J Mol Evol (1988) 27:203–211

Journal of Molecular Evolution

Concerted Gene Duplications in the Two Keratin Gene Families

Miroslav Blumenberg

Departments of Dermatology and Biochemistry, N.Y.U. Medical Center, 550 First Avenue, New York, New York 10016, USA

Summary. Evolutionary trees were derived from the keratin protein sequences using the Phylogeny Analysis Using Parsimony (PAUP) set of programs. Three major unexpected conclusions were derived from the analysis: The smallest keratin protein subunit, K#19 (Moll et al. 1982), is not the most primitive one, but has evolved to fulfill a highly specialized function, presumably to redress the unbalanced synthesis of keratin subunits. Second, the ancestors of keratins expressed in the early embryonic stages, K#8 and K#18, were the first to diverge from the ancestors of all the other keratins. The branches leading to these two keratins are relatively short, indicating a comparatively strong selection against changes in the sequences of these two proteins. Third, the two keratin families show extraordinary parallelism in their patterns of gene duplications. In both families the genes expressed in embryos diverged first, later bursts of gene duplications created the subfamilies expressed in various differentiated cells, and relatively recent gene duplications gave rise to the hair keratin genes and separated the basal cell-specific keratin from those expressed under hyperproliferative conditions. The parallelism of gene duplications in the two keratin gene families implies a mechanism in which duplications in one family influence duplication events in the other family.

Key words: Evolutionary trees — Gene families — Concerted gene duplications — Keratins — Trophoblasts — Intermediate filaments

Offprint requests to: M. Blumenberg

Introduction

The proteins of each of the three filamentous networks of the eukaryotic cytoplasmic cytoskeleton are encoded by gene families. There are six expressed mammalian actin genes (Vandekerckhove and Weber 1979) and five known expressed tubulin genes (Cleveland and Sullivan 1985). The third cytoskeletal component, the intermediate filaments, constitutes the most diverse and, in this respect, the most interesting family of eukaryotic cytoskeletal proteins. Within the intermediate filament proteins several families have been recognized by immunologic and nucleic acid hybridization techniques (for recent reviews see Steinert and Parry 1985; Traub 1985; Geisler and Weber 1986). The nonepithelial intermediate filament proteins constitute a separate family (Weber and Geisler 1985), whereas the keratins belong to two families, acidic or type I and basic or type II (Crewther et al. 1978). About 25 different keratin proteins have been described so far in mammals, each with specific regulation of expression during development and differentiation (Moll et al. 1982; Cooper and Sun 1986).

All intermediate filament proteins have three domains: a central alpha-helical rod domain of constant size that derives from a common ancestor in all intermediate filament proteins, and two end-domains of variable structure (Steinert and Parry 1985). The variable domains have independent evolutionary histories (Klinge et al. 1987). Specific for basic keratins are two homologous globular subdomains, H1 and H2, bracketing the rod domain. A schematic diagram of this structure is shown in Fig. 1.

The cell and molecular biology of intermediate filament proteins has recently become the focus of a considerable amount of research (Hanukoglu and

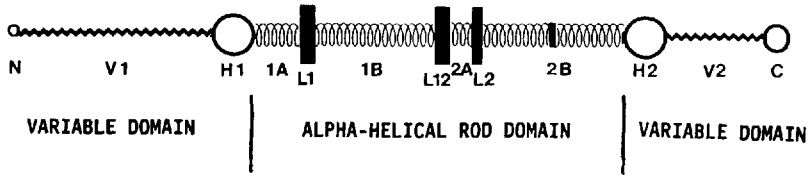

Fig. 1. Structure of intermediate filament proteins. All intermediate filament proteins have a central alpha-helical rod domain derived from a common ancestor. Alpha-helical subdomains 1A, 1B, 2A, and 2B are separated by nonhelical linker subdomains L1, L12, and L2. The variable end domains differ in size, composition, and organization of subdomains. The globular H1 and H2 subdomains, for instance, are characteristic of basic keratins, but not present in acidic keratins and nonepithelial intermediate filaments.

Fuchs 1982; Quax-Jeuken et al. 1983; Steinert et al. 1983; Lewis et al. 1984; Quax et al. 1985). However, the evolutionary events that generated this diverse gene family have not been explored systematically. Although subfamilies have been proposed among the acidic keratins, this proposal was based on orthologous relationships among four sequences and not on systematic comparisons of many keratin sequences (Jorcano et al. 1984b). The incremental model of keratin evolution, in which small keratins are ancestral to larger ones, which in turn gave rise to the largest ones (Sun et al. 1984), cannot be reconciled with the data presented in this manuscript.

Distance matrices have been used extensively to develop evolutionary trees of organisms (Fitch and Margoliash 1967) and intermediate filament sequences have also been compared using this method (Fraser et al. 1985; Hanukoglu and Fuchs 1986; Lewis and Cowan 1986). However, when sequence data are available, use of distance matrices that reduce the wealth of sequence comparison information into a single numerical value is inferior to the use of new computer programs capable of deriving evolutionary trees directly from sequence data (Fink 1986). The Phylogeny Analysis Using Parsimony (PAUP) set of programs (Swofford 1985) seems best suited for the analysis of keratin sequences (Fink 1986) and I have used it to derive evolutionary trees of keratin protein sequences. This analysis has demonstrated extraordinary parallelism of gene duplications in the alpha-helical regions of the two keratin gene families.

