The EMBO Journal vol.4 no.9 pp.2261 – 2267, 1985

Complete sequence of a bovine type I cytokeratin gene: conserved and variable intron positions in genes of polypeptides of the same cytokeratin subfamily

Michael Rieger[1], Jose L.Jorcano[1,2] and Werner W.Franke[1]

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

Communicated by Werner W.Franke

The complete sequence of a bovine gene encoding an epidermal cytokeratin of mol. wt. 54 500 (No VIb) of the acidic (type I) subfamily is presented, including an extended 5′ upstream region. The gene (4377 bp, seven introns) which codes for a representative of the glycine-rich subtype of cytokeratins of this subfamily, is compared with genes coding for: (i) another subtype of type I cytokeratin; (ii) a basic (type II) cytokeratin gene; and (iii) vimentin, a representative of another intermediate filament (IF) protein class. The positions of the five introns located within the highly homologous α-helix-rich rod domain are identical or equivalent, i.e., within the same triplet, in the two cytokeratin I genes. Four of these intron positions are also identical with intron sites in the vimentin gene, and three of these intron positions are identical or similar in the type I and type II cytokeratin subfamilies. On the other hand, the gene organization of both type I cytokeratins differs from that of the type II cytokeratin in the rod region in five intron positions and in the introns located in the carboxy-terminal tail region, with the exception of one position at the rod-tail junction. Remarkably, the two type I cytokeratins also differ from each other in the positions of two introns located at and in the region coding for the hypervariable, carboxy-terminal portion. The introns and the 5′ upstream regions of the cytokeratin VIb gene do not display notable sequence homologies with the other IF protein genes, but sequences identical with – or very similar to – certain viral and immunoglobulin enhancers have been identified. The results show that two major domains in IF proteins, the α-helix-rich rod and the non-α-helical, carboxy-terminal tail differ not only in the degree of amino acid sequence conservation but also in the exon patterns of their genes and that different intron positions can occur even in cytokeratin genes of the same subfamily. The findings suggest that the two domains differ in their evolutionary origin and stability, and that the two cytokeratin subfamilies as well as different subtypes of the same subfamily have diverged in gene organization relatively early in evolution.

Key words: intermediate filaments/keratins/multigene families/gene structure/regulatory elements

Introduction

Most cytoskeletal proteins occur as multigene families of polypeptides which display high structural homology and are differentially expressed in a manner related to cell differentiation. A particularly striking example is provided by the proteins which form the intermediate-sized filaments (IF) which usually are divided into five major classes: vimentin IF typical of mesenchymally derived cells; desmin IF of myogenic cells; glial filaments typical of certain glial cells; neurofilaments of neurons; and cytokeratin IF which contain proteins related to α-keratins of epidermis and its appendages and characteristic of epithelial cells (for reviews, see Franke *et al.*, 1982; Lazarides, 1982; Weber and Geisler, 1984).

The cytokeratins are a particularly complex class of IF proteins. For example, human cytokeratins comprise at least 19 different polypeptides which are expressed in different combinations of 2 – 11 members in different cell types (Moll *et al.*, 1982; Fuchs *et al.*, 1984; Sun *et al.*, 1984). Similar complexities and principles of cell type-specific expression have been described in the cytokeratins of various animal species, notably rodents and cow (Franke *et al.*, 1981a, 1981b, 1982; Schiller *et al.*, 1982; Tseng *et al.*, 1982). The cytokeratins can be divided into two large subfamilies, i.e., the acidic type I cytokeratins and the more basic type II cytokeratins, which share only limited sequence homology (<30%) which is restricted to the α-helical 'rod' domain (Fuchs *et al.*, 1981, 1984; Crewther *et al.*, 1983; Hanukoglu and Fuchs, 1983; Sun *et al.*, 1984; Weber and Geisler, 1984). This subdivision is of functional significance since at least one representative of either subfamily is required for the formation of the heterotypic subunit complexes of cytokeratin IF (Moll *et al.* 1982; Crewther *et al.*, 1983; Franke *et al.*, 1983; Hanukoglu and Fuchs, 1983; Fuchs *et al.* 1984; Jorcano *et al.*, 1984b; Quinlan *et al.*, 1984; Steinert *et al.*, 1984; Sun *et al.*, 1984). In addition, however, pronounced differences can be observed even between members of the same cytokeratin subfamily. For example, different type I cytokeratins expressed in the same tissue of the same species can differ greatly in their non-α-helical carboxy-terminal region ('tail'; for nomenclature see Geisler and Weber, 1982; Weber and Geisler, 1984).

Recent analyses of cloned cytokeratin genes (Lehnert *et al.*, 1984; Marchuck *et al.*, 1984, 1985) have shown marked conservation of exon-intron patterns not only between different cytokeratin genes but also between cytokeratin genes and the gene encoding vimentin (Quax *et al.*, 1983). Here we present the complete nucleotide sequence of the bovine gene coding for a type I cytokeratin of mol. wt. 54 500 ('keratin VIb'; Franke *et al.*, 1981a, 1981b) and its adjacent regions. Comparison of this gene with the recently published sequences of a human cytokeratin of the same subfamily (No. 14 of mol. wt. 50 000; Marchuk *et al.*, 1984, 1985) and of a type II cytokeratin gene (Johnson *et al.*, 1985) has revealed principles of gene organization common to IF proteins as well as profound differences, even between cytokeratins of the same subfamily. Our results indicate that introns are more conserved in the gene portion coding for the conserved α-helical 'rod' domain whereas intron positions corresponding to the non-α-helical, hypervariable carboxy-terminal tail vary considerably.

© IRL Press Limited, Oxford, England.

