The EMBO Journal vol.3 no.13 pp.3279–3287, 1984

Characterization of bovine keratin genes: similarities of exon patterns in genes coding for different keratins

Marianne E. Lehnert[1], José L. Jorcano[1,2], Hanswalter Zentgraf[3], Manfred Blessing[2], Juergen K. Franz[1] and Werner W. Franke[1]

[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 W.W. Franke

Four different genomic clones which contain the genes coding for epidermal keratins Ia (mol. wt. ~68 000), Ib (68 000), III (60 000) and VIb (54 500) have been selected using cDNA probes and identified by hybrid-selection translation. The genes vary considerably in length, primarily due to differences in intron sizes: keratin Ia, 9.3 kb (~2.55 kb total exons); keratin Ib, 6.0 kb (2.25 kb exons); keratin III, 6.0 kb (2.2 kb exons); keratin VIb, 4.4 kb (1.85 kb exons). The genes for all three representatives of the basic (type II) cytokeratin subfamily, i.e., keratins Ia, Ib and III, contain eight introns of variable sizes (0.1–1.8 kb) and their exon patterns are very similar. The gene coding for keratin VIb, a representative of the acidic (type I) subfamily, contains seven introns, and the size pattern of its five innermost exons closely resembles that of the genes of the type II keratins. Most of the introns are located in regions coding for the α-helical cores of these proteins. Mapping of the intron positions by the S1 nuclease technique and sequencing of some exon-intron boundaries has revealed that some of the introns of all four keratin genes have similar positions to each other and to those of the hamster vimentin gene. The correlation of intron 7 of the type II keratin genes, which corresponds to intron 6 of the gene of the type I keratin VIb, with intron 6 of the vimentin gene is particularly striking: they are located in a position disrupting the third amino acid after the TYR(R,K,S,T)LLEGE consensus sequence that defines the carboxy-terminal end of the α-helical core portion of all intermediate filament (IF) proteins. These common principles of genomic organization indicate that the members of the cytokeratin multigene family are related in their gene structure and support the hypothesis that the different IF proteins are derived from a common ancestral gene.

Key words: intermediate filaments/keratins/multigene families/gene structure/exon-intron patterns

Introduction

The intermediate-sized filaments (IF; 7–12 nm diameter) are self-assembly polymer structures which occur in the cytoplasm of most vertebrate cells. In different cell types they are composed of different proteins (for reviews, see Lazarides, 1982; Osborn and Weber, 1983): (i) epithelial and carcinoma cells contain cytokeratin IF, which comprise a group of proteins related to α-keratins of hair, wool and epidermis (Franke *et al.*, 1978a, 1978b, 1979a; Sun and Green, 1978; Sun *et al.*, 1979); (ii) mesenchymally-derived cells, including sarcomas, and certain other non-epithelial cells and tumors contain IF of the vimentin type (Bennett *et al.*, 1978; Franke *et al.*, 1978a; Hynes and Destree, 1978); (iii) most, but not all types of myogenic cells and myosarcomas are characterized by the expression of another IF protein, desmin (Lazarides and Hubbard, 1976; Small and Sobieszek, 1977); (iv) astrocytes and gliomas produce IF-containing glial filament protein (Liem *et al.*, 1978; Dahl and Bignami, 1983); (v) neuronal cells and certain tumors derived thereof possess IF containing one or more of the three neurofilament polypeptides (Bennett *et al.*, 1978; Liem *et al.*, 1978; Osborn and Weber, 1983). In addition, certain kinds of cells have been shown to co-express two different types of IF proteins which are then either segregated into different filaments (Franke *et al.*, 1978a, 1979b, 1979c; Osborn *et al.*, 1980) or form true heteropolymer filaments (Steinert *et al.*, 1981; Quinlan and Franke, 1982, 1983, Schmid *et al.*, 1982; Sharp *et al.*, 1982).

The cytokeratins present a special case among the IF proteins as they are coded by a large multigene family of ~20 different polypeptides ranging in mol. wt. from 40 000 to 68 000 and in isoelectric pH from ~5 to 8 (Franke *et al.*, 1981; Schiller *et al.*, 1982). In human tissues, for example, 19 cytokeratin polypeptides have been distinguished by biochemical and immunological criteria (Moll *et al.*, 1982; Tseng *et al.*, 1982). Moreover, this family can be subdivided into two 'subfamilies', i.e., the 'acidic' cytokeratins related to type I α-keratins of sheep wool (Crewther *et al.*, 1983) and the more 'basic' cytokeratins related to type II wool keratins. These two subfamilies can be distinguished by peptide maps (Schiller *et al.*, 1982), by certain monoclonal antibodies (Gigi *et al.*, 1982; Tseng *et al.*, 1982) and by differences in cross-hybridization of their mRNAs (Fuchs *et al.*, 1981; Kim *et al.*, 1983). The evidence to date indicates that at least one representative each of the acidic and of the basic subfamily are necessary to form distinct heterotypic tetramer complexes which are the subunits that spontaneously form IF *in vivo* and *in vitro* (Crewther *et al.*, 1983; Gruen and Woods, 1983; Franke *et al.*, 1983; Quinlan *et al.*, 1984).

A comparison of the amino acid sequences of different types of IF proteins has revealed a basic principle of organization. A central 'rod' of coiled coil α-helical arrangement of 300–330 amino acids is always found flanked by non-α-helical 'head' and 'tail' regions of variable lengths (Geisler and Weber, 1981, 1982; Weber and Geisler, 1984; Hanukoglu and Fuchs, 1982, 1983; Quax-Jeuken *et al.*, 1983; Steinert *et al.*, 1983). However, significant sequence homology between different IF proteins is restricted to the α-helical core portion, the non-α-helical head and tail regions of different IF proteins varying greatly in amino acid sequence. Moreover, cytokeratin representatives of the basic (type II) subfamily display relatively little amino acid sequence homology to polypeptides of the acidic (type I) subfamily (<30%). Again this homology is limited to the α-helical region (Hanukoglu and Fuchs, 1983). In addition, we have found that different members of the acidic cytokeratin subfamily can be

distinguished from each other by their grossly different extra-α-helical carboxyterminal tail sequences (Jorcano *et al.*, 1984a, 1984b).

To study the relationship of the diverse cytokeratin polypeptides to each other and to other IF proteins and to understand the regulation of their cell type-specific expression at the gene organization level, we have selected and compared genomic DNA clones. Here, we show the exon-intron organization of four different bovine keratin genes.

Results

Identification of genomic keratin clones

We have recently described (Jorcano *et al.*, 1984a, 1984b) cDNA clones for the major bovine epidermal cytokeratin polypeptides Ia (mol. wt. 68 000; No. 1 of the bovine cytokeratin catalog of Schiller *et al.*, 1982), Ib (mol. wt. 68 000; No. 2), III (mol. wt. 60 000; No. 6), IV (mol. wt. 59 000; No. 7), VIb (mol. wt. 54 500; No. 14) and VII (mol. wt. 50 000; No. 16). In contrast to previous reports (Fuchs *et al.*, 1981; Roop *et al.*, 1983) we have also shown that, under stringent hybridization conditions, the different mRNAs can be separated, indicating the absence of extensive nucleotide sequence homology (Franke *et al.*, 1984; Jorcano *et al.*, 1984a, 1984b; see also Kim *et al.*, 1983).