Sources and Methods

The sequences used in this study come from published sources and from unpublished work of my laboratory and several others. The sequences were entered into an IBM PC XT computer using DNA sequence analysis programs from International Biotechnologies Inc. (Pustel and Kafatos 1984). Table 1 contains a list of all the proteins analyzed as well as the source of the sequences and indicates the subdomains available for analysis. In the preparation of complementary DNA (cDNA) libraries, the sequences corresponding to the 3′ ends are overrepresented. Consequently, all reported sequences contain carboxy-proximal segments, whereas the sequences of the amino-terminal regions are not available for many of the reported proteins.

To format the sequences for analysis by the PAUP set of programs (Swofford 1985), a "desktop utilities" editor from "Homebase" or "Sidekick" was used so that files containing sequences of different families or subfamilies and different domains or subdomains could be constructed easily and conveniently. Sequences could be added to or removed from existing files as well. When only partial sequences are available, these are included only in analyses of the appropriate subdomains. The very similar termini of the central domains were used to align all the sequences. Gaps were introduced where necessary to keep the sequences aligned. These gaps occurred only in the linker subdomains.

The appropriately formatted files containing aligned intermediate filament protein sequences were first converted into numerical files using the "REDUCSEQ" program. The "noninformative" amino acid positions are defined as those in which all sequences are identical except one, and the mutations can be assigned to that one terminal branch of the evolutionary tree without affecting the branching order. Although their inclusion increases the computation time, these amino acid positions were included because they indicate the relative lengths of the terminal branches.

The PAUP program was always run using the "MULPARS" option, which ensures that the shortest tree is found. The "MULPARS" feature develops an initial tree from the sequences and then starts exchanging branches of this tree, "swapping," until all possible trees have been tested and the shortest tree is found. When several trees of identical length were found, they were integrated into consensus trees using the "CONTREE" program. This usually occurs when several branches separated from the common ancestor within a very short time interval so that the relative branching order cannot be inferred with confidence.

Wherever appropriate, the roots of the evolutionary trees of basic keratins were designated using acidic keratin sequences as an outgroup and vice versa. In the cases of the H2 subdomain sequences, the midpoint between the first two branchings was designated as the root.

Results

Evolutionary Tree of Acidic Keratins

The earliest gene duplication in the acidic keratin gene family was the segregation of the 45-kd protein ancestor gene from the ancestor of the rest of the acidic keratin genes (Fig. 2). No further gene duplications occurred in the 45-kd keratin branch.

At a significantly later date several subfamilies

Table 1. Keratin protein sequences compiled in the data bank

Number	Organism	Molecular weight[a]	Moll number[b]	Expressed in	Subdomains[c]	Reference
Acidic keratins						
1	Bovine	54 kd	10–11	Stratum corneum	1A, 1B, L12, 2B, V2	Rieger et al. (1985)
2	Murine	59 kd	10	Stratum corneum	1A, 1B, L12, 2B, V2	Steinert et al. (1983), Krieg et al. (1985)
3	Murine	47 kd	13	Esophagus	1B, L12, 2B, V2	Knapp et al. (1986)
4	Bovine	50 kd	14	Epidermis	V2	Jorcano et al. (1984b)
5	Human	50 kd	14	Basal cells	1A, 1B, L12, 2B, V2	Marchuk et al. (1984, 1985)
6	Human	Pseudo-50 kd	"14"	Not expressed	1A, 1B, L12, 2B, V2	Savtchenko et al. (in preparation)
7	Murine	52 kd	15	Basal cells	1A, 1B, L12, 2B, V2	Knapp et al. (1987)
8	Murine	52 kd′	15 (?)	Supabasal nonepidermal cells	V2	Knapp et al. (1987)
9	Murine	50 kd	16	Hyperproliferation	2B, V2	Knapp et al. (1987)
10	Human	46 kd	17	Simple epithelia	1A, 1B, L12, 2B, V2	RayChaudhury et al. (1986)
11	Human	45 kd	18	Trophoblasts	L12, 2B, V2	Leube et al. (1986), Oshima et al. (1986), Romano et al. (1986)
12	Murine	45 kd	18	Trophoblasts	1A, 1B, L12, 2B, V2	Singer et al. (1986)
13	Bovine	40 kd	19	Cultured cells	1A, 1B, L12, 2B, V2	Bader et al. (1986)
14	Ovine	"8C2"	—	Wool/hair	1A, 1B, L12, 2B, V2	K. Ward (personal communication)
15	Ovine	"8C1"	—	Wool/hair	1A, 1B, L12, 2B, V2	K. Ward (personal communication), Dowling et al. (1986)
16	Anuran	47 kd	—	Tadpole	1A, 1B, L12, 2B, V2	Winkles et al. (1985)
17	Anuran	"81"	—	Tadpole	1A, 1B, L12, 2B, V2	Jonas et al. (1985)
18	Anuran	"81B"	—	Tadpole	1A, 1B, L12, 2B, V2	Miyatani et al. (1986)
Basic keratins						
1	Bovine	67 kd	1	Stratum corneum	H2, V2	Jorcano et al. (1984a)
2	Human	67 kd	1	Stratum corneum	1A, 1B, L12, 2B, H2, V2	Johnson et al. (1985)
3	Murine	67 kd	1	Stratum corneum	1A, 1B, L12, 2B, H2, V2	Steinert et al. (1985)
4	Bovine	65 kd	3	Cornea	H2, V2	Jorcano et al. (1984a)
5	Human	65 kd	3	Cornea	1A, 1B, L12, 2B, H2, V2	Klinge et al. (1987)
6	Murine	57 kd	4	Esophagus	1A, 1B, L12, 2B, H2, V2	Knapp et al. (1986)
7	Bovine	60 kd	5	Basal cells	H2, V2	Jorcano et al. (1984a)
8	Murine	60 kd	5	Basal cells	1A, 1B, L12, 2B, H2, V2	Steinert et al. (1984)
9	Bovine	59 kd	6	Hyperproliferation	2B, H2, V2	Jorcano et al. (1984a)
10	Human	56 kd	6	Hyperproliferation	1B, L12, 2B, H2, V2	Hanukoglu and Fuchs (1983)
11	Human	56 kd′	6′	Hyperproliferation	1A, 1B, L12, 2B, H2, V2	Tyner et al. (1985)
12	Human	55 kd	7	Mesothelium	1A, 1B, L12, 2B, H2, V2	Glass et al. (1985)
13	Bovine	52 kd	8	Trophoblasts	1A, 1B, L12, 2B, H2, V2	Magin et al. (1986)
14	Ovine	"3"	—	Wool/hair	2B, H2, V2	K. Ward (personal communication)
15	Ovine	"7"	—	Wool/hair	2B, H2, V2	K. Ward (personal communication)
16	Ovine	"CHC6"	—	Wool/hair	1B	Crewther et al. (1978)
17	Anuran	56 kd	—	Oocytes	1A, 1B, L12, 2B, H2, V2	Franz and Franke (1986)
18	Anuran	"76"	—	Tadpole	1A, 1B, L12, 2B, H2, V2	T. Sargent (personal communication)
19	Anuran	64 kd	—	Adult skin	1B, L12, 2B, H2, V2	Hoffmann et al. (1985)
20	Anuran	64 kd′	—	Adult skin	1B, L12, 2B, H2, V2	Hoffmann et al. (1985)