Results

Determination of number of gene copies

When restriction fragments of calf thymus DNA were analyzed by blot-hybridization with the cDNA clone pKBVIb[1], only one strongly hybridizing band was observed on digestion with *Pst*I, which does not cleave within the gene, and with *Eco*RI, which cleaves once within the gene but outside of the region represented by the cDNA probe (Figure 1, arrowheads). Some minor bands, which appeared faint under stringent hybridization conditions (Figure 1, arrows), showed somewhat increased hybridization at standard conditions. The same minor bands were also observed when bovine DNA was similarly probed with cDNAs coding for other type I cytokeratins (data not shown), indicating that they were due to cross-hybridization with genes of other members of this subfamily, as has been described for human (Fuchs *et al.*, 1981) and murine (Roop *et al.*, 1983) epidermal keratins.

A strongly hybridizing band of the same mol. wt. and hybridization intensity as that shown in Figure 1 for calf thymus DNA (a diploid tissue not expressing cytokeratin VIb) was also observed for DNA from bull sperm (haploid cells not expressing any cytokeratins) and from calf muzzle epidermis (a tissue expressing large amounts of cytokeratin VIb; Franke *et al.*, 1981a; Schiller *et al.*, 1982; Jorcano *et al.*, 1984a, 1984b). This indicates that the haploid bovine genome contains only one gene copy for cytokeratin VIb and that no genomic re-arrangement takes place during cell differentiation. We have also found no indication for the existence of pseudogenes as have been described for murine cytokeratin A (Vasseur *et al.*, 1985).

Nucleotide sequence of the gene and amino acid sequence deduced therefrom

The complete nucleotide sequence of the gene is shown in Figure 2, together with the predicted amino acid sequence of the polypeptide. Intron positions were determined by comparisons with the sequences of the corresponding cDNA (Jorcano *et al.*, 1984b) and the highly homologous murine cytokeratin of mol. wt. 59 000 (Steinert *et al.*, 1983) as well as by S1 nuclease mapping as previously described (Lehnert *et al.*, 1984). Where comparable, the genomic sequence was found to be identical to the partial cDNA sequence published previously (Jorcano *et al.*, 1984b), with the exception of the region of nucleotides 4490–4600 in which the cDNA sequence published (Jorcano *et al.*, 1984b) contained some sequencing errors due to the high degree of DNA secondary structure in this region, which is refractory to the elongation reaction at the lower temperature used at that time (see Materials and methods). The corrected amino acid sequence is shown in Figure 3.

The gene contains seven introns, and the lengths of the exons and introns, as determined by sequencing, are generally in fair agreement with previous data from electron microscopic analyses of heteroduplex molecules (Lehnert *et al.*, 1984). The sizes of the individual exons are as follows: A (from cap site), 595 bp; B, 83 bp; C, 157 bp; D, 162 bp; E, 126 bp; F, 218 bp; G, 258 bp; H (up to polyadenylation signal), 354 bp. Exons B–G, which are included by introns, display a relatively narrow size distribution with an average of 167 ± 57 (S.D.) bp, an exon size range typical for many eukaryotic genes (Naora *et al.*, 1982). In contrast, the introns vary greatly in size, from 83 bp (intron 3) to 798 bp (intron 1).

The total gene comprises 4377 bp from the cap site to the minimal 3′ end, i.e., 18 bp 3′ downstream from the polyadenylation signal, and the corresponding mRNA region contains 1971 nucleotides without the poly(A) tail. The latter value corresponds well to the 2100 nucleotides estimated from Northern blots for the poly(A)-containing mRNA (Jorcano *et al.*, 1984a). The amino acid coding region includes, without the codon for the initial methionine, 1575 bp and thus defines a polypeptide of 525 amino acids and a total mol. wt. value of 54 712 which is in excellent agreement with the value of 54 500 previously estimated from SDS-polyacrylamide gel electrophoresis (Franke *et al.*, 1978, 1981a; Schiller *et al.*, 1982).

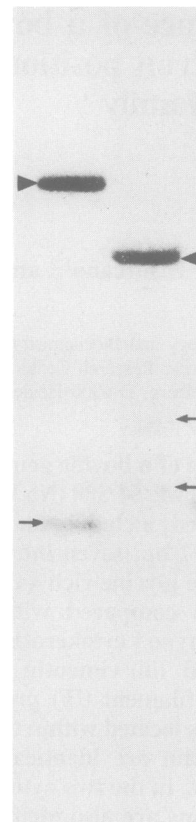

Fig. 1. Distribution of sequences coding for cytokeratin VIb in the bovine genome. Calf thymus DNA was digested with *Pst*I (**left**) and *Eco*RI (**right**), separated on 0.7% agarose gels, blotted on nitrocellulose and hybridized under stringent conditions to nick-translated cDNA clone pKBVIb[1]. Note the single strong hybridizing bands (arrowheads), indicating that this gene is represented only once in the bovine genome. The faint bands (arrows) appear to be due to cross-hybridization with sequences coding for other cytokeratin genes of the same subfamily.

When the amino acid sequence of this bovine type I cytokeratin was compared with the other two type I cytokeratin sequences available (Figure 3), the high homology with the murine cytokeratin of mol. wt. 59 000 (Steinert *et al.*, 1983) was striking. This supports our conclusion (Jorcano *et al.*, 1984b) that these polypeptides represent equivalent cytokeratins in the two species, both belonging to the subtype characterized by a very high glycine content and $^{S}_{G}GG^{S}_{G}YGG$ repeats in their carboxy-terminal portion and an identical terminal heptapeptide. While both the bovine and the murine proteins showed considerable homology to the human cytokeratin No. 14 of mol. wt. 50 000 in the α-helical rod region, the head and the tail portions were clearly different, confirming our conclusion that at least two types of cytokeratins can be distinguished within this subfamily (Jorcano *et al.*, 1984b).