Using these cDNA clones and very stringent conditions we screened a genomic bovine library containing fragments obtained by partial digestion with *Sau*IIIA of calf thymus DNA and integration into λ phage EMBL 3. From a total of 20 positive clones we studied in greater detail several clones which contained the total information for the mRNAs coding for keratins Ia, Ib, III and VIb and which probably represented complete genes. Figure 1a–d presents examples of the identification of these clones on the basis of their specific hybridization with cDNA probes in Southern blot analyses. In addition, we have identified the genomic clones by hybrid-selection-translation assays, using poly(A)$^+$ RNA from bovine muzzle epidermis as previously described for the characterization of the corresponding cDNA clones (Jorcano *et al.*, 1984a). Figures 1e, e′ and f, f′ show examples of the results obtained with the clone coding for the largest (λKBIa^1) and the smallest (λKBVIb^1) of the four genes presented in this paper. The two other clones, i.e., λKBIb^1 and λKBIII^1, displayed a similar high selectivity in this assay (data not shown).

When DNA isolated from cow muzzle, calf thymus or bull sperm was digested with different restriction nucleases such as *Bam*HI, *Eco*RI and *Pst*I, and fragments obtained were separated by gel electrophoresis on agarose and examined by Southern blot hybridization with the different cDNAs to epidermal keratins, one, or in some cases two and three, positive bands were observed, indicating that the corresponding genes most likely occur as single copy genes (data not shown). As similar hybridization bands were found in DNA from both sperm and differentiated tissues appears to exclude considerable re-arrangements of these keratin genes during embryogenesis and organ development.

Heteroduplex analysis

After determining the optimal melting conditions, which were relatively high for all clones (see Materials and methods), we analyzed the heteroduplex molecules formed after hybridization of the genomic clones with total poly(A)$^+$ RNA from bovine muzzle epidermis by electron microscopy. Representative examples are shown in Figure 2A–D, with interpretative drawings (A′–D′). For each genomic clone at least 30 heteroduplices were morphometrically evaluated. In all clones the 3′ end was demarcated by the non-hybridizing poly(A) tail. In addition, extended regions upstream from the 5′ end of the heteroduplex were found, indicating that the inserts contained the promoter elements and probably represented full-length gene clones. Figure 3 presents the exon-intron maps of these genes.

All three genes coding for keratins of the basic (type II) subfamily revealed eight introns whereas only seven introns were detected in clone λKBVIb^1, a representative of the acidic (type I) subfamily. Intron lengths varied greatly within a given gene as did the lengths of corresponding introns in different genes. Total gene lengths determined, by electron microscopy, as the sum of all exons and introns showed great differences, primarily due to differences of intron lengths. The size of the gene coding for keratin Ia contained ~9.3 (± 0.4) kb of which ~6.75 kb were occupied by introns. The gene coding for keratin Ib, a polypeptide very similar in size to keratin Ia (Franke *et al.*, 1981; Schiller *et al.*, 1982) and amino acid sequence (Jorcano *et al.*, 1984c), was much shorter, i.e., 6.0 kb, 3.75 kb of which was represented by introns. The gene coding for keratin III, which is a considerably smaller polypeptide (mol. wt. 60 000), was of nearly the same total size as the gene for keratin Ib and also showed a similar proportion of total intron length (~3.8 kb). The gene for the acidic (type I) keratin VIb of mol. wt. 54 500 was considerably shorter (4.4 kb) and contained a total of 2.55 kb of intron sequences. These data should be compared with those of the only other IF protein gene analyzed so far, i.e., the gene coding for hamster vimentin (Quax *et al.*, 1983) which comprises ~10 kb of which ~8.1 kb are represented by the eight introns of this gene.

Figure 4a and b present the locations of the introns in the specific bovine cytokeratin genes, in comparison with the hamster vimentin gene. The pattern of exons B–H of all three representatives of the basic (type II) cytokeratin subfamily (Figure 4a), showed close similarity, although not identity, whereas the terminal exons (A and I) displayed more prominent size differences. Whether the slight tendency of exon size reduction from keratin Ia to keratins Ib and III which is seen in Figure 4a is real can only be decided by determination of the entire sequence; the differences shown in Figure 4a are still within the limits of error of the heteroduplex measurements. Comparison of the exon pattern of type II keratin genes with that of the gene coding for keratin VIb, the representative of the type I keratin subfamily (Figure 4b), showed similar lengths for exons C–G of subfamily II keratin genes and exons B–F of the type I keratin gene. In contrast, introns 1 and 8 of the type II keratin genes and intron 7 of the type I keratin gene had no corresponding counterpart in the other keratin gene subfamily. Comparison with the exon pattern of the hamster vimentin gene (bottom panel of Figure 4b) indicated similar arrays of some exons, notably the innermost ones, whereas other introns appeared to occur in locations unique to either the keratins or vimentin.

Determination of locations of introns

Comparison of our intron maps as determined by electron microscopy with the sequence data available for the corresponding bovine keratin cDNA clones (Jorcano *et al.*, 1984b, 1984c) as well as with the sequences of related keratins from other species (Hanukoglu and Fuchs, 1982, 1983;

