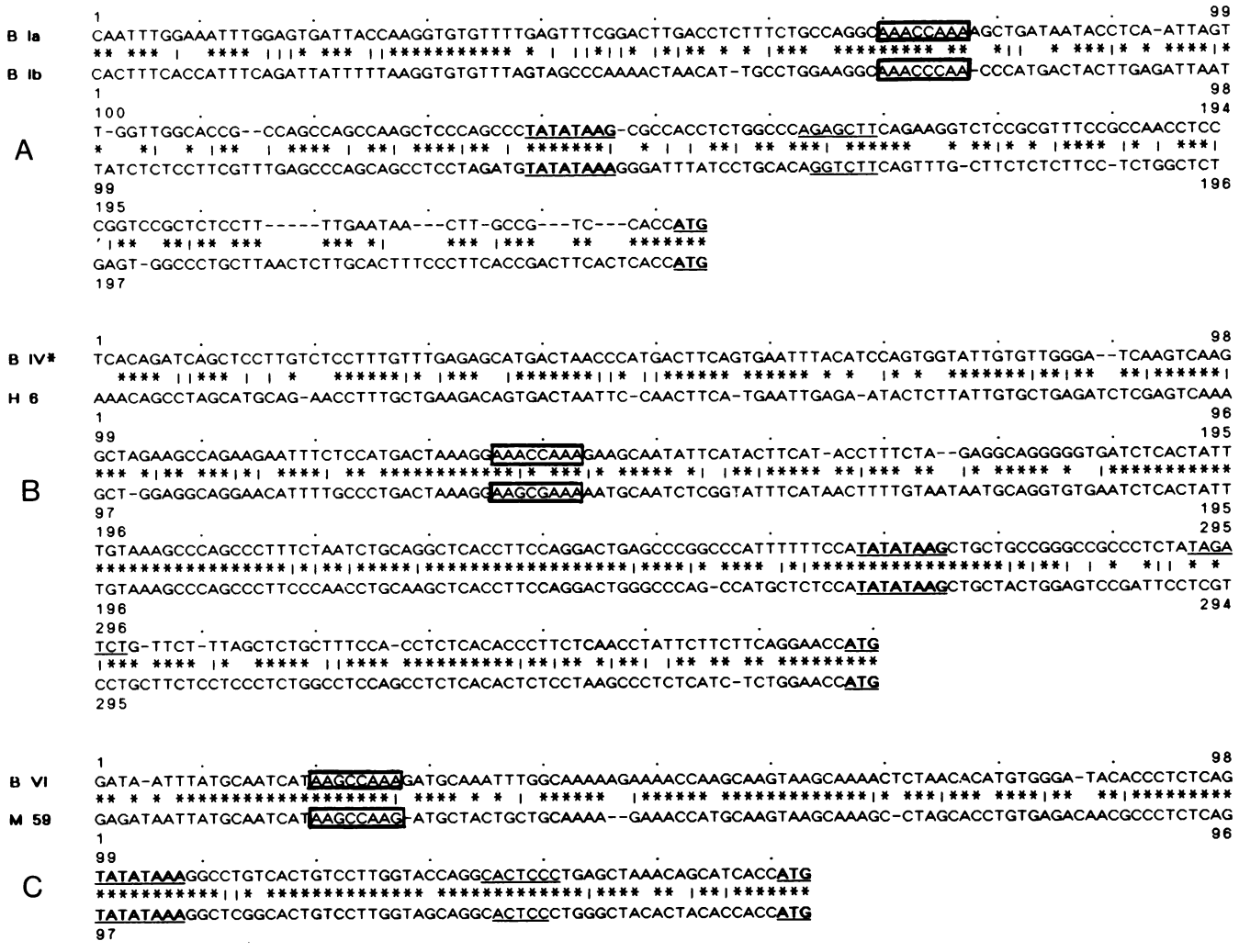

Fig. 8. Comparision of the nucleotide sequence of the 5'-untranslated and flanking regions of the genes coding for different bovine epidermal cytokeratins (**A**) or coding for the same cytokeratin in two different species (**B,C**) showing in this last case the evolutionary conservation of these regions. (**A**) Bovine cytokeratin Ia (B Ia) versus bovine cytokeratin Ib (B Ib). (**B**) Bovine epidermal cytokeratin IV (B IV*) versus human epidermal cytokeratin no. 6 (H 6; Tyner *et al.*, 1985). (**C**) Bovine epidermal cytokeratin VI (B VI; Rieger *et al.*, 1985) versus mouse epidermal cytokeratin of M_r 59 000 (M 59; Krieg *et al.*, 1985). Underlined are the ATG start codons, the cap sites as determined by SI nuclease mapping (results not shown) and the TATA boxes. Asterisks denote identical nucleotides and vertical bars purine → purine and pyrimidine → pyrimidine transitions. The boxes contain an octanucleotide AAPuCCAAA found upstream of the TATA box in all epidermal cytokeratin genes.

II genes see Jorcano *et al.*, 1984c; Magin *et al.*, 1986; for human genes see Glass *et al.*, 1985; Johnson *et al.*, 1985; Tyner *et al.*, 1985; Leube *et al.*, 1986). We are currently extending our chromosomal walks in order to determine the numbers and linkage forms of clustered cytokeratin genes. The results shown in Figure 1, together with other preliminary data, suggest that many of the genes encoding type II cytokeratins are close together but probably separated from those encoding type I cytokeratins. This agrees with recent results of Chaudhury *et al.* (1986) who found three type I cytokeratin genes (one coding for cytokeratin 17 and two others that are yet unidentified) closely linked in the human genome. Recent data (Powell *et al.*, 1986) also indicate the separated linkage of several genes coding for these subgroups of cytokeratins in the sheep genome. This clustering of different subfamilies into separate loci is also found in other multigene families such as the α- and β-globins (Proudfoot *et al.*, 1980) and the light and heavy chains of immunoglobulins (e.g. Davis *et al.*, 1980; Tonegawa, 1983) which, as with cytokeratins, code for two types of polypeptides whose stoichiometric interaction is required to produce the functional protein. Such clustered arrangements probably reflect the evolutionary history of these gene families which may be derived by an early duplication of a single precursor gene, followed by the insertion of one of the copies at a new chromosomal locus.