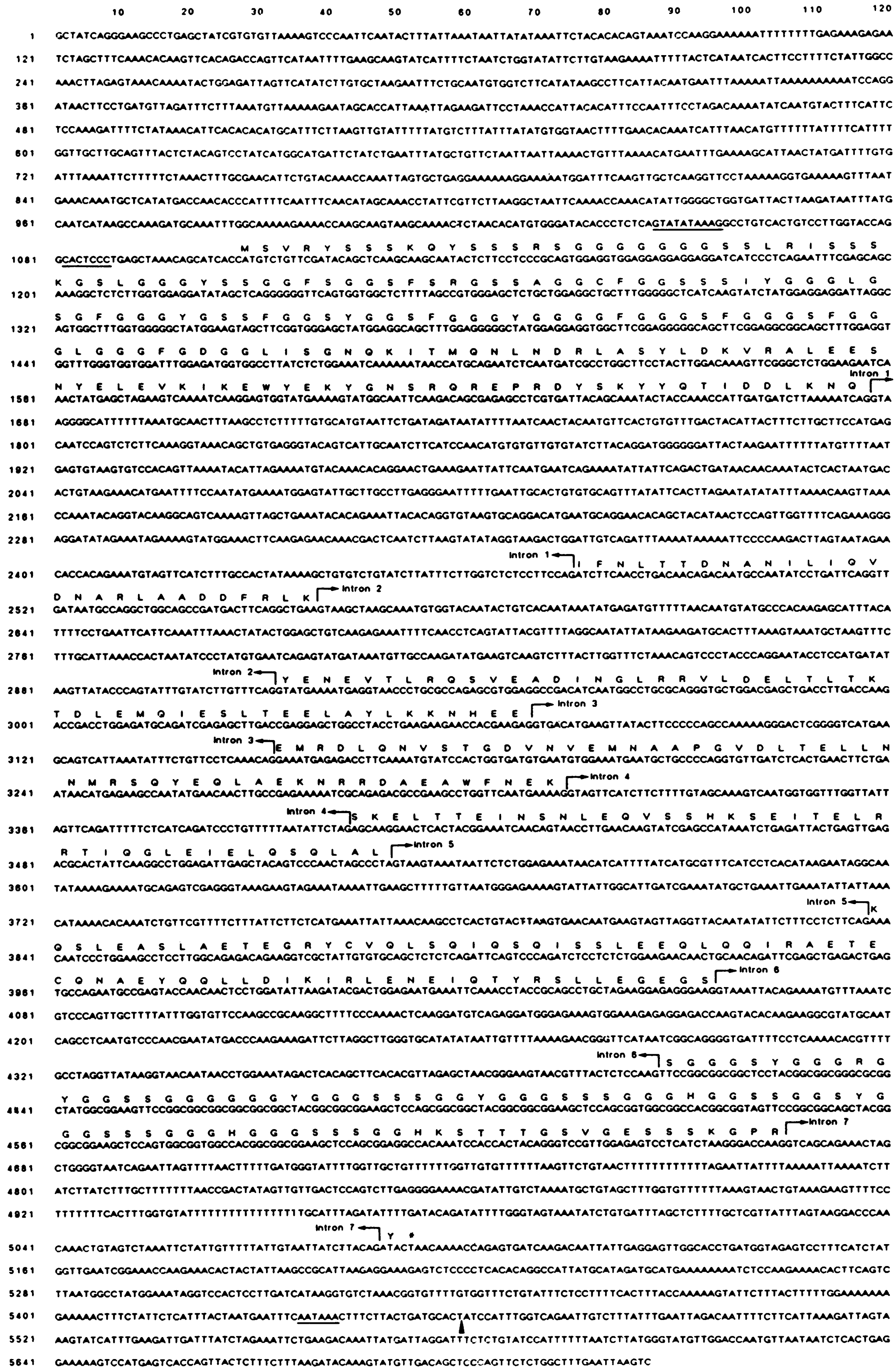

Fig. 2. Nucleotide sequence of the gene coding for cytokeratin VIb and adjacent regions, together with the amino acid sequence deduced therefrom. Nucleotides are numbered from 1 to 5726. The amino acid sequences encoded by the exons are indicated above the DNA sequence (one letter code). The demarcations of the introns are indicated by arrows pointing toward the center of the intron. TATA-box, cap-site sequence and polyadenylation signal are underlined. Stop codon is denoted by asterisk; the arrowhead denotes the end of the corresponding cDNA clone.