[a] Apparent molecular weight in SDS polyacrylamide gels

[b] For catalog numbers see Moll et al. (1982)

[c] For the designation of subdomains see Steinert and Parry (1985). The amino terminal domains have not been analyzed. Acidic keratins do not have H2 globular domains. In individual proteins only those subdomains for which the entire subdomain sequence is known have been analyzed

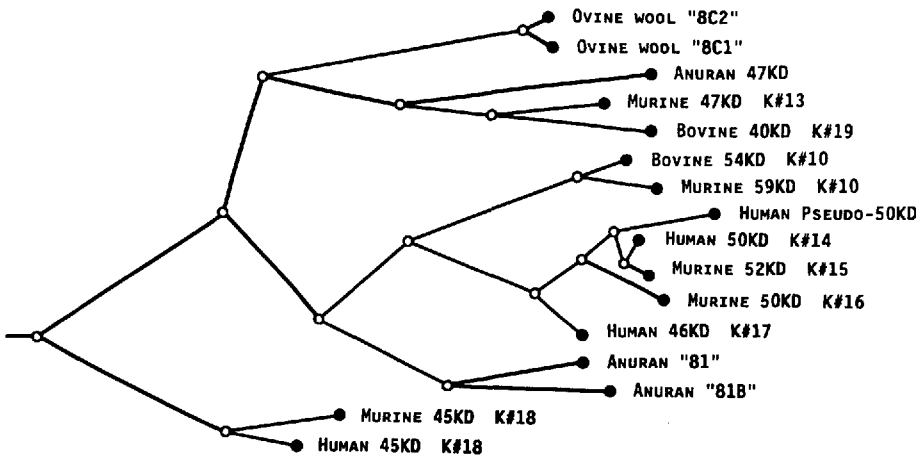

Fig. 2. Evolutionary tree of 2B subdomains of acidic keratins. The tree was rooted using basic keratins as an outgroup. Bovine 54-kd and murine 59-kd keratins are orthologs, as are the two 45-kd proteins.