In contrast to other multigene families, the cytokeratin genes do not seem to be arranged in an order related to their expression during development and differentiation. So, the gene coding for cytokeratin Ib, which is expressed at the protein and mRNA level in the epidermis of the cow muzzle and in cornea, is flanked by the gene coding for cytokeratin 6* found in tongue and esophagus, i.e. the two cytokeratins exhibit a totally different cell-type pattern of expression (Figure 7; see also Jorcano *et al.*, 1984a; Schiller, 1985; Cooper and Sun, 1986). The regulation of the linked cytokeratin pair III/IV is even more complex. These two keratins and their mRNAs are co-expressed, although at different concentrations, in a series of bovine tissues (muzzle and heel pad epidermis and tongue mucosa) and cultured cell lines (BMGE+H) but in cornea only cytokeratin III is expressed (Figure 7; Jorcano *et al.*, 1984a; Schiller, 1985; Cooper and Sun, 1986; Knapp *et al.*, 1986). Clearly, these two linked cytokeratin genes are co-expressed in some cells but differentially expressed in others. On the other hand, some typical epidermal cytokeratins

such as Ia and Ib (Ib is found also in cornea) do not seem to be linked (Figure 1B and unpublished results).

It is also worth noting that none of the four type II cytokeratin genes reported here seems to be linked to the gene coding for the corresponding type I complex partner, i.e. cytokeratins VI, VII and 18 (Schiller, 1985; Cooper and Sun, 1986). Thus, the linkage of certain cytokeratin genes seems to be an evolutionary consequence rather than a means to control the co-ordinate expression of this multigene family. Corresponding results have been reported for the rat sarcomeric myosin heavy chain (MHC) gene family (Mahdavi *et al.*, 1986). The genes encoding the cardiac α- and β-MHC, which are linked in tandem, are co-expressed in the ventricles at almost all stages of development, from fetal stages to adult. However, in adult animals the α-MHC mRNA is strongly expressed in the atrial muscle but not found in other muscles. In contrast, the β-MHC mRNA is found in several muscles containing slow fibers but barely detected in the atrial muscle.