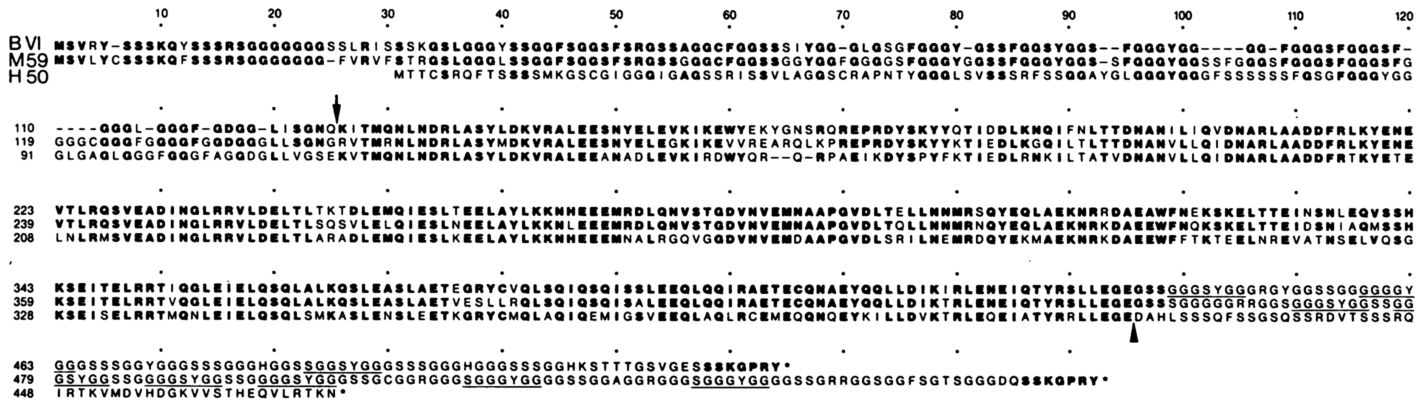

Fig. 3. Comparison of the amino acid sequences of three representatives of the acidic (type I) subfamily of cytokeratins, as deduced from the corresponding nucleotide sequences (the initial methionine residue is included): BVI, bovine cytokeratin VIb of mol. wt. 54 500; M59, murine cytokeratin of mol. wt. 59 000 (data from Steinert *et al.*, 1983); H50, human cytokeratin No. 14 of mol. wt. 50 000 (data from Marchuk *et al.*, 1984, 1985). Sequences are aligned to obtain maximal homology; deletions introduced for this purpose are denoted by horizontal bars. The canonical repeats of glycine-rich heptapeptides in BVI and M59 are underlined. Amino acids which the bovine cytokeratin has in common with one of the other two cytokeratins are in boldface type. No homology arrangements have been made for most of the carboxy-terminal portion as the glycine-richness of this region would result in the appearance of fortuitous homologies. Arrow denotes the start of the α-helix-rich rod domain, arrowhead demarcates the end of the helical region. Asterisk, stop codon.

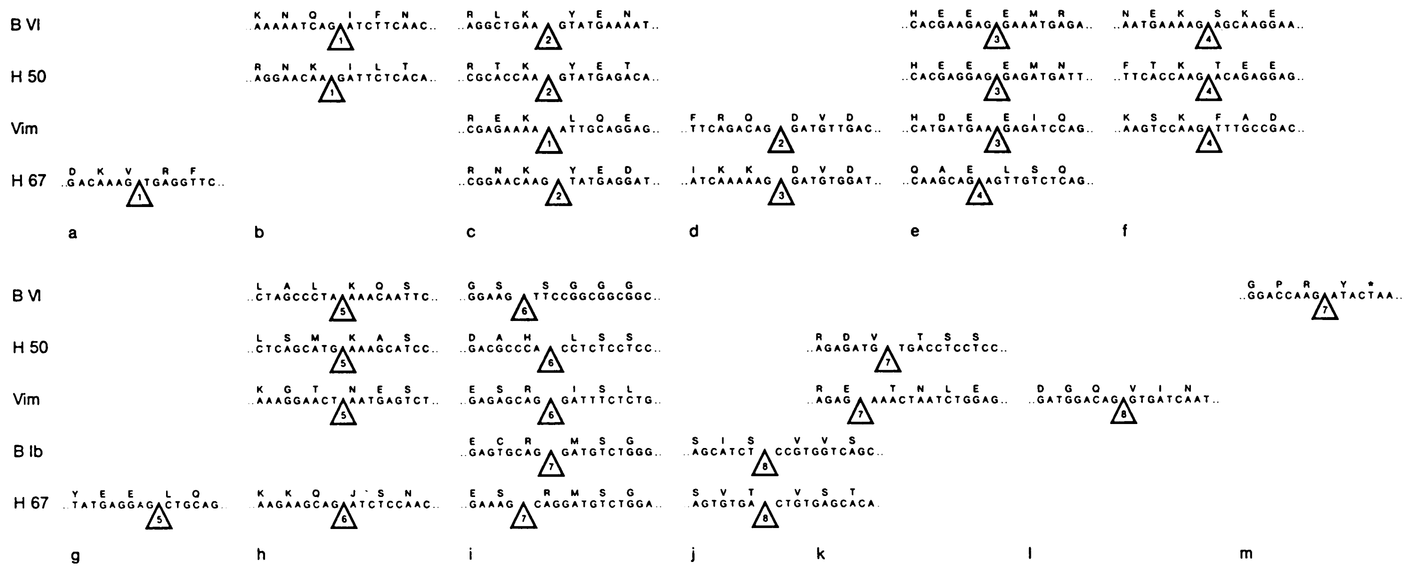

Fig. 4. Positions of the introns in different IF protein genes. The nucleotide and amino acids sequences surrounding the introns (insertion site denoted by triangle) are shown. The specific numbers of the introns are given in the triangles. Alignment and symbols as in Figure 3. Vim, hamster vimentin (Quax *et al.*, 1983); BIa, partial sequence of bovine epidermal cytokeratin Ib; H67, human cytokeratin No. 1 mol. wt. 67 000 (Johnson *et al.*, 1985).

Comparison of intron positions

Comparison of the nucleotide sequences of the bovine gene BVI with the human type I gene coding for cytokeratin No. 14 (gene 'H50'; Marchuk *et al.*, 1984, 1985) revealed no significant sequence homologies within introns whereas the positions of most of the introns were conserved (Figures 4 and 5). Introns 2–5 occurred at identical positions, all disrupting the same (introns 2–4) or a corresponding triplet after the second nucleotide. Intron 1 was also located in the same amino acid position but the insertion site was slipped by one nucleotide. Intron 6 was found in a similar region in both cytokeratin genes but its insertion site was displaced upstream by one triplet in the bovine gene BVI. In contrast, the bovine gene BVI did not contain an intron corresponding to intron 7 of the human gene; instead it contained a long intron (422 bp) disrupting the penultimate triplet of the carboxy terminus, a region without introns in the human gene, for cytokeratin No. 14.