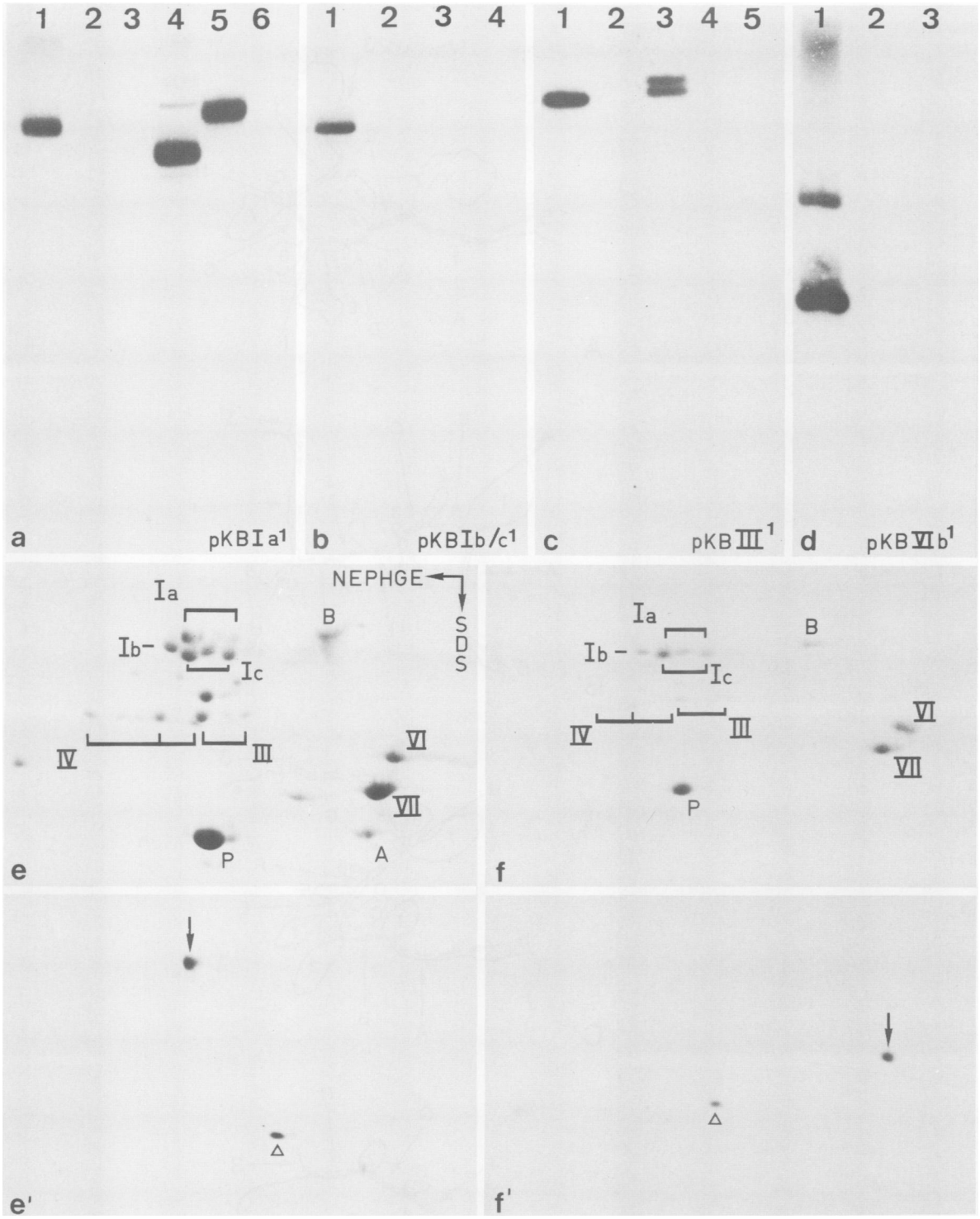

Fig. 1. Identification and characterization of genomic clones for bovine epidermal keratins by Southern blot analysis (**a**–**d**) and hybridization-selection-translation, using total poly(A)$^+$ RNA from bovine muzzle epidermis (**e**, **e'**, **f**, **f'**). (**a**–**d**) 2 μg of DNA isolated from clones positive in the screening of the bovine genomic λ-phage library were restricted, electrophoresed on 1% agarose gels, blotted to nitrocellulose filters and hybridized with the cDNA probes indicated at the bottom of the figures. Only positive clones are mentioned; lanes not showing a hybridizing band contain DNA from different clones negative with the specific cDNA probe used. **(a)** Clones coding for keratin Ia: **lane 1**, λKBIa^2; **lane 4**, λKBIa^3; **lane 5**, λKBIa^1; **lanes 2** and **6** contain DNA of negative clones. **(b) Lane 1**, clone λKBIb^1, coding for keratin Ib. **(c)** Clones coding for keratin III: **lane 1**, λKBIII^2; **lane 3**, λKBIII^1. **(d) Lane 1**, clone λKBVIb^1, coding for keratin VIb. (**e**,**e'**) Two-dimensional electrophoresis (NEPHGE, non-equilibrium pH gradient in the first dimension; SDS, electrophoresis in the presence of SDS used in the second dimension) of [^{35}S]methionine labelled products of translation *in vitro* using poly(A)$^+$ RNA selectively hybridized to clone λKBIa^1 in co-electrophoresis with total unlabelled bovine muzzle epidermal keratins (Roman numerals; cf. Schiller *et al.*, 1982; Jorcano *et al.*, 1984a) and reference proteins phosphoglycerokinase (P), bovine serum albumin (B) and α-actin (A). Coomassie blue staining **(e)** and fluorography **(e')** of the same gel is shown. The arrow in **e'** denotes keratin Ia synthesized *in vitro* showing the specific selection of this RNA by λKBIa^1. (**f**,**f'**) Co-electrophoresis (similar conditions as in **e** and **e'**) of *in vitro* translated [^{35}S]methionine-labelled products from mRNA selected by clone λKBVIb^1 with epidermal keratins (symbols as in **e**). Coomassie blue staining **(f)** is compared with fluorography **(f')** of the same gel. Arrow in **f'** denotes labelled keratin polypeptide VIb, indicating the specificity of hybridization of this mRNA to the genomic clone examined. Open triangles denote a labelled product endogenous to the rabbit reticulocyte system used. The gel system used here does not resolve components VIa and b (for separation see Kreis *et al.*, 1983; Jorcano *et al.*, 1984a).