The 5′-flanking regions have been shown to play an important role in the transcriptional regulation of many eukaryotic genes. In different species (bovine, murine and human) these regions are highly homologous in genes coding for the same, or equivalent, cytokeratin (Figure 8). A similar evolutionary conservation has also been detected in several other eukaryotic genes (e.g. Yang *et al.*, 1984; Miyatake *et al.*, 1985; Robert *et al.*, 1986), including members of other multigene families (e.g. Minty and Kedes, 1986), and it has been shown that the conserved 5′-flanking sequences are essential for an efficient cell type-specific expression of some of these genes (e.g. Walker *et al.*, 1983; Minty and Kedes, 1986). As the 5′-flanking sequences of the cytokeratin genes have withstood mutational pressure during evolution these regions are probably of functional importance for expression. Indeed, when tested in chloramphenicol acetyltransferase (CAT) assays according to Gorman *et al.* (1982), the 2 kb preceding the 5′-end of the gene encoding cytokeratin IV directed the CAT synthesis as strongly as the SV40 promoter (results not shown).

Chaudhury *et al.* (1986) have reported that the human genes encoding two acidic (nos. 14 and 17) and one basic (no. 6) cytokeratins that are co-expressed in cultured epidermal keratinocytes exhibit a significant degree of sequence homology in the 5′-untranslated and flanking regions, suggesting that these sequences play an important role in the co-ordinated expression of the three genes. However, the corresponding parts of the genes encoding cytokeratins Ia, Ib, IV and VIb, all of which are co-expressed in epidermis and some of which also occur in other bovine epithelial cells (Schiller *et al.*, 1982; Jorcano *et al.*, 1984a; Schiller, 1985; Cooper and Sun, 1986) do not show extensive homology. The only motif consistently conserved besides the TATA box is a AAPuCCAAA box found upstream of the TATA box in all genes encoding bovine, human and murine epidermal cytokeratins analyzed so far. This symmetrical sequence can be located at somewhat variable distances from the TATA box in different genes, but its position in a given gene is conserved in different species (Figure 8). This motif is not found in the genes encoding several human, bovine and murine cytokeratins that are expressed in simple epithelia (Vasseur *et al.*, 1985; Bader *et al.*, 1986; Chaudhury *et al.*, 1986) nor is it seen in the gene encoding the neurofilament polypeptide NF-L (Lewis and Cowan, 1986). Interestingly, however, a very similar sequence lies upstream of the gene coding for involucrin, a protein also synthesized by the epidermal keratinocytes (Table I; cf. Eckert and Green, 1986). Although some of the genes carrying this box are also expressed in other stratified epithelia, four are expressed only in epidermis (KB Ia, KB VIb, KH 1, KM 59K). The functional relevance of this sequence for the epidermal expression of these genes will be studied using biological tests and confirmed by additional sequence information from other epidermal and non-epidermal cytokeratin genes.

This lack of extensive homology in the 5′-flanking region is also found among members of other multigene families like the chicken cardiac and skeletal α-actin genes which are co-expressed in skeletal muscle during embryonic development (Eldridge *et al.*, 1985) or other co-ordinately expressed genes such as those coding for the several constituents of the striated muscle (Robert *et al.*, 1986). This suggests that in spite of the coincidental expression of several cytokeratin genes in some tissues there may be different factors interacting with the regulatory elements of the genes which specify their spatially and temporally differential expression.

Materials and methods

Isolation and characterization of genomic clones

The screening of the bovine genomic library and the isolation of the phages hybridizing with the pUC plasmids carrying genomic restriction fragments or cDNA sequences coding for the bovine epidermal keratins III, IV and Ib (Jorcano *et al.*, 1984a) and the identification of the genomic clones by means of hybrid-selection–translation assays, were performed as previously described (Lehnert *et al.*, 1984). Selected regions of the genomic clones were sequenced by the chemical cleavage method of Maxam and Gilbert (1980). The position and polarity of the genes inside the clones was established by Southern blot analysis (Southern, 1975) probing sequentially the nitrocellulose filters with nick-translated clones from the 3′-end of the cDNA and with calf muzzle epidermis poly(A)$^+$ RNA, ^{32}P-labelled with T4 polynucleotide kinase (Maizels, 1976). Hybridization and washing conditions were as described by Lehnert *et al.* (1984), except that when end-labelled RNA was used as a probe, the filters were washed at 55°C. In some experiments, radioactive strand-specific RNA probes from genomic fragments subcloned in the vectors pTZ18R or pTZ19R (Pharmacia, Freiburg, FRG) were synthesized using T7 RNA polymerase (Biolabs, Schwalbach, FRG) and [^{32}P]-UTP (NEN, Dreieich, FRG) according to the manufacturer's instructions. In this case, the hybridization and washing conditions were as described for Northern hybridizations (see below).