Comparison of the bovine type I cytokeratin gene with the hamster gene of another IF protein, vimentin, also showed the positions of introns 2, 3, 4 and 5 of the cytokeratin gene to be identical with the introns 1, 3, 4 and 5, respectively, of vimentin. Moreover, vimentin intron 6 was found at the same position as intron 6 of the human type I cytokeratin gene and intron 7 of two bovine type II cytokeratin genes (Lehnert *et al.*, 1984). Intron 7 of the vimentin gene was located in a similar region as intron 7 of the human type I cytokeratin gene (Figure 4) but was displaced by four nucleotides. In contrast, introns 2 and 8 of the vimentin gene had no positional counterpart in the type I cytokeratin genes, and intron 1 of both type I cytokeratin genes as well as intron 7 of the bovine gene were located in regions devoid of introns in the vimentin gene.

Remarkably, the recently published sequence of a gene coding for the human type II cytokeratin (No. 1) of mol. wt. 67 000 (gene 'H67', Figure 4; Johnson *et al.*, 1985) did not show more

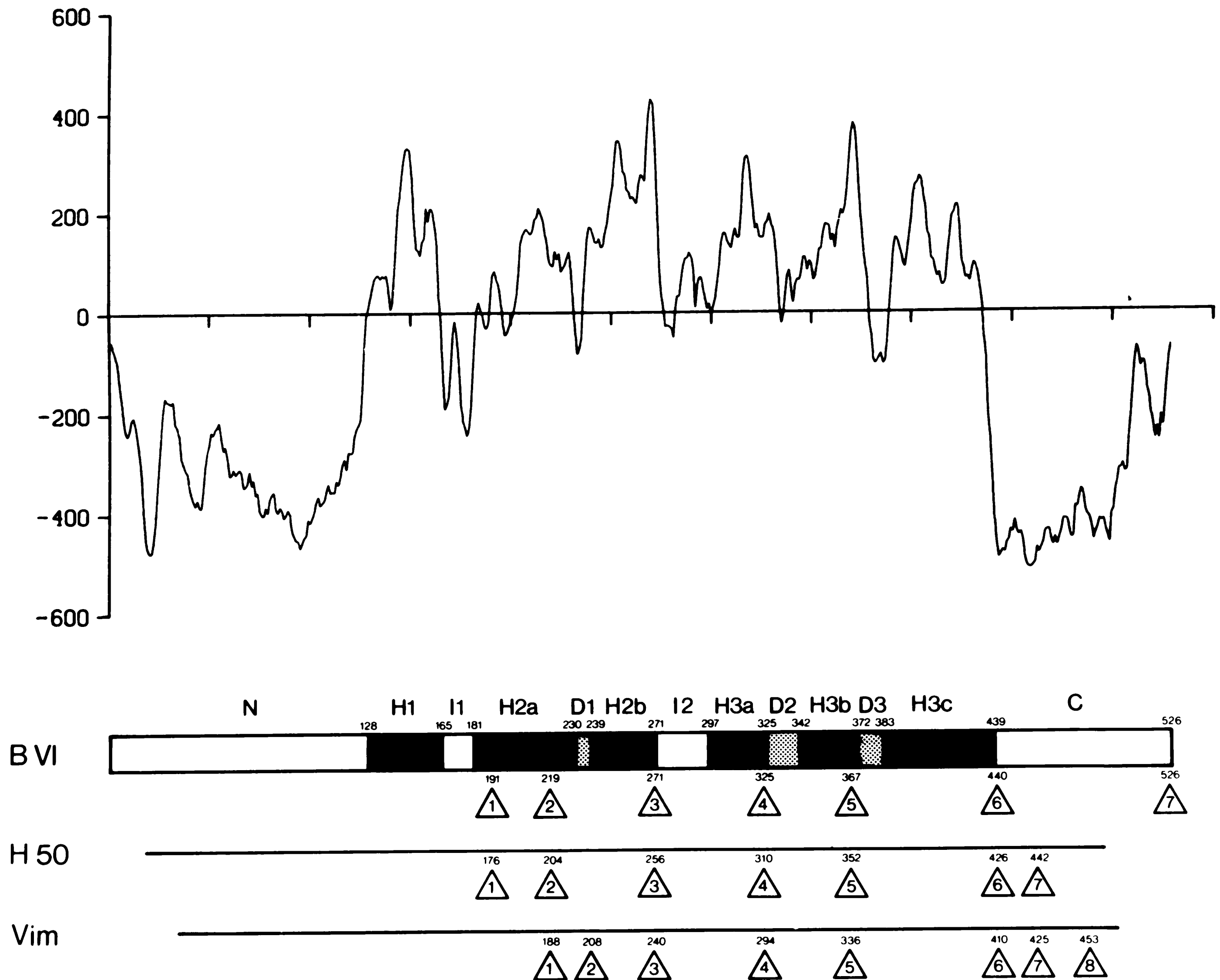

Fig. 5. Relationship between structural domains of IF proteins and intron positions of the corresponding genes. The upper part shows an α-helix probability profile for the amino acid sequence of bovine epidermal cytokeratin VIb. The program of Garnier *et al.* (1978) was used with a sliding average of eight amino acids (plotted as a function of amino acid position). The values on the ordinate present relative units of probability of α-helical conformation in information theoretical units (centinats). The polypeptide domains of bovine epidermal keratin VIb (BVI), also representing the gross domain organization of IF proteins in general, is drawn to the same scale below. Helical domains (H1, H2, H3) are shown as black boxes, open boxes represent non-α-helical regions (N,C; amino- and carboxy-terminal parts; I1, I2, helix-interrupting sequences). Dotted areas (D1 – D3) represent regions of disturbance of α-helicity. The amino acids at the boundaries of the domains are given by small numbers. The intron positions of three IF protein genes are marked by triangles, the intron numbers are given within the triangles. In the specific polypeptides the number of the amino acid following a given intron is indicated below the lines. The abbreviations for the three polypeptides shown are given in Figure 4.