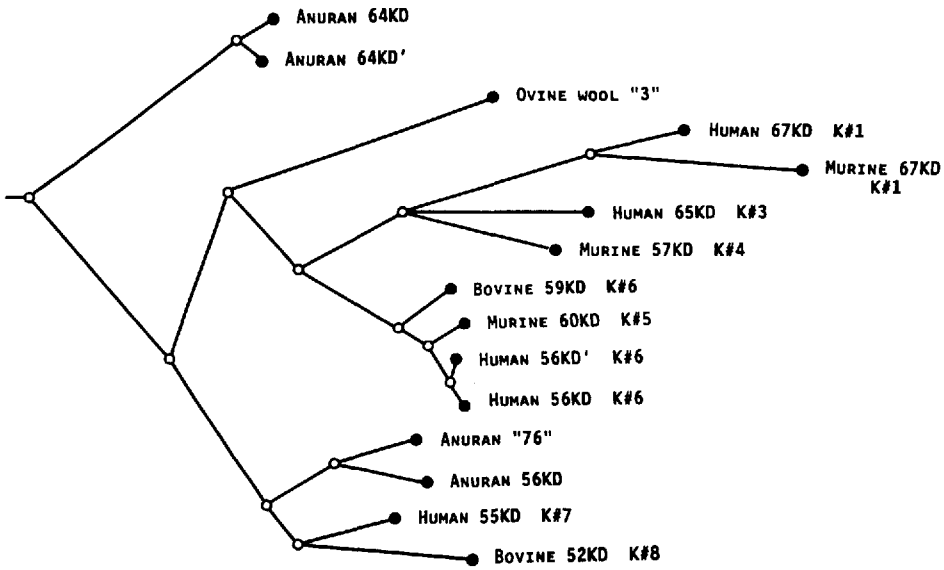

Fig. 3. Evolutionary tree of 2B subdomains of basic keratins. The tree was rooted using acidic keratins as an outgroup. The 67-kd, 65-kd, and 57-kd keratins branched virtually simultaneously from their common ancestor.

diverged within the alternative branch. Each keratin expressed in a different differentiation pathway represents a separate evolutionary subfamily, e.g., the hair/wool-specific keratins, or the murine 59-kd and bovine 54-kd keratins found in the suprabasal layers of skin. Keratins not associated with a particular differentiation pathway, but expressed in the basal layers of skin and other stratified epithelia, and keratins expressed under hyperproliferative conditions (i.e., human 50-kd and 46-kd keratins) represent another subfamily.

The divergence of mammals created identifiable orthologous sets of keratin genes (e.g., the murine 59-kd/bovine 54-kd pair). This occurrence was followed by at least one additional gene duplication event, the branching of the human 50-kd gene from a sequence we have designated the human 50-kd pseudogene. This last gene duplication occurred relatively recently, approximately 20 million years (Myr) ago (Savtchenko et al., in preparation).

Evolutionary Trees of Basic Keratins

The branch that contains the mammalian 52-kd and 55-kd keratins expressed in embryonic tissue and in mesothelial cells (Glass et al. 1985; Magin et al. 1986) also contains the two basic anuran keratins expressed in oocytes and in the tadpole (Franz and Franke 1986; T. Sargent, personal communication).

The genes coding for the two 64-kd keratins expressed in adult frog epidermis are very similar to each other (Fig. 3). These two keratins accumulated a large number of differences from their mammalian counterparts in the 2B subdomains (Fig. 1), although in the remaining parts of the molecules they are associated with the differentiation-associated branch (e.g., in Fig. 4).

The dispersal of the differentiation-specific basic keratins, as was the case in the acidic keratins, occurred within a relatively short evolutionary period. This dispersal created the branches of keratins ex-

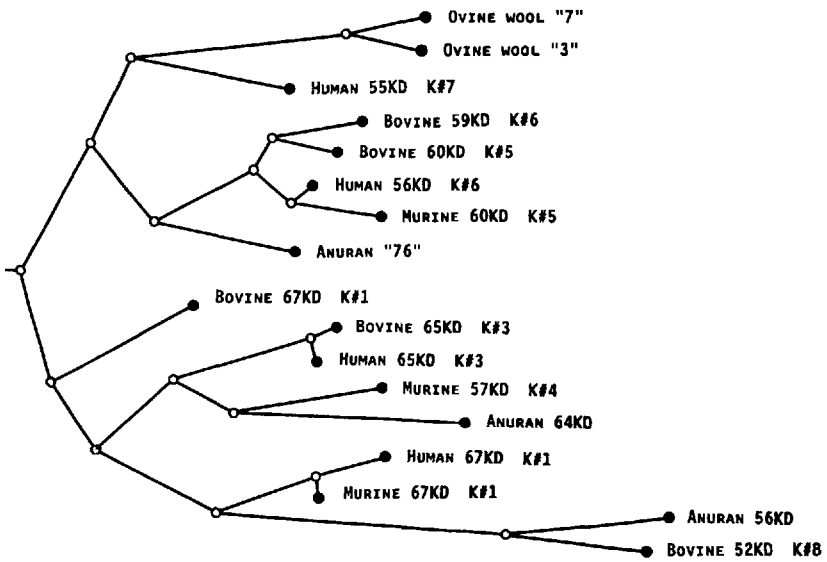

Fig. 4. Evolutionary tree of H2 subdomains of basic keratins. The tree was rooted halfway between the skin and hair keratin sequences. It illustrates the close relationships of orthologous keratin pairs.

pressed in hair, skin, cornea, esophagus, and other stratified, nonkeratinizing epithelia and the branch that later diverged again to give rise to the keratins expressed in the basal layer and under normal and hyperproliferative conditions. The branch leading to hair keratins has recently also undergone another burst of gene duplications to create the set of hair-specific basic keratin genes (Fig. 4).

The central alpha-helical domains of basic keratins derive from the common ancestor of all intermediate filaments, including acidic keratins. The variable end domains, on the other hand, are characteristic of individual proteins and they are expected to have different evolutionary histories. The nonhelical subdomains H2 (see Fig. 1) are unique to basic keratins and, at least in part, are responsible for the consistent size difference between the acidic and the basic keratins (Sun et al. 1984). The H2 subdomains may have a family-specific function. Therefore, I developed an evolutionary tree using only the H2 subdomain sequences (Fig. 4). For several basic keratins only carboxy-terminal sequences are available and these have been included in the analysis. Because of the larger number of sequences available, the tree in Fig. 4 is particularly illustrative of the close relationship among the orthologous human, bovine, and murine keratins.