Northern and slot–blot hybridization

RNA was isolated from different bovine tissues and cultured cell lines following the guanidinium hydrochloride method as described (Kreis *et al.*, 1983); 10 μg of total RNA were glyoxylated, electrophoresed on 1.2% agarose gels and transferred to nitrocellulose filters according to the method of Thomas (1980), as described (Jorcano *et al.*, 1984a). ^{32}P-Labelled anti-sense RNAs were synthesized by T7 RNA polymerase from restriction fragments containing the 3′-ends of the genes coding for cytokeratins Ib, III, IV and 6* cloned in vector pTZ18(19)R (Pharmacia) according to the manufacturer's instructions and used as hybridization probes. The hybridization took place in 50% formamide, 5 × SSPE [1 × SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4 (pH 7.4), 1 mM EDTA (pH 7.4)], 5 × Denhardt's solution, 100 μg/ml yeast tRNA and 0.1% SDS at 60°C for 12–18 h. The filters were washed in 2 × SSPE, 0.1% SDS for 40 min at room temperature, incubated for 20 min at room temperature in the same buffer containing 10 μg/ml RNase A and finally washed in 0.1 × SSPE, 0.1% SDS at 73°C for 2 h.

For slot-blot hybridization, 10 μg of glyoxylated total RNA was immobilized on nitrocellulose filters with the help of a Manifold II device (Schleicher and Schuell, Dassel, FRG). The probes and the hybridization and washing conditions were the same as those used for Northern hybridization.

Heteroduplex analysis

Heteroduplexes were formed between the genomic clones and tongue mucosa RNA and analyzed by electron microscopy as described (Lehnert *et al.*, 1984), except that five to ten times higher amounts of total RNA were used instead of poly(A)$^+$ mRNA.

Acknowledgements

We are indebted to our colleagues S.Rupert and Dr G.Schütz for making available to us the genomic library, T.M.Magin and Dr P.Cowin for helpful discussions and critically reading the manuscript, and C.T.Bock, M.Freudemann and R.Zimbelmann for skillful technical assistance. We thank M.Rieger for attrac-

ting our attention to the sequence conservation in the 5'-upstream region of the corresponding bovine and murine genes. In particular we thank W.W.Franke for valuable discussions, encouragement and support and critical reading of the manuscript. The work has been supported in part by the German Ministry for Research and Technology (ZMBH Project) and by the Deutsche Forschungsgemeinschaft (DFG).

References

Bader,B.L., Magin,T.M., Hatzfeld,M. and Franke,W.W. (1986) *EMBO J.*, **5**, 1866–1875.

Balcarek,J.M. and Cowan,N.J. (1985) *Nucleic Acids Res.*, **13**, 5527–5543.

Blessing,M. (1985) Diploma Thesis. University of Heidelberg.

Chaudhury,A.R., Marchuk,D., Lindhurst,M. and Fuchs,E. (1986) *Mol. Cell. Biol.*, **6**, 539–548.

Cooper,D. and Sun,T.-T. (1986) *J. Biol. Chem.*, **261**, 4646–4654.

Cooper,D., Schermer,A. and Sun,T.-T. (1985) *Lab. Invest.*, **52**, 243–256.

Crewther,W.G., Dowling,L.M., Gough,K.H., Marshall,R.C. and Sparrow,L.G. (1980) In Parry,D.A.D. and Creamer,L.K. (eds), *Fibrous Proteins: Scientific Industrial and Medical Aspects*. Academic Press, London, Vol. 2, pp. 151–159.

Crewther,W.G., Dowling,L.-M., Steinert,P.M. and Parry,D.A.D. (1983) *Int. J. Macromol.*, **5**, 267–274.

Davis,M.M., Stuart,K.K. and Hood,L. (1980) *Cell*, **22**, 1–2.

Eldridge,J., Zehner,Z. and Paterson,B.M. (1985) *Gene*, **36**, 55–63.

Eckert,R.L. and Green,H. (1986) *Cell*, **46**, 583–589.