similarities in its intron pattern with the type I cytokeratin genes than it did with the vimentin gene. Only one intron position (Figure 4h) was exactly identical between the two cytokeratin subfamilies, and two other introns were located in the same codon but displaced by one or two nucleotides, respectively (introns 2 and 4; Figure 4c and e). Intron 7 of the gene for human type II cytokeratin No. 1 was not in exactly the same position as in the corresponding bovine No. 1 cytokeratin gene (BIb in Figure 4i), in the hamster vimentin gene, and in the human type I cytokeratin No. 14 gene (H50) but, surprisingly, was identical in position to intron 6 in the bovine BVI gene of the other (type I) subfamily (Figure 4i). Moreover, two introns of the human type II cytokeratin gene one of which was identical in position to a hamster vimentin gene intron were located in regions in which both the human and the bovine type I cytokeratin genes were devoid of introns (Figure 4d and j). Moreover, two introns of the human type II cytokeratin gene occurred in places in which none of the other IF protein genes contained introns.

To examine the hypothesis that introns tend to occur near borders of structural or functional domains of proteins (Gilbert, 1978; Doolittle, 1978; Lonberg and Gilbert, 1985; Stone *et al.*, 1985) we compared the exon patterns of the IF protein genes with the patterns of α-helical domains of the proteins as predicted by the program of Garnier *et al.* (1978). Figure 5 presents the comparison of the genes coding for the three negatively charged IF proteins. All three genes revealed a pronounced correspondence of intron positions to borders of structural domains or turns of conformational character. In the two type I cytokeratin genes, introns 1, 3, 4, 5 and 6 could be correlated to slopes of regions of high α-helicity. Remarkably, four of these five in-

tron positions were also found in the vimentin gene. In addition, intron 7 of the bovine type I cytokeratin gene (BVI) was obviously related to the carboxy terminus (Figures 4m and 5). No domain could so far be correlated with the position common to intron 2 of the two type I cytokeratin genes and intron 1 of the vimentin gene. Comparison of the human type II cytokeratin gene (H67) with predicted secondary structure (Johnson *et al.*, 1985) has also shown a relationship of some introns to regions of changes of α-helix probability whereas other introns could not be related to distinct changes of structure. Interestingly, three of the four intron positions maintained in all IF protein genes sequenced so far can be associated with edges of α-helical structure profiles (positions aligned in Figure 4e, h and i).

Possible signals and regulatory sequences

The only polyadenylation signal found (Figure 2) was the one already identified in the cDNA sequence (Jorcano *et al.*, 1984b). A typical TATA-box (Breathnach and Chambon, 1981) was recognized at position −35 upstream from the cap site (Figure 2). Searching 1046 bp upstream from the TATA-box and for possible regulatory signals we found two sequences with some homology (six out of eight nucleotides) to the 'enhancer consensus sequence' (Weiher *et al.* 1983) which were located at positions −19 and −497 from the TATA-box (for the human type I cytokeratin gene H50 see also Marchuk *et al.*, 1985). However, the functional significance of these sequences is unproven, and it should be noted that several similar sequences occur in the transcribed gene region, including two complete enhancer-type sequences in introns 4 and 6, respectively. At position −67 from the TATA-box we noted the sequence AAATTTGGC reported to occur in a similar position in all estrogen-regulated genes of chicken studied so far (cf. Renkawitz *et al.*, 1984). We also compared the sequence of the epidermis-specific bovine cytokeratin gene BVI with that of bovine papilloma virus (Chen *et al.*, 1982) and have found a surprisingly high number of homologies the functional significance of which is currently under study.

Discussion

The sequence data presented allow the comparison of gene organization of different major classes of IF proteins such as type I and type II cytokeratin subfamilies and the non-epithelial IF proteins which are collectively termed 'type III' IF proteins by some authors (Johnson *et al.*, 1985). They also allow, for the first time, comparison of two genes of the same subfamily, i.e., the type I cytokeratins. Our results confirm and extend the concept that different cytokeratins and IF proteins in general form not only the same structure and share principles of arrangement of protein domains (Geisler and Weber, 1981, 1982; Quax *et al.*, 1983; Hanukoglu and Fuchs, 1982, 1983; Crewther *et al.*, 1983; Steinert *et al.*, 1983; Weber and Geisler, 1984) but also display common exon-intron arrangements of their genes (Lehnert *et al.*, 1984; Marchuk *et al.*, 1984; Johnson *et al.*, 1985). In the two type I cytokeratin genes compared five out of seven introns are in the same codon positions (one of them shifted by one nucleotide) and another intron (no. 6) is displaced by only one codon. Moreover, four identical intron positions of the type I cytokeratins are also present in the hamster vimentin gene (Quax *et al.*, 1983), and three are found identical or with one or two nucleotides moved over in the human type II cytokeratin gene (Johnson *et al.*, 1985). Another intron position is shared by vimentin and the type II cytokeratin gene but is lacking in both cytokeratin I genes. Apparently, the genes coding for IF proteins represent a large family with some precisely conserved exon-intron patterns. Similar observations have been made in other multigene families, including cytoskeletal filament proteins such as actins, myosins and tubulins (e.g., Breathnach and Chambon, 1981; Hamada *et al.*, 1982; Nudel *et al.*, 1982; Gwo-Shu Lee *et al.*, 1983; Strehler *et al.*, 1985).