There are several interesting differences in the topologies of the trees derived from the sequences of the 2B subdomains and the H2 subdomains: the adult frog keratins, which diverged extensively in the 2B subdomains, do not appear to be very distant from other basic keratins in their H2 subdomains. The embryonic, K#8 keratin, on the other hand, has diverged more extensively in its H2 subdomain.

A common feature of the trees presented in Figs. 3 and 4 is that each differentiation-specific keratin is on a separate branch except for the very closely related wool/hair subfamily and the subfamily of keratins expressed in the basal layer and under hyperproliferative conditions.

Discussion

Orthologous Relationships among Keratins

By definition, two genes (or proteins) are orthologous if they diverged at the same time that the species harboring them diverged. Patterns of expression and particularly the sequences of the 3′ untranslated segments (Yaffe et al. 1985) can be used as indicators of orthologous relationships among keratins, as evidenced by the human and bovine 65-kd keratins (Fig. 4) or the acidic murine 59- and bovine 54-kd pair (Fig. 2).

The relationships among the three 67-kd keratins from different species are puzzling. Although the orthologous relationship between the human and the murine species is beyond doubt, the bovine 67-kd keratin does not appear to be very closely related to the other two (Figs. 3 and 4). If the sequencing artifacts are excluded, three possibilities are suggested. First, the bovine 67-kd keratinization-specific protein may have diverged extensively during evolution. This possibility is supported by the fact that the human and murine 67-kd keratins also diverged extensively judging from the lengths of their branches (Fig. 3). Second, it may be that there are several genes in the bovine genome (and possibly human and murine genomes as well) coding for a set of basic keratin proteins of 67 kd. This possibility is supported by the precedents of two human 56-kd keratin genes (Hanukoglu and Fuchs 1983; Tyner et al. 1985) and the two very similar anuran keratins designated "81" and "88." The third possibility is

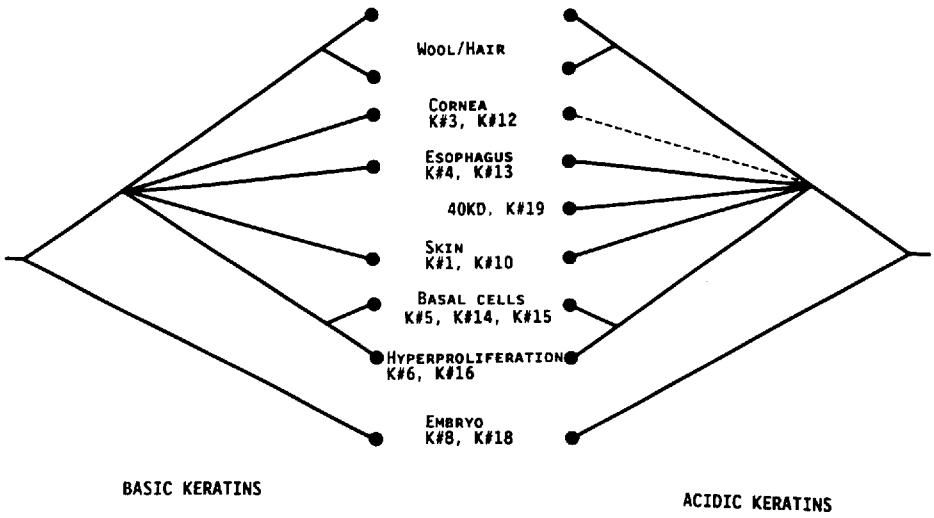

Fig. 5. Schematic representation of parallel gene duplications in the acidic and basic keratin families. Branches are not drawn to scale. The acidic cornea-specific keratin sequence is not available. The 40-kd acidic keratin specifically evolved to counteract overproduction of any one of the basic keratins.

that the bovine 67-kd keratin is a bovine snout-specific keratin not found in the skin and therefore not orthologous to the human and the murine 67-kd proteins, a possibility supported by the biochemical analysis of those tissues (Cooper and Sun 1986).

Gene duplications are very common and have occurred throughout the evolution of keratin genes. The paralogous sheep wool keratins are very similar to each other, which implies a recent divergence of their genes. Alternatively, the two genes have been kept uniform by gene conversion. From analysis of nonhelical keratin domains, it is already concluded that gene conversion mechanisms play a significant role in keratin evolution (Klinge et al. 1987).

Some gene duplications have occurred after species divergence. For instance, the two acidic murine 52-kd keratin genes (Knapp et al. 1987) appear to be orthologous to the human 50-kd keratin gene, which underwent a gene duplication event about 20 Myr ago (Savtchenko et al., in preparation). Since mice and men diverged about 80 Myr ago (Goodman 1981), either the genes duplicated independently in both the human and in the murine line, or, if the gene duplication occurred before the divergence of mammalian species, the human and the murine gene pairs evolved homogeneously by mutual gene conversions. I favor the first hypothesis, independent gene duplications, because one of the two human genes is a nonfunctional sequence, a pseudogene, and it is difficult to envision a process of natural selection that keeps a functional gene and a pseudogene homogeneous.