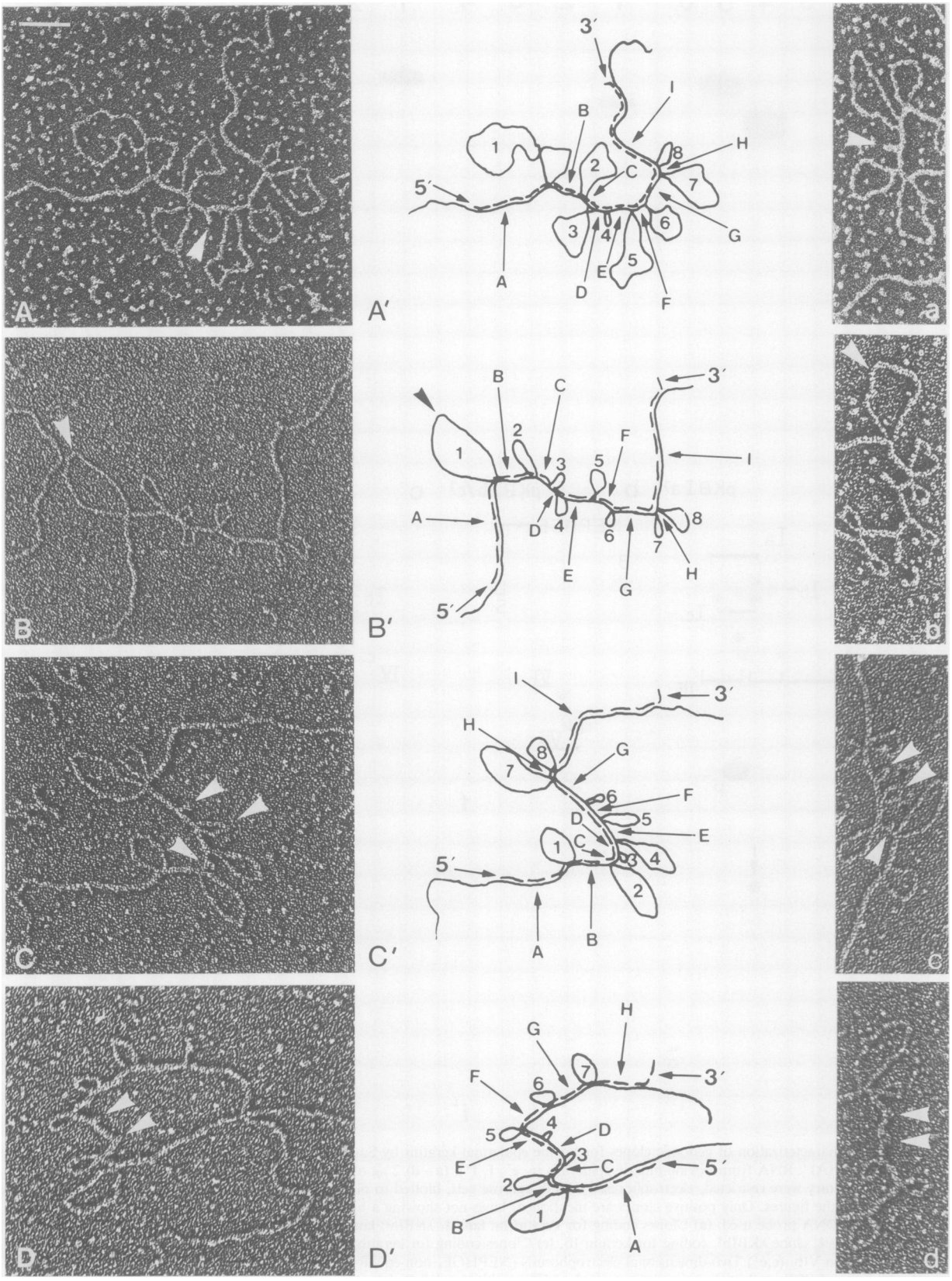

Fig. 2. Representative electron micrographs showing heteroduplex molecules of poly(A)$^+$ RNA from bovine muzzle epidermis with different genomic clones **(A – D)**, together with interpretative drawings **(A′ – D′)**, and details of certain introns **(a – d)**. In **A′ – D′**, DNA is represented by a continuous line, RNA by an interrupted line; 5′ ends of mRNA are identified by a change in molecular diameter, 3′ ends are demarcated by the projecting, i.e., non-hybridized poly(A) tails. Exons are denoted by capital letters, introns by arabic numerals. **(A,A′,a)** Clone λKBIa[1] (intron 4, denoted by arrowhead, is shown in greater clarity in the molecule presented in **a**); **(B,B′,b)** Clone λKBIb[1] (the intron nature of the loop designated intron 1, denoted by the arrowhead, is better resolved in the molecule shown in **b**). **(C,C′,c)** Clone λKBIII[1] (introns 3,5,6 denoted by the arrowheads are better resolved in the molecule shown in **c**). **(D,D′d)** Clone λKBVIb[1] (the very small introns 3 and 4 are shown in a different molecule in **d**). For contrast enhancement pictures have been printed as negatives.

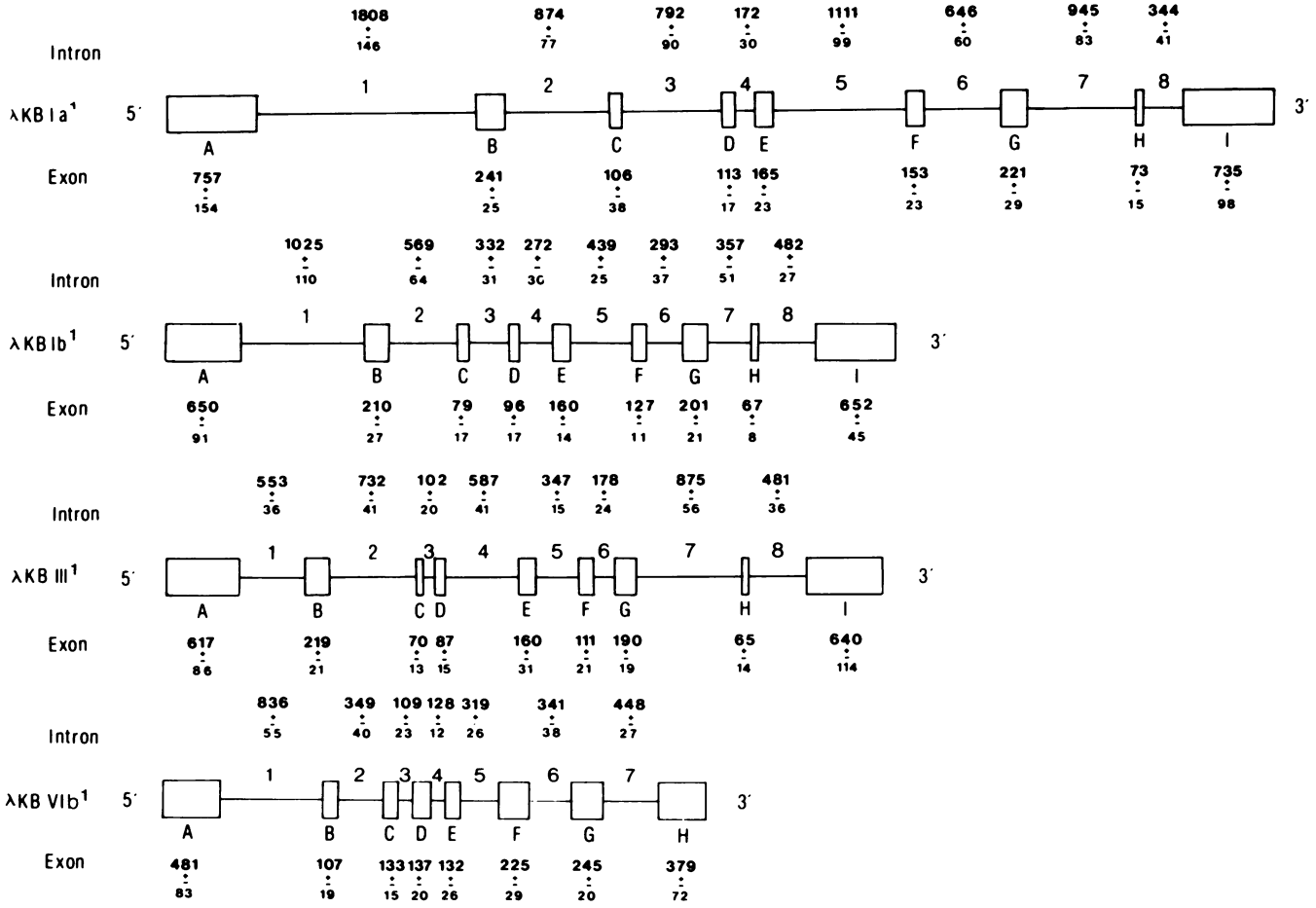

Fig. 3. Schematic summary of electron microscopic length determination of the exons and introns of the four bovine genomic keratin clones examined by heteroduplex formation. Symbols are as in Figure 2. Mean values of lengths (in base pair units) and standard deviations are indicated.

Steinert *et al.*, 1983; Hoffmann and Franz, 1984) indicated that all introns, except for the most 3′ one, are located within the gene region corresponding to the α-helical core characteristic of all IF proteins. For example, intron 7 of the genes coding for the type II keratins Ia, Ib and III and intron 6 of the gene for the type I keratin VIb appeared to be located close to the carboxy-terminal end of the α-helical core. Therefore, we have mapped the position of the 3′ ends of some of the introns of the genomic clones by the S1 nuclease method, using appropriate restriction fragments. Figure 5 presents an example, showing that the insertion site of intron 7 of λKBIa¹ maps 149 nucleotides downstream of the *Hind*III cleavage site present in exon G (for cDNA sequence data see Jorcano *et al.*, 1984c).

In order to confirm the S1 nuclease map results, we sequenced, in some cases, the same restriction fragments by the method of Maxam and Gilbert (1980). Figure 6 presents a part of the sequence determined for the same *Hind*III-*Sac*II fragment used for the S1 mapping shown in Figure 5. The exon-intron boundary is located at the same position predicted from the S1 nuclease mapping, i.e., eight nucleotides after the region coding for the consensus sequence TYR-(R,K,S,T)LLEGE common to all IF proteins (Geisler *et al.*, 1983), disrupting an arginine triplet (Figure 6, upper lines). Comparison with the hamster vimentin gene sequence reported by Quax *et al.* (1983) shows that exactly the same position is occupied by intron 6 of the vimentin gene (Figure 6, bottom lines). The same location was found for intron 7 of clone λKBIb¹. Detailed sequence comparisons of diverse introns of keratin genes will be published elsewhere.