It is obvious that the conserved intron positions in the various IF protein genes examined so far are located within the α-helical rod, i.e., in the part of the protein which is characterized by considerable sequence homology (Geisler and Weber, 1982; Crewther *et al.*, 1983; Hanukoglu and Fuchs, 1983; Quax *et al.*, 1983; Jorcano *et al.*, 1984b; Steinert *et al.*, 1984; Weber and Geisler, 1984). Our finding that most of these positionally conserved introns in IF protein genes are located in regions near edges of predicted α-helical subdomains adds support to the debated concept of intron involvement in the assembly of proteins during evolution (Lonberg and Gilbert, 1985; Doolittle, 1978; Gilbert, 1978; Stone *et al.*, 1985). The reason why such a correlation has been noted for only one or two (Quax *et al.*, 1983; Johnson *et al.*, 1985) or none (Marchuck *et al.*, 1984) of the introns in analyses of other authors may reflect differences of the specific secondary structure prediction programs used.

The conserved positions of the introns in the α-helix-rich rod domain is in striking contrast to the exon patterns of the carboxy-terminal tail region which is widely regarded as hypervariable in amino acid sequence (Geisler and Weber, 1981, 1982; Quax *et al.*, 1983; Fuchs *et al.*, 1984; Jorcano *et al.*, 1984b; Weber and Geisler, 1984). Here differences of intron positions are seen not only between two different IF protein classes such as vimentin and the cytokeratins or between the two cytokeratin subfamilies (see also Lehnert *et al.*, 1984) but also between different cytokeratins of the same subfamily, as shown by the two type I cytokeratin genes compared in this study. Clearly, the last introns are in grossly different positions in the genes of all three IF protein classes examined. Remarkably, the position of the intron which is right at the end of the α-helical-rod domain (Figure 4i) is conserved in the genes of IF proteins as diverse as vimentin (intron 6), two bovine type II cytokeratins (intron 7; see also Lehnert *et al.*, 1984) and the human type I cytokeratin No. 14 (intron 6) but is not conserved in the glycine-rich subtype of type I cytokeratins (intron 6 of bovine gene BVIb) and the human type II cytokeratin No. 1 (intron 7). Instead it is moved over in the latter by exactly one codon (see Lehnert *et al.*, 1984), pointing to an unexpected coincidence between a type I and a type II cytokeratin. We interpret these differences to indicate that the two domains, the highly conserved rod and the hypervariable tail, have undergone different evolutionary processes at the genomic level. Apparently, the gene portion corresponding to the α-helical rod domain contains regions highly conservative in intron sites, besides other regions which have lost (or gained) introns during evolution. In contrast, the tail domain may be more susceptible to intron losses, or greatly different tail-equivalent portions may have been joined to the rod domain, thus forming the IF proteins as fusion proteins of different structural domains. Moreover, the extensive differences in sequence and in exon patterns in the tail domain between two polypeptides of the same cytokeratin type I subfamily suggest that genomic divergence between these two subtypes has been established relatively early, a notion which is also supported by preliminary analyses of the corresponding genes in the frog, *Xenopus laevis* (unpublished data). It is also noteworthy that the two type I cytokeratin genes, which are both expressed in epidermis, as well as the epidermis-specific, human No. 1 cytokeratin do not exhibit obvious sequence or signal homologies which could lead to an understan-

ding of the regulation of their cell type-specific expression. Clearly, more systematic analyses of other IF protein genes, including those of taxonomically distant species, together with functional assays, are required to provide a better basis for our understanding of the evolutionary development of this multigene family.

Materials and methods

The genomic clone λKB VIb[1] (Lehnert *et al.*, 1984) was digested with the restriction enzymes *Eco*RI, *Hind*III, *Sal*I and *Pst*I and the fragments hybridizing to the cDNA clone pKB VIb[1] were subcloned into pEMBL8+ and pEMBL9+ (Dente *et al.*, 1983). From each plasmid clone a set of shortened inserts was generated following a DNase I deletion strategy (Frischauf *et al.*, 1980).

DNA sequencing was performed using the dideoxy chain termination procedure (Sanger *et al.*, 1977), with some modifications which allowed readings of up to 850 nucleotides from one reaction (M.Rieger, in preparation). In brief, end-labelled M13 sequencing primers (New England Biolabs, Beverly, MA) were hybridized either to single-stranded templates (Dente *et al.*, 1983) or to nicked plasmid DNA templates. The primer elongation reaction was performed at 27–30°C, instead of room temperature, to avoid artefacts due to secondary structures of the DNA template. The latter was found to be especially important in sequencing regions containing glycine-serine-rich repeats. The reaction products were analyzed on 60 cm and 95 cm sequencing gels. For computer analysis of the sequences the programs of Osterburg *et al.* (1982) were used.

For Southern hybridizations (Southern, 1975), 30 μg of high mol. wt. DNA from calf thymus, calf muzzle or bull sperm were digested with 150 units of *Eco*RI or *Pst*I. The restricted DNAs were separated on 0.7% agarose gels and blotted on nitrocellulose filters. The filters were hybridized with nick-translated DNA of the cDNA clone pKBVIb[1] (Jorcano *et al.*, 1984a, 1984b) in a solution containing 50% deionized formamide, 5 × SSPE (SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, 2 mM EDTA, pH 7.4), 5 × Denhardt's solution (Maniatis *et al.*, 1982), 100 μg/ml *Escherichia coli* tRNA and 0.1% SDS for 20 h at 42°C (standard conditions) or 50°C (stringent conditions). The filters were washed in 0.1 × SSPE, 0.1% SDS at 65°C for 90 min, dried and exposed to Kodak X-AR films.