Evolution of Keratin Genes

The 45-kd keratin has a very special place in the family of acidic keratins that may explain its particular evolution. It was the first to diverge in the acidic keratin family (Fig. 2) and this protein, designated "endo B" in mouse, is the first keratin to appear during development. It is detectable in the 4- and 8-cell embryo (Oshima 1982), while all other keratins appear much later (Dale et al. 1985).

The branch containing the 45-kd keratins is significantly shorter than the other branches. This means that the 45-kd keratin remains similar to what its ancestor was at the time of gene duplication and that fewer mutations have been tolerated in this branch than in the other branches (Fig. 2).

The basic 52-kd keratin is most commonly found coexpressed with the acidic 40-kd and the 45-kd keratins. In data presented here, it is found on the same branch with the 55-kd protein, which is also generally found in mesothelial cells (Moll et al. 1982). The same branch contains the two anuran basic keratins expressed in oocytes and the tadpole, and the keratins expressed in adult frog skin are on a different branch (Figs. 3 and 4). Thus, the embryonic keratins are on a separate branch in the basic keratin family as well as in the acidic family.

It is perhaps more surprising that the evolution of the 40-kd keratin (Bader et al. 1986) does not stand out among the acidic keratins. This protein is expressed in very few tissues of the adult, but it is present in the fetal periderm (Dale et al. 1985), often in epithelial tumors and in epithelial cells in culture (Moll et al. 1982; Cooper and Sun 1986). It has been regarded as a marker of "dedifferentiation" of epithelial cells and it was proposed to be the most primitive keratin (Bader et al. 1986).

I would like to offer an alternative explanation: the 40-kd keratin has evolved to perform a highly specialized function. Temporary, unbalanced overproduction of basic keratin has been reported in cultured corneal cells and in stratifying epidermis (Schermer et al. 1986). In at least one case (Schermer et al. 1986), this imbalance is concomitant with expression of the 40-kd gene. Recent experiments

suggest that artificial induction of a basic keratin protein can turn on expression of an acidic keratin protein (Giudice and Fuchs 1987). When this control mechanism becomes disturbed, which results in unbalanced overproduction of (any!) basic keratin, the 40-kd acidic keratin may be synthesized to redress the imbalance.

This conclusion is supported by the fact that the 40-kd protein has a very short amino-terminal domain and does not have a carboxy-terminal domain. The tissue-specific function of keratins resides, at least in part, in the variable terminal domains. For instance, stratum corneum keratins have long hydrophobic glycine-rich stretches and hair keratins have cystine-rich regions capable of forming disulfide cross-links. Thus, the 40-kd keratin could assemble into intermediate filaments by pairing with any one of the basic keratins (Hatzfield and Franke 1985) without contributing a potentially harmful variable domain. Moreover, the 40-kd keratin is the only one that does not have, and according to this hypothesis does not need, a coordinately expressed keratin protein from the other family. If this hypothesis is correct, the synthesis of the 40-kd keratin in transformed epithelial cell lines is not surprising.

Concerted Gene Duplications in the Acidic and Basic Keratin Gene Families

The embryonic keratins were the first to diverge in both families, as discussed above. In the nonembryonic branches of both the acidic and the basic families, hair, epidermis, and esophageal keratins each are associated with a separate evolutionary branch. The most recent gene duplications in both families are those that created the two subfamilies of hair keratins and the two subfamilies consisting of keratins that are expressed in basal cells and under hyperproliferative conditions. This parallelism of gene duplications is schematically represented in Fig. 5.

Because keratin intermediate filaments are obligate heteropolymers, consisting of at least one member from each of the two families, it can be expected that members of the two families influence each others evolution. Coevolution of acidic and basic keratins is evident in the variable terminal domains of glycine-rich keratins and we have proposed gene conversion between acidic keratin genes and the basic keratin genes as one of the possible mechanisms for this coevolution (Klinge et al. 1987). But conversion cannot explain concerted gene duplications in two families. Concerted gene duplications are known in several multisubunit protein systems, e.g., silk moth chorion genes (Burke and Eickbush 1986) and mouse urinary proteins (Ghazal et al. 1985), systems with the two components extremely tightly linked in the genome. The duplication events simultaneously amplify the genes of both components. Although some keratin genes appear linked, they are not linked in acidic–basic keratin gene pairs (Blumenberg and Savtchenko 1986; Powell et al. 1986; RayChaudhury et al. 1986) and therefore the same mechanism does not apply to keratin genes. The concerted duplication of acidic and basic keratin genes must proceed via a different, as yet unknown, mechanism.

Acknowledgments. Thanks to K. Ward, T. Sargent, R. Liem, and W.W. Franke for making sequences available before publication. I also thank I.M. Freedberg for ideas and encouragement. The research was supported by grants AM30682, AM35290, and RR05399 from the National Institutes of Health, an American Cancer Society Institutional Grant to N.Y.U., a Dermatology Foundation Lila Gruber Research Award, and a grant from the R.L. Baer Foundation.