Discussion

The genomic clones described in this study allow a comparison of the gene organization of different IF proteins to be made both between closely related proteins such as the epidermal cytokeratins of the basic (type II) subfamily and between more distantly related IF proteins such as between cytokeratins and vimentin. Our results demonstrate that different cytokeratin genes can differ greatly in size and that there is no simple correlation between gene size and polypeptide size. For example, the genes coding for epidermal keratin polypeptides Ia and Ib, i.e. proteins which are similar in size (mol. wt. 68 000), peptide maps and amino acid sequence, differ considerably in size. These differences can be even more pronounced when different IF proteins are compared: for example, the gene coding for hamster vimentin of mol. wt. 53 500 (Quax *et al.*, 1983) is larger (~10 kb) than each of the four bovine keratin genes. Such differences in gene sizes, which primarily reflect differences in the sizes of introns and the 3′-non-translated regions, have been reported for other multigene family proteins, one of the most striking examples being the actin family (e.g., Hamada *et al.*, 1982; Nudel *et*

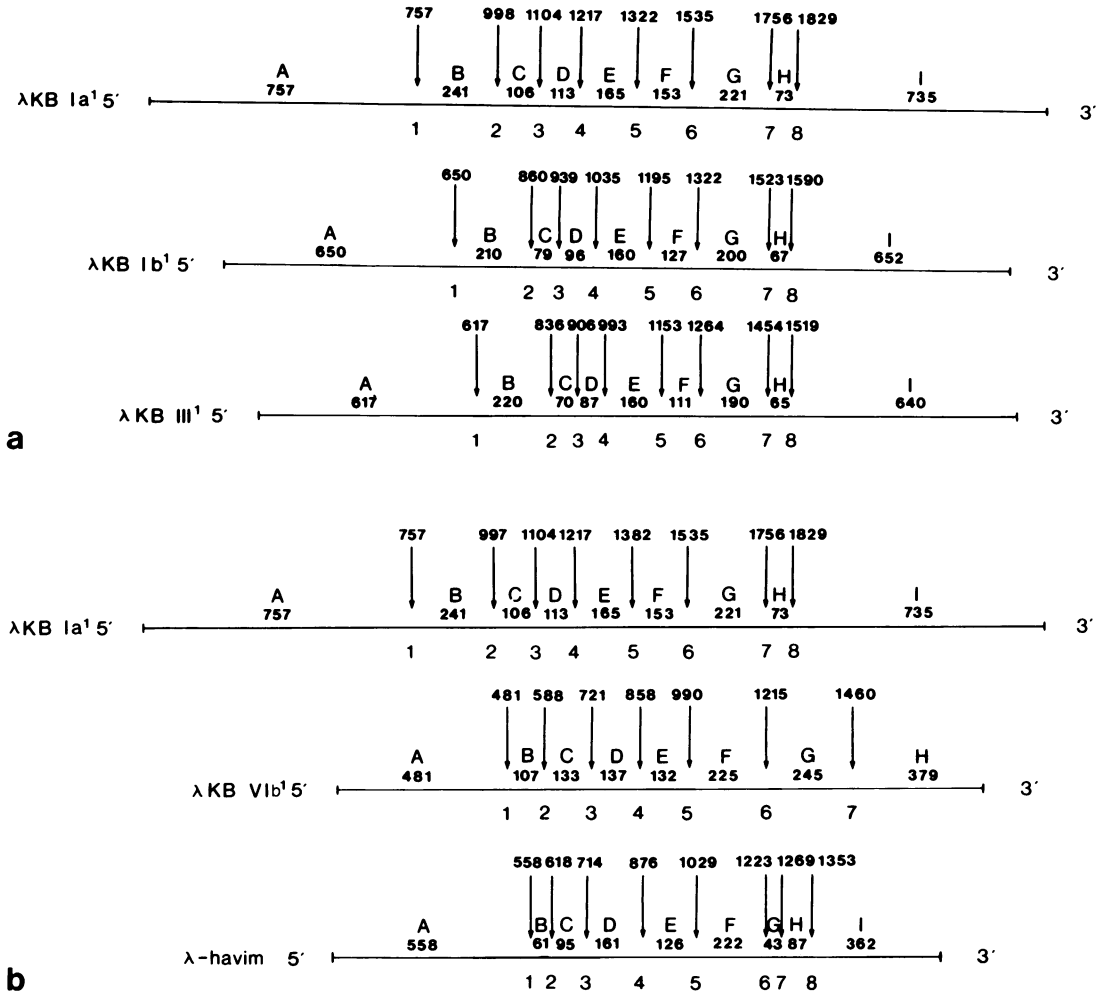

Fig. 4. Schematic presentation of locations of introns in genes coding for different intermediate filament proteins (symbols as in Figures 2 and 3). Insertion sites of introns are denoted by vertical arrows and the specific distances from the 5′ ends are indicated by the mean values (bp); numbers on the horizontal axes indicate the lengths of the specific exons (in bp). **(a)** Comparison of exon patterns in the bovine genes of three representatives of the basic (type II) cytokeratin subfamily, showing similar intron insertion points in all three genes. **(b)** Comparison of the gene organization of two bovine keratin-genes, i.e., a representative of the basic (type II) cytokeratin subfamily (upper panel, clone λKBIa[1]) and a representative of the acidic (type I) subfamily (central panel, clone λKBVIb[1]), with that of another intermediate filament protein, hamster vimentin (bottom panel, clone λ-havim; data from Quax *et al.*, 1983). In **a** and **b**, the different genes have been aligned such that the position of intron 7 of the basic (type II) subfamily keratin genes corresponds to that of intron 6 of the gene coding for keratin VIb and the hamster vimentin gene.

al., 1982; Ueyama *et al.*, 1984).

Our results show considerable variation of total intron lengths and intron length patterns among cytokeratin genes. Considerable variation occurs even between genes of similar total sizes and between genes coding for very similar polypeptides of the same subfamily. Again this is a feature reported for other multigene families (for a review, see Breathnach and Chambon, 1981), including non-IF cytoskeletal proteins such as actins (see references above) and tubulins (Gwo-Shu Lee *et al.*, 1983; see there for further refs.).

The differences in lengths of introns of keratin genes is in contrast to the apparent constancy of intron numbers in a given IF protein category: eight introns have been noted in the neutral-basic epidermal bovine cytokeratins of the type II family, the same number in vimentin (Quax *et al.*, 1983), whereas the type I keratin gene contains seven introns. The similarities of intron insertion sites and the resulting exon patterns are conspicuous not only between cytokeratin genes of the same basic (type II) subfamily but also, though less strikingly, between the two cytokeratin subfamilies. The heteroduplex analyses, S1 nuclease mapping and partial sequence data indicate that some introns occur in similar sequence positions in the genes of both subfamilies, whereas the positions of introns 1 and 8 of the basic (type II) cytokeratin genes do not have corresponding introns in the gene for keratin VIb[1], the representative of the acidic (type I) cytokeratin subfamily. *Vice versa*, intron 7 of the latter gene is located in a region where no intron has been detected in any of the type II cytokeratin genes. The remarkable example of precision of a common intron position (intron 7 of the basic keratin subfamily) not only in genes of the same IF protein type, i.e., cytokeratins, but also in a non-epithelial IF protein of a taxonomically distant vertebrate species, i.e. hamster vimentin, indicates that at least some intron positions have been conserved during IF protein diversification and species evolution. Similar observations of identical or corresponding intron sites have been reported for several other multigene family proteins of vertebrates such as globins, vitellogenins, ovalbumin and related proteins, lens crystallins, actins and tubulins (for references see Breathnach and Chambon, 1981;