Acknowledgements

The work has been supported in part by the Deutsche Forschungsgemeinschaft (DFG, grant to W.W.F.) and the Ministry for Research and Technology (ZMBH-Project).

References

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 329-383.
Chen,E.Y., Howley,P.M., Levinson,A.D. and Seeburg,P.H. (1982) *Nature*, **299**, 529-534.
Crewther,W.G., Dowling,L.M., Steinert,P.M. and Parry,D.A. (1983) *Int. J. Biol. Macromol.*, **5**, 267-274.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.
Doolittle,W.F. (1978) *Nature*, **272**, 581-582.
Franke,W.W., Weber,K., Osborn,M., Schmid,E. and Freudenstein,C. (1978) *Exp. Cell Res.*, **116**, 429-445.
Franke,W.W., Schiller,D.L., Moll,R., Winter,S., Schmid,E., Engelbrecht,I., Denk,H., Krepler,R. and Platzer,B. (1981a) *J. Mol. Biol.*, **153**, 933-959.
Franke,W.W., Winter,S., Grund,C., Schmid,E., Schiller,D.L. and Jarasch,E.-D. (1981b) *J. Cell Biol.*, **90**, 116-127.
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.
Frischauf,A.M., Garoff,H. and Lehrach,H. (1980) *Nucleic Acids Res.*, **8**, 5541-5549.
Fuchs,E.V., Coppock,S.M., Green,H. and Cleveland,D.W. (1981) *Cell*, **27**, 75-84.
Fuchs,E., Grace,M.P., Kim,K.H. and Marchuk,D. (1984) in Levine,A.J., Van de Woude,G.F., Topp,W.C. and Watson,J.D. (eds.), *Cancer Cells 1, The Transformed Phenotype*, Cold Spring Harbor Laboratory Press, NY, pp. 161-167.
Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97-120.
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.
Gilbert,W. (1978) *Nature*, **271**, 501.
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. (1983) *Cell*, **33**, 915-924.
Johnson,L.D., Idler,W.W., Zhou,X.-M., Roop,D.R. 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.*, **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.
Lazarides,E. (1982) *Annu. Rev. Biochem.*, **51**, 219-250.
Lehnert,M.E., Jorcano,J.L., Zentgraf,H., Blessing,M., Franz,J.K. and Franke,W.W. (1984) *EMBO J.*, **3**, 3279-3287.
Lonberg,N. and Gilbert,W. (1985) *Cell*, **40**, 81-90.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Marchuk,D., McCrohon,S. and Fuchs,E. (1984) *Cell*, **39**, 491-498.
Marchuk,D., McCrohon,S. and Fuchs,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1609-1613.
Moll,R., Franke,W.W., Schiller,D.L., Geiger,B. and Krepler,R. (1982) *Cell*, **31**, 11-24.
Naora,H. and Deacon,N.J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6196-6200.
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.
Osterburg,G., Glatting,K.H. and Sommer,R. (1982) *Nucleic Acids Res.*, **10**, 207-216.
Quax,W.J., Egberts,W.V., Hendriks,W., Quax-Jeuken,Y.E.F.M. and Bloemendal,H. (1983) *Cell*, **35**, 215-223.
Quinlan,R.A., Cohlberg,J.A., Schiller,D.L., Hatzfeld,M. and Franke,W.W. (1984) *J. Mol. Biol.*, **178**, 365-388.
Renkawitz,R., Schütz,G., Von der Ahe,D. and Beato,M. (1984) *Cell*, **37**, 503-510.
Roop,D.R., Hawley-Nelson,P., Cheng,C.K. and Yuspa,S.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 716-720.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acd. Sci. USA*, **74**, 5463-5467.
Schiller,D.L., Franke,W.W. and Geiger,B. (1982) *EMBO J.*, **1**, 761-769.
Southern,E. (1975) *J. Mol. Biol.*, **98**, 503-517.
Steinert,P.M., Rice,R.H., Roop,D.R., Trus,B.L. and Steven,A.C. (1983) *Nature*, **302**, 794-800.
Steinert,P.M., Parry,D.A.D., Racoosin,E.L., Idler,W.W., Steven,A.C., Trus,B.L. and Roop,D.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5709-5713.
Stone,E.M., Rothblum,K.N. and Schwartz,R.J. (1985) *Nature*, **313**, 498-500.
Strehler,E.E., Mahdavi,V., Periasamy,M. and Nadal-Ginard,B. (1985) *J. Biol. Chem.*, **260**, 468-471.
Sun,T.-T., Eichner,R., Schermer,A., Cooper,D., Nelson,W.G. and Weiss,R.A. (1984) in Levine,A.J., Van de Woude,G.F., Topp,W.C. and Watson,J.D. (eds.), *Cancer Cells 1, The Transformed Phenotype*, Cold Spring Harbor Laboratory Press, NY, pp. 169-176.
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.
Vasseur,M., Duprey,P., Brûlet,P. and Jacob,M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1155-1159.
Weber,K. and Geisler,N. (1984) in Levine,A.J., Van de Woude,G.F., Topp,W.C. and Watson,J.D. (eds.), *Cancer Cells 1, The Transformed Phenotype*, Cold Spring Harbor Laboratory Press, NY, pp. 153-159.
Weiher,H., König,M. and Gruss,P. (1983) *Science (Wash.)*, **219**, 626-631.

Received on 5 June 1985