References

Bader BL, Magin TM, Hatzfield M, Franke WW (1986) Amino acid sequence and gene organization of cytokeratin no. 19, an exceptional tail-less intermediate filament protein. EMBO J 5:1865–1875

Blumenberg M, Savtchenko ES (1986) Linkage of human keratin genes. Cytogenet Cell Genet 42:65–71

Burke WD, Eickbush TH (1986) The silkmoth late chorion locus. I Variation within two paired multigene families. J Mol Biol 190:343–356

Cleveland DW, Sullivan KF (1985) Molecular biology and genetics of tubulins. Annu Rev Biochem 54:331–365

Cooper D, Sun T-T (1986) Monoclonal antibody analysis of bovine epithelial keratins. J Biol Chem 261:4646–4654

Crewther WG, Inglis AS, McKern NM (1978) Amino acid sequences of alpha-helical segments from *S*-carboxymethylkeratin-A. Biochem J 173:365–371

Dale BA, Holbrook KA, Kimball JR, Hoff M, Sun T-T (1985) Expression of epidermal keratins and filaggrin during human fetal skin development. J Cell Biol 101:1257–1269

Dowling LM, Crewther WG, Inglis AS (1986) The primary structure of component 8C-1, a subunit protein of intermediate filaments in wool keratins. Biochem J 236:695–703

Fink WL (1986) Microcomputers and phylogenetic analysis. Science 234:1135–1139

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Franz J, Franke WW (1986) Cloning of cDNA and amino acid sequence of a cytokeratin expressed in oocytes of *Xenopus laevis*. Proc Natl Acad Sci USA 83:6475–6479

Fraser RDB, MacRae TP, Suzuki E, Parry DAD, Trajstman AC, Lucas I (1985) Intermediate filament structure: molecular interactions in the filament. Int J Biol Macromol 7:214–230

Geisler N, Weber K (1986) Structural aspects of intermediate filaments. In: Shay JW (ed) Cell and molecular biology of the cytoskeleton. Plenum, New York, pp 41–68

Ghazal P, Clark AJ, Bishop JO (1985) Evolutionary amplification of a pseudogene. Proc Natl Acad Sci USA 82:4182–4185

Giudice GJ, Fuchs E (1987) The transfection of epidermal keratin genes into fibroblasts and simple epithelial cells: evidence for inducing a type I gene by a type II gene. Cell 48:453–463

Glass C, Kim KH, Fuchs E (1985) Sequence and expression of

a human type II mesothelial keratin. J Cell Biol 101:2366–2373

Goodman M (1981) Decoding the pattern of protein evolution. Prog Biophys Mol Biol 37:105–164

Hanukoglu I, Fuchs E (1982) The cDNA sequence of a human epidermal keratin: divergence of sequence but conservation of structure among intermediate filament proteins. Cell 31: 243–252

Hanukoglu I, Fuchs E (1983) The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. Cell 33:915–924

Hanukoglu I, Fuchs E (1986) Structural diversity and evolution of intermediate filament proteins. In: Karlin S, Nevo E (eds) Evolutionary processes and theory. Academic Press, New York, pp 69–98

Hatzfeld M, Franke WW (1985) Pair formation and promiscuity of cytokeratins: formation in vitro of heterotypic complexes and intermediate-sized filaments by homologous and heterologous recombination of purified polypeptides. J Cell Biol 101:1826–1841

Hoffmann W, Franz JK, Franke WW (1985) Amino acid sequence microheterogeneities of basic (type II) cytokeratins of *Xenopus laevis* epidermis and evolutionary conservativity of helical and non-helical domains. J Mol Biol 184:713–724

Johnson LD, Idler WW, Zhou X-M, Roop DR, Steinert PM (1985) Structure of a gene for the human epidermal 67-kDa keratin. Proc Natl Acad Sci USA 82:1896–1900

Jonas E, Sargent TD, Dawid IB (1985) Epidermal keratin gene expressed in embryos of *Xenopus laevis*. Proc Natl Acad Sci USA 82:5413–5417

Jorcano JL, Franz JK, Franke WW (1984a) Amino acid sequence diversity between bovine epidermal cytokeratin polypeptides of the basic (type II) subfamily as determined from cDNA clones. Differentiation 28:155–163

Jorcano JL, Rieger M, Franz JK, Schiller DL, Moll RL, Franke WW (1984b) Identification of two types of keratin polypeptides within the acidic cytokeratin subfamily I. J Mol Biol 179:257–281

Klinge EM, Sylvestre YR, Freedberg IM, Blumenberg M (1987) Evolution of keratin genes: different protein domains evolve by different mechanisms. J Mol Evol 24:319–329

Knapp B, Rentrop M, Schweizer J, Winter H (1986) Nonepidermal members of the keratin multigene family: cDNA sequences and in situ localization of the mRNAs. Nucleic Acids Res 14:751–763

Knapp B, Rentrop M, Schweizer J, Winter H (1987) Three cDNA sequences of mouse type I keratins. J Biol Chem 262: 938–945

Krieg TM, Schafer MP, Cheng CK, Filpula D, Flaherty P, Steinert PM, Roop DR (1985) Organization of a type I keratin gene. J Biol Chem 260:5867–5870

Leube RE, Bosch FX, Romano V, Zimbelmann R, Hoefler H, Franke WW (1986) Cytokeratin expression in simple epithelia. III Detection of mRNAs encoding human cytokeratins nos. 8 and 18 in normal and tumor cells by hybridization with cDNA sequences in vitro and in situ. Differentiation 33: 69–85

Lewis SA, Cowan NJ (1986) Anomalous placement of introns in a member of the intermediate filament multigene family: an evolutionary conundrum. Mol Cell Biol 6:1529–1534