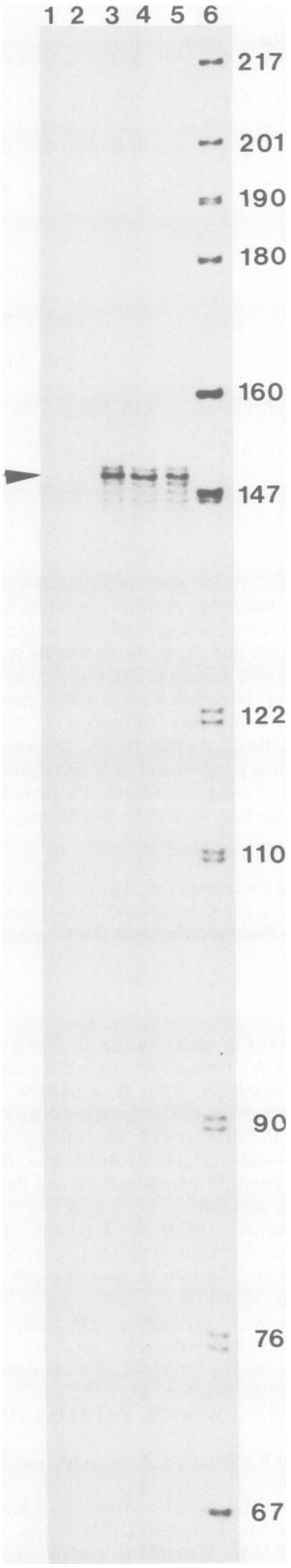

Fig. 5. Determination of the position of intron 7 in the gene coding for bovine keratin Ia by S1 nuclease mapping. A 2-kb *Hind*III-*Sac*I restriction fragment ^{32}P-labelled at the *Hind*III end, was hybridized with cow snout epidermal poly(A)$^+$ mRNA at different temperatures. The hybrids were digested with S1 nuclease and the protected DNA fragments were analyzed on a DNA sequencing gel. **Lane 1**, control lacking RNA; **lane 2**, control with tRNA; **lanes 3–5**, DNA-RNA hybrids formed at 56°C **(3)**, 58°C **(4)** or 60°C **(5)**; **lane 6**, DNA size markers (bp units). The arrowhead points to the S1-protected band, representing a DNA fragment 149 bases long.

Gwo-Shu Lee *et al.*, 1983; Nudel *et al.*, 1982; Hosbach *et al.*, 1983; Inana *et al.*, 1983; Ueyama *et al.*, 1984).

With the exception of the last intron, all introns of the bovine keratin genes examined so far are located in the coding region with most of them occurring in the gene intercept which codes for the α-helical central rod portion of the polypeptide. When the exon arrays are compared with current models of organization of the α-helical core and its local interruptions (Geisler and Weber, 1981, 1982; Hanukoglu and Fuchs, 1983; Steinert *et al.*. 1983; for review, see Weber and Geisler, 1984), some intron positions seem to be related to certain aspects of protein structure. This is clearly so for intron 7 of the basic (type II) cytokeratins which corresponds to intron 6 of the type I cytokeratin VI and hamster vimentin. This intron is located precisely at the end of the α-helical rod, i.e., in a position defining a marked change in protein conformation. This analysis of cytokeratin gene structure adds a remarkable example to the list of intron locations assumed to demarcate structurally and/or functionally defined protein subunits (Gilbert, 1978; Artymiuk *et al.*, 1981; Inana *et al.*, 1983; Yamada *et al.*, 1980; for review, see Breathnach and Chambon, 1981). Whether such similarities of certain exon patterns are a general feature of all IF proteins will become clear when information on the gene organization of other IF proteins is available. However, the observation of several similarities of exon patterns in different bovine keratin genes and in the hamster vimentin gene supports our earlier suggestion that the different IF proteins may be 'derived from a common ancestral gene' (Franke *et al.*, 1978a).

Materials and methods

Isolation and identification of genomic clones

A bovine genomic library (Ruppert *et al.*, 1984) was kindly provided by S. Ruppert and G. Schütz (this institute). A number of plaques encompassing three times the size of the bovine genome were screened essentially according to Benton and Davis (1977). As hybridization probes we used plasmids pKBIa[1], pKBIb/c[1], pKBIII[1] and pKBVIb[1], which code for the bovine epidermal keratins Ia, Ib and c, III and VIb, respectively (Jorcano *et al.*, 1984a), that were ^{32}P-labelled by nick-translation (Maniatis *et al.*, 1982; Rigby *et al.*, 1977). The hybridized filters were extensively washed in 2 x SSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% (w/v) SDS and finally twice for 1 h each in 0.1 x SSC, 0.1% SDS at 65°C. Under these conditions, the cDNA clones used as probes did not show apparent cross-hybridization with RNA or DNA coding for other members of the keratin multigene family (Jorcano *et al.*, 1984a, 1984b; and, unpublished observations).

The clones positive in the screening were further characterized in two ways. (i) DNA was isolated from the phages according to Maniatis *et al.* (1982), restricted with different enzymes, run on agarose gels and blotted to nitrocellulose filters (Southern, 1975). The filters were hybridized with the nick-translated cDNA clones under the same stringency conditions used in the screening of the genomic library, and the genomic clones hybridizing specifically with the different cDNA clones were visualized by autoradiography. (ii) 10 μg of recombinant phage DNA was bound to small nitrocellulose filters and hybridized to cow muzzle poly(A)$^+$ mRNA as described (Jorcano *et al.*, 1984a). The RNA remaining bound to the filters after washing at 65°C in 0.1 x SSC was eluted by boiling in water, precipitated with ethanol and translated *in vitro* as described (Jorcano *et al.*, 1984a; Magin *et al.*, 1983). The products synthesized *in vitro* were analyzed by two-dimensional gel electrophoresis, using non-equilibrium pH gradient electrophoresis in the first dimension, as reported previously (Franke *et al.*, 1981; Kreis *et al.*, 1983; Magin *et al.*, 1983).

Heteroduplex analysis

Heteroduplex molecules of purified cloned DNA (10 μg/ml) and epidermal cow muzzle poly(A)$^+$ mRNA (10 μg/ml) were formed in 70% formamide 0.3 M NaCl, 20 mM Tris-HCl (pH 8.5) and 1 mM EDTA. The sample containing clone λKBIa[1] was kept for 17 h at 62°C, followed by incubation at 64°C for 2–4 h. Clones λKBIb[1] and λKBIII[1] were incubated at 61°C for 17 h, and clone λKBVI[1] was incubated at 60°C for the same time. The hybrids were prepared for electron microscopy as described (Herrmann *et al.*, 1978). Single-stranded fd DNA (6408 bp) and double-stranded pUC8 DNA (2715 bp)