Franke,W.W., Schmid,E., Schiller,D.L., Winter,S., Jarasch,E.-D., Moll,R., Denk,H., Jackson,B.W. and Illmensee,K. (1982) *Cold Spring Harbor Symp. Quant. Biol.*, **46**, 431–453.

Franke,W.W., Schiller,D.L., Hatzfeld,M. and Winter,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7113–7117.

Fraser,R.D.B., MacRae,T.P., Parry,D.A.D. and Suzuki,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1179–1183.

Fuchs,E. and Hanukoglu,I. (1983) *Cell*, **34**, 332–334.

Fuchs,E., Coppock,S.M., Green,H. and Cleveland,D.W. (1981) *Cell*, **27**, 75–84.

Geisler,N. and Weber,K. (1982) *EMBO J.*, **1**, 1649–1656.

Glass,C., Kim,K.H. and Fuchs,E. (1985) *J. Cell Biol.*, **101**, 2366–2373.

Gorman,C.M., Moffat,L.F. and Howard,B. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.

Hanukoglu,I. and Fuchs,E. (1983) *Cell*, **33**, 915–924.

Heid,H.W., Werner,E. and Franke,W.W. (1986) *Differentiation*, **32**, 101–119.

Johnson,L.D., Idler,W.W., Zhou,X.-M., Roop,D. and Steinert,P.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1896–1900.

Jorcano,J.L., Magin,T.M. and Franke,W.W. (1984a) *J. Mol. Biol.*, **175**, 21–37.

Jorcano,J.L., Rieger,M., Franz,J.K., Schiller,D.L., Moll,R. and Franke,W.W. (1984b) *J. Mol. Biol.*, **179**, 257–281.

Jorcano,J.L., Franz,J.K. and Franke,W.W. (1984c) *Differentiation*, **28**, 155–163.

Kim,K.H., Rheinwald,J.G. and Fuchs,E. (1983) *Mol. Cell. Biol.*, **3**, 495–502.

Knapp,A.C., Franke,W.W., Heid,H., Hatzfeld,M., Jorcano,J.L. and Moll,R. (1986) *J. Cell Biol.*, **103**, 657–667.

Kreis,T.E., Geiger,B., Schmid,E., Jorcano,J.L. and Franke,W.W. (1983) *Cell*, **32**, 1125–1137.

Krieg,T.M., Schafer,M.P., Cheng,C.K., Filpula,D., Flaherty,P., Steinert,P.M. and Roop,D.R. (1985) *J. Biol. Chem.*, **260**, 5867–5870.

Lazarides,E. (1982) *Annu. Rev. Biochem.*, **51**, 219–250.

Lehnert,M.E., Jorcano,J.L., Zentgraf,H., Blesing,M., Franz,J.K. and Franke, W.W. (1984) *EMBO J.*, **3**, 3279–3287.

Leube,R.E., Bosch,F.X., Romano,V., Zimbelmann,R., Höfler,H. and Franke, W.W. (1986) *Differentiation*, **33**, 69–85.

Lewis,S.A. and Cowan,N.J. (1986) *Mol. Cell. Biol.*, **6**, 1529–1534.

Magin,T.M., Jorcano,J.L. and Franke,W.W. (1986) *Differentiation*, **30**, 254–264.

Mahdavi,V., Strehler,E.E. Periasamy,M., Wieczorek,D., Izumo,S., Grund,S., Strehler,M.-A. and Nadal-Ginard,B. (1986) In Emerson,C., Fischman,D., Nadal-Ginard,B. and Siddiqui,M.A.Q. (eds), *Molecular Biology of Muscle Development*. UCLA Symposia on Molecular and Cellular Biology, **29**, pp. 345–361.

Maizels,N. (1976) *Cell*, **9**, 431–438.

Marchuk,D., McCrohon,S. and Fuchs,E. (1984) *Cell*, **39**, 491–498.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.

McKeon,F.D., Kirschner,M.W. and Caput,D. (1986) *Nature*, **319**, 463–468.

Minty,A. and Kedes,L. (1986) *Mol. Cell. Biol.*, **6**, 2125–2136.

Miyatake,S., Otsuka,T., Yokota,T., Lee,F. and Arai,K. (1985) *EMBO J.*, **4**, 2561–2568.