Lewis SA, Balcarek JM, Krek V, Shelanski M, Cowan NJ (1984) Sequence of a cDNA clone encoding mouse glial fibrillary acidic protein: structural conservation of intermediate filaments. Proc Natl Acad Sci USA 81:2743–2746

Magin TM, Jorcano JL, Franke WW (1986) Cytokeratin expression in simple epithelia. II. cDNA cloning and sequence characteristics of bovine cytokeratin A (no. 8). Differentiation 30: 254–264

Marchuk D, McCrohon S, Fuchs E (1984) Remarkable conservation of structure among intermediate filament genes. Cell 39:491–498

Marchuk D, McCrohon S, Fuchs E (1985) Complete sequence of a gene encoding a human type I keratin: sequences homologous to enhancer elements in the regulatory region of the gene. Proc Natl Acad Sci USA 82:1609–1613

Miyatani S, Winkles JA, Sargent TD, Dawid IB (1986) Stage-specific keratins in *Xenopus laevis* embryos and tadpoles: the XK81 gene family. J Cell Biol 103:1957–1965

Moll R, Franke WW, Schiller DL, Geiger B, Krepler R (1982) The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell 31:11–24

Oshima RG (1982) Developmental expression of murine extra-embryonic endodermal cytoskeletal proteins. J Biol Chem 257:3414–3421

Oshima RG, Millan JL, Cecena G (1986) Comparison of mouse and human keratin 18: a component of intermediate filaments expressed prior to implantation. Differentiation 33:61–68

Powell BC, Cam GR, Fietz MJ, Rogers GE (1986) Clustered arrangement of keratin intermediate filament genes. Proc Natl Acad Sci USA 83:5048–5052

Pustel J, Kafatos FC (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res 12:643–655

Quax W, van den Broek L, Egberts WV, Ramaekers F, Bloemendal H (1985) Characterization of the hamster desmin gene: expression and formation of desmin filaments in non-muscle cells after gene transfer. Cell 43:327–338

Quax-Jeuken YEFM, Quax WJ, Bloemendal H (1983) Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence. Proc Natl Acad Sci USA 80:3548–3552

RayChaudhury A, Marchuk D, Lindhurst M, Fuchs EV (1986) Three tightly linked genes encoding human type I keratins: conservation of sequence in the 5′-untranslated leader and 5′-upstream regions of coexpressed keratin gene. Mol Cell Biol 6:539–548

Rieger M, Jorcano JL, Franke WW (1985) Complete sequence of a bovine type I cytokeratin gene: conserved and variable intron positions in genes of polypeptides of the same cytokeratin subfamily. EMBO J 4:2261–2267

Romano V, Hatzfeld M, Magin TM, Zimbelmann R, Franke WW, Maier G, Ponstingl R (1986) Cytokeratin expression in simple epithelia. I. Identification of mRNA coding for human cytokeratin no. 18 by a cDNA clone. Differentiation 30:244–253

Schermer A, Galvin S, Sun T-T (1986) Differentiation-related expression of a major 64K corneal keratin in vivo and in culture suggests limbal location of corneal epithelial stem cells. J Cell Biol 103:49–62

Singer PA, Trevor K, Oshima RG (1986) Molecular cloning and characterization of the endo B cytokeratin expressed in preimplantation mouse embryos. J Biol Chem 261:538–547

Steinert PM, Parry DAD (1985) Intermediate filaments. Annu Rev Cell Biol 1:41–65

Steinert PM, Rice RH, Roop DR, Trus BL, Steven AC (1983) Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. Nature 302:794–799

Steinert PM, Parry DAD, Racoosin EL, Idler WW, Steven AC, Trus BL, Roop DR (1984) The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin of 60,000 Da: analysis of sequence differences between type I and type II keratins. Proc Natl Acad Sci USA 81:5709–5713

Steinert PM, Parry DAD, Idler WW, Johnson LD, Steven AC,

Roop DR (1985) Amino acid sequences of mouse and human epidermal type II keratins of M_r 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits. J Biol Chem 260:7142–7149

Sun T-T, Eichner R, Schermer A, Cooper D, Nelson DW, Weiss RA (1984) Classification, expression and possible mechanism of evolution of mammalian keratin: a unifying model. In: Levine AJ, Topp WC, Van de Woude GF, Watson JD (eds) The cancer cell, vol 1. Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp 169–176

Swofford DL (1985) PAUP phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign IL

Traub P (1985) Intermediate filaments. Springer-Verlag, New York

Tyner AL, Eichman MJ, Fuchs E (1985) The sequence of a type II keratin gene expressed in human skin: conservation of structure among all intermediate filament genes. Proc Natl Acad Sci USA 82:4683–4687

Vandekerckhove J, Weber K (1979) Complete amino acid sequence of actins from bovine aorta; bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. Differentiation 14:123–133

Weber K, Geisler N (1985) Intermediate filaments: structural conservation and divergence. Ann NY Acad Sci 455:126–143

Winkles JA, Sargent TD, Parry DAD, Jonas E, Dawid IB (1985) Developmentally regulated cytokeratin gene in *Xenopus laevis*. Mol Cell Biol 5:2575–2581

Yaffe D, Nudel V, Mayer Y, Neuman S (1985) Highly conserved sequences in the 3′ untranslated regions of mRNAs coding for homologous proteins in distantly related species. Nucleic Acids Res 13:3723–3727

Received September 24, 1987/Revised and accepted February 5, 1988