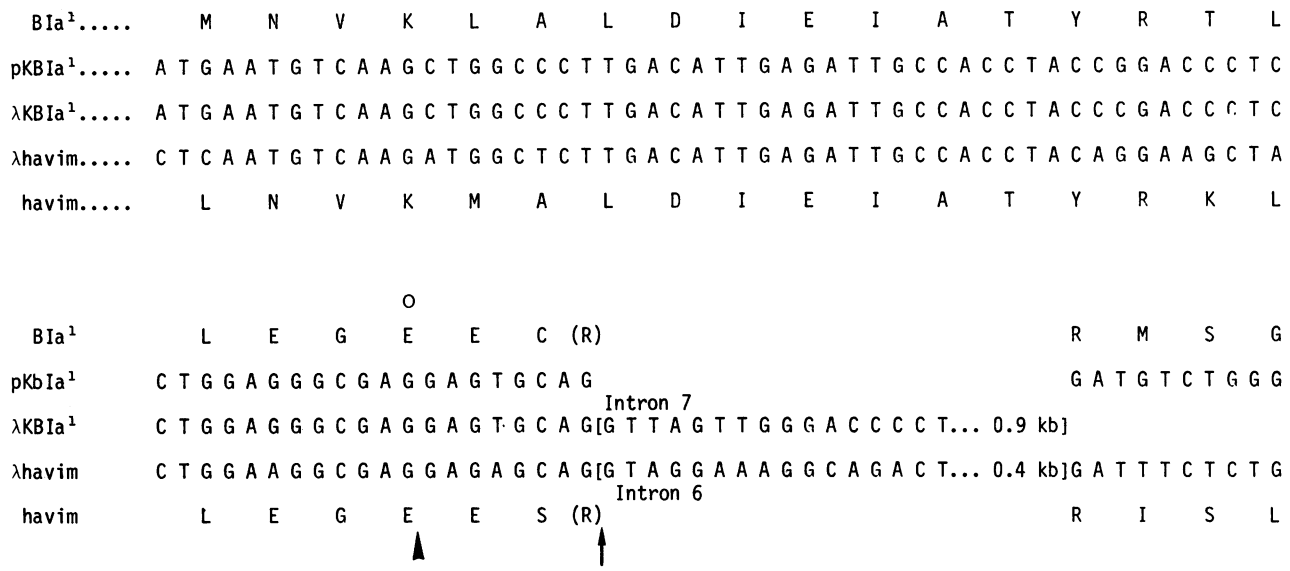

Fig. 6. Comparison of amino acid and nucleotide sequences at the boundaries between exon G and intron 7 of the gene coding for bovine keratin Ia and between exon F and intron 6 of the gene for hamster vimentin (data from Quax *et al.*, 1983). The upper line (BIa¹) presents the amino acid sequence of bovine keratin Ia as determined from a cDNA clone (cf. Jorcano *et al.*, 1984c); the second line gives the corresponding nucleotide sequence of this cDNA clone (pKBIa¹); the third line presents the sequence of the corresponding genomic clone (λKBIa¹); the fourth line shows the nucleotide sequence of a genomic clone of hamster vimentin (λhavim); the bottom line presents the corresponding vimentin amino acid sequence. The exon-intron boundary is denoted by an arrow; the end of the TYR(T,K)LLEGE consensus sequence at the carboxy-terminal end of the α-helical core ('rod') region of all intermediate filament proteins is denoted by an arrowhead and is designated sequence position '0' as in previous papers (Jorcano *et al.*, 1984b, 1984c). Note that intron 7 of the bovine keratin gene and intron 6 of the hamster vimentin gene are located exactly at the same position, i.e., disrupting an arginine coding triplet. Note also that both introns are of grossly different sizes. The sequence of both intermediate filament proteins and their corresponding genes is highly similar in this region of the α-helical core of the proteins (shown here until position −20 of the amino acid sequence) whereas the homology is lower in the subsequent intron and the extra-helical portion of the protein (for detailed comparisons of amino acid sequences see Geisler and Weber, 1981, 1982; Hanukoglu and Fuchs, 1983; Quax *et al.*, 1983; Steinert *et al.*, 1983; Weber and Geisler, 1984).

were included in the spreading solution to serve as internal length standards. Micrographs were taken with a Zeiss EM10 electron microscope. Length measurements and statistical evaluations were performed as described (Scheer and Zentgraf, 1978) and >30 individual heteroduplex molecules were evaluated per clone.

S1 nuclease mapping

The positions of several introns in the 3′ half of the keratin genes were determined by S1 nuclease mapping of heteroduplex molecules using appropriate restriction fragments. As an example we describe the mapping of the position of intron 7 in the gene coding for keratin Ia (see Figures 2 and 3), using the procedure of Berk and Sharp (1977) as modified by Weaver and Weismann (1979). 5 μg DNA from clone λKBIa¹ were digested with *Hind*III and labelled at the 3′ ends using the Klenow fragment of *Escherichia coli* DNA polymerase (Maniatis *et al.*, 1982). The digested DNA was electrophoresed on a 0.8% low-melting agarose gel and a gel slice containing a fragment of 3.2 kb was melted at 56°C, phenol extracted, and the DNA was precipitated with ethanol. The purified DNA was digested with *Sac*II and a fragment of ~2 kb was isolated from a low melting-temperature agarose gel. The labeled *Hind*III end of this fragment maps inside exon G. Aliquots of the labelled DNA (6000 c.p.m.) were hybridized with 2 μg of cow muzzle poly(A)$^+$ mRNA for 12 h at 56°, 58° and 60°C, respectively, in 10 μl of 80% formamide, 0.4 M NaCl, 40 mM Pipes, 1 mM EDTA, pH 6.4 (hybridization buffer). At the end of the hybridization, they were quickly diluted with 250 μl of chilled 'S1 buffer' [0.25 M NaCl, 0.03 M sodium acetate buffer (pH 4.6), 1 mM $ZnSO_4$, 10 μg/ml denatured salmon sperm DNA], containing 100 units of S1 nuclease (Sigma). After a 45 min incubation at 30°C, the samples were extracted with phenol:chloroform (1:1, v:v) and precipitated with ethanol in the presence of 10 μg carrier tRNA. The pellets were re-dissolved in 6 μl of 90% formamide, 1 mM EDTA, heated to 90°C for 2 min, chilled on ice, and 2 μl were loaded on a DNA sequencing gel.

DNA sequencing

Fragments obtained by restriction nuclease digestion were sequenced by the chemical cleavage method of Maxam and Gilbert (1980).

Acknowledgements

We are indebted to our colleagues S. Ruppert, G. Schütz, W. Keller and D. Bohman for valuable discussions, to T. Bock, S. Schreiber and M. Schrenk for expert technical assistance, as well as to R. Quinlan and P. Cowin for reading and correcting the manuscript. The work has been supported in parts by the German Ministry for Research and Technology (ZMBH-Project) and by the Deutsche Forschungsgemeinschaft (DFG, grant to W.W.F.).