Moll,R., Franke,W.W., Schiller,D.L., Geiger,B. and Krepler,R. (1982) *Cell*, **32**, 11–24.

Osborn,M. and Weber,K. (1983) *Lab. Invest.*, **48**, 372–393.

Parry,D.A.D., Steven,A.C. and Steinert,P.M .(1985) *Biochem. Biophys. Acta Commun.*, **127**, 1012–1018.

Proudfoot,N.J., Shander,M.H.M., Manley,J.L., Gefter,M.L. and Maniatis,T. (1980) *Science*, **209**, 1329–1336.

Powell,C.P., Cam,G.R., Fietz,M.J. and Rogers,G.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5048–5052.

Quax,W., Egberts, W.V., Hendriks,W., Quax-Jeuken,Y. and Bloemendal,H. (1983) *Cell*, **35**, 215–223.

Quax,W., van den Broek,L., Egberts,W.V., Ramaekers,F. and Bloemendal,H. (1985) *Cell*, **43**, 327–338.

Quinlan,R.A., Cohlberg,J.A., Schiller,D.L., Hatzfeld,M. and Franke,W.W. (1984) *J. Mol. Biol.*, **178**, 365–388.

Quinlan,R.A., Schiller,D.L., Hatzfeld,M., Achtstaetter,T., Moll,R., Jorcano, J.L., Magin,T.M. and Franke,W.W. (1985) In Wang,E., Fischman,D., Liem,R.K.H. and Sun,T.-T. (eds), *Intermediate Filaments. Ann. N.Y. Acad. Sci.*, **455**, 282–306.

Rieger,M., Jorcano,J.L. and Franke,W.W. (1985) *EMBO J.*, **4**, 2261–2267.

Robert,B., Barton,P., Alonso,S., Cohen,A., Daubas,P., Garnier,I., Weydert,A. and Buckingham,M. (1986) In Emerson,C., Fischman,D., Nadal-Ginard,B. and Siddiqui,M.A.Q. (eds), *Molecular Biology of Muscle Development*. UCLA Symposia on Molecular and Cellular Biology, Vol. 29, pp. 487–506.

Rogers,G.E. (1985) In Wang,E., Fischman,D., Limen,R.K.H. and Sun,T.-T. (eds), *Intermediate Filaments. Ann. N.Y. Acad. Sci.*, **455**, 403–425.

Roop,D.R., Hawley-Nelson,P., Cheng,C.K. and Yuspa,S.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 716–720.

Schiller,D.L. (1985) Doctoral Thesis. University of Heidelberg.

Schiller,D.L., Franke,W.W. and Geiger,B. (1982) *EMBO J.*, **1**, 761–769.

Schmid,E., Schiller,D.L., Grund,C., Stadler,J. and Franke,W.W. (1983) *J. Cell Biol.*, **96**, 37–50.

Southern,E. (1975) *J. Mol. Biol.*, **98**, 503–517.

Steinert,P.M., Steven,A.C. and Roop,D.R. (1985) *Cell*, **42**, 411–419.

Sun,T.-T., Eichner,R., Schermer,A., Cooper,D., Nelson,W.G. and Weiss,R.A. (1984) In Levine,A.J., Vande Woude,G.F. and Watson,J.D. (eds), *Cancer Cells. 1. The Transformed Phenotype*. Cold Spring Harbor Laboratory, pp. 169–176.

Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.

Tonegawa,S. (1983) *Nature*, **302**, 575–581.

Tyner,A.L., Eichman,M.J. and Fuchs,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4683–4687.

Vasseur,M., Duprey,P., Brûlet,P. and Jacob,F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1155–1159.

Walker,M.D., Edhund,T., Boulet,A. and Rutter,W.J. (1983) *Nature*, **306**, 557–561.

Weber,K. and Geisler,N. (1984) In Levine,A.J., Vande Woude,G.F. and Watson,J.D. (eds), *Cancer Cells. 1. The Transformed Phenotype*. Cold Spring Harbor Laboratory, pp. 153–159.

Woods,E.F. and Inglis,A.S. (1984) *Int. J. Biol. Macromol.*, **6**, 277–283.

Yang,J.K., Masters,J.N. and Attardi,G. (1984) *J. Mol. Biol.*, **176**, 169–187.

Received on November 12, 1986; revised on January 6, 1987