References

Artymiuk,P.J., Blake,C.C.F. and Sippel,A.E. (1981) *Nature*, **290**, 287-288.
Bennett,G.S., Fellini,S.A., Croop,J.M., Otto,J.J., Bryan,J. and Holtzer,H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4364-4368.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Crewther,W.G., Dowling,L.M., Steinert,P.M. and Parry,D.A. (1983) *Int. J. Biol. Macromol.*, **5**, 267-274.
Dahl,D. and Bignami,A. (1983) in Lajtha,A. (ed.), *Handbook of Neurochemistry*, Vol. **5**, Plenum Publishing Corporation, pp. 127-151.
Franke,W.W., Schmid,E., Osborn,M. and Weber,K. (1978a) *Proc. Natl. Acad. Sci. USA*, **75**, 5034-5038.
Franke,W.W., Weber,K., Osborn,M., Schmid,E. and Freudenstein,C. (1978b) *Exp. Cell Res.*, **116**, 429-445.
Franke,W.W., Appelhans,B., Schmid,E., Freudenstein,C., Osborn,M. and Weber,K. (1979a) *Differentiation*, **15**, 7-25.
Franke,W.W., Schmid,E., Weber,K. and Osborn,M. (1979b) *Exp. Cell Res.*, **118**, 95-109.
Franke,W.W., Schmid,E., Winter,S., Osborn,M. and Weber,K. (1979c) *Exp. Cell Res.*, **123**, 25-46.
Franke,W.W., Schiller,D.L., Moll,R., Winter,S., Schmid,E., Engelbrecht,I., Denk,H., Krepler,R. and Platzer,B. (1981) *J. Mol. Biol.*, **153**, 933-959.
Franke,W.W., Schiller,D.L., Hatzfeld,M. and Winter,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7113-7117.
Franke,W.W., Schiller,D.L., Hatzfeld,M., Magin,T.M., Jorcano,J.L., Mittnacht,S., Schmid,E., Cohlberg,J.A. and Quinlan,R.A. (1984) *Cancer Cells 1, The Transformed Phenotype*, published by Cold Spring Harbor Laboratory Press, NY, pp. 177-190.
Fuchs,E.V., Coppock,S.M., Green,H. and Cleveland,D.W. (1981) *Cell*, **27**, 75-84.
Geisler,N. and Weber,K. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4120-4123.
Geisler,N. and Weber,K. (1982) *EMBO J.*, **1**, 1649-1656.
Geisler,N., Kaufmann,E., Fischer,S., Plessmann,U. and Weber,K. (1983)

EMBO J., **2**, 1295-1302.
Gigi,O., Geiger,B., Eshhar,Z., Moll,R., Schmid,E., Winter,S., Schiller,D.L. and Franke,W.W. (1982) *EMBO J.*, **1**, 1429-1437.
Gilbert,W. (1978) *Nature*, **271**, 501.
Gruen,L.C. and Woods,E.F. (1983) *Biochem. J.*, **209**, 587-595.
Gwo-Shu Lee,M., Lewis,S.A., Wilde,C.D. and Cowan,N.J. (1983) *Cell*, **33**, 477-487.
Hamada,H., Petrino,M.G. and Kakanuga,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5901-5905.
Hanukoglu,I. and Fuchs,E. (1982) *Cell*, **31**, 243-252.
Hanukoglu,I. and Fuchs,E. (1983) *Cell*, **33**, 915-924.
Herrmann,R., Neugebauer,K., Zentgraf,H. and Schaller,H. (1978) *Mol. Gen. Genet.*, **159**, 171-178.
Hoffmann,W. and Franz,J.K. (1984) *EMBO J.*, **3**, 1301-1306.
Hynes,R.O. and Destree,A.T. (1978) *Cell*, **13**, 151-163.
Inana,G., Piatigorsky,J., Norman,B., Slingsby,C. and Blundell,T. (1983) *Nature*, **302**, 310-315.
Jorcano,J.L., Magin,T.M. and Franke,W.W. (1984a) *J. Mol. Biol.*, **176**, 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*, in press.
Kim,K.H., Rheinwald,J.G. and Fuchs,E. (1983) *Mol. Cell Biol.*, **3**, 495-502.
Kreis,T.E., Geiger,B., Schmid,E., Jorcano,J.L. and Franke,W.W. (1983) *Cell*, **32**, 1125-1137.
Lazarides,E. (1982) *Annu. Rev. Biochem.*, **51**, 219-250.
Lazarides,E. and Hubbard,B.D. (1976) *Proc. Natl. Acad. Sci. USA*, **76**, 4344-4348.
Liem,R.K.H., Yen,S.-H., Salomon,G.D. and Shelanski,M.L. (1978) *J. Cell Biol.*, **79**, 637-645.
Magin,T.M., Jorcano,J.L. and Franke,W.W. (1983) *EMBO J.*, **2**, 1387-1392.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Moll,R., Franke,W.W., Schiller,D.L., Geiger,B. and Krepler,R. (1982) *Cell*, **31**, 11-24.
Nudel,U., Katcoff,D., Zakut,R., Shani,M., Carmon,Y., Finer,M., Czosnek, H., Ginsburg,I. and Yaffe,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2763-2767.
Osborn,M. and Weber,K. (1983) *Lab. Invest.*, **48**, 372-394.
Osborn,M., Franke,W.W. and Weber,K. (1980) *Exp. Cell Res.*, **125**, 37-46.
Quax,W.J., Egberts,W.V., Hendriks,W., Quax-Jeuken,Y.E.F.M. and Bloemendal,H. (1983) *Cell*, **35**, 215-223.
Quax-Jeuken,Y.E.F.M., Quax,W.J. and Bloemendal,H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3548-3552.
Quinlan,R.A. and Franke,W.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3452-3456.
Quinlan,R.A. and Franke,W.W. (1983) *Eur. J. Biochem.*, **132**, 477-484.
Quinlan,R.A., Cohlberg,J.A., Schiller,D.L., Hatzfeld,M. and Franke,W.W. (1984) *J. Mol. Biol.*, **178**, 365-388.
Rigby,P.W.J., Dieckmann,M., Rodes,C. and Berg, P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Roop,D.R., Hawley-Nelson,P., Cheng,C.K. and Yuspa,S.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 716-720.
Ruppert,S., Scherer,G. and Schütz,G. (1984) *Nature*, **308**, 554-557.
Scheer,U. and Zentgraf,H. (1978) *Chromosoma*, **69**, 243-254.
Schiller,D.L., Franke,W.W. and Geiger,B. (1982) *EMBO J.*, **1**, 761-769.
Schmid,E., Osborn,M., Rungger-Brändle,E., Gabbiani,G., Weber,K. and Franke,W.W. (1982) *Exp. Cell Res.*, **137**, 329-340.
Sharp,G., Osborn,M. and Weber,K. (1982) *Exp. Cell Res.*, **141**, 385-394.
Small,J.V. and Sobieszek,A. (1977) *J. Cell Sci.*, **23**, 243-268.
Southern,E. (1975) *J. Mol. Biol.*, **98**, 503-517.
Steinert,P.M., Idler,W.W., Cabral,F., Gottesman,M.M. and Goldman,R.D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3692-3696.
Steinert,P.M., Rice,R.H., Roop,D.R., Trus,B.L. and Steven,A.C. (1983) *Nature*, **302**, 794-800.
Sun,T.-T. and Green,H. (1978) *Cell*, **14**, 469-476.
Sun,T.-T., Shih,C. and Green,H. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2813-2817.
Tseng,S.C.G., Jarvinen,M.J., Nelson,W.G., Huang,J.-W., Woodcock-Mitchell,J. and Sun,T.-T. (1982) *Cell*, **30**, 361-372.
Ueyama,H., Hamada,H., Battula,N. and Kakunaga,T. (1984) *Mol. Cell. Biol.*, **4**, 1073-1078.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.
Weber,K. and Geisler,N. (1984) *Cancer Cells 1, The Transformed Phenotype*, published by Cold Spring Harbor Laboratory Press, NY, pp. 153-159.
Yamada,Y., Avvedimento,V.E., Mudryj,M., Ohkubo,H., Vogeli,G., Irani, M., Pastan,I. and de Crombrugghe,B. (1980) *Cell*, **22**, 887-892.

Received on 15 October